Workshop on Uncertainty Quantification in Climate Modeling and Projection
ITCP, Trieste, Italy, July 13-17, 2015

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science
GCESS

Quantification of Parametric Uncertainty of Large Complex Geophysical Models

Qingyun Duan

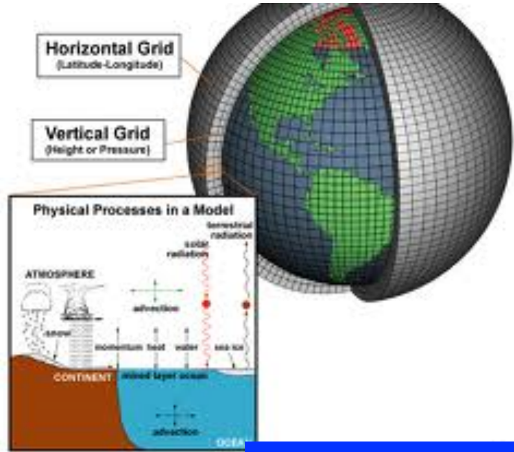
GCESS/Beijing Normal University

July 13, 2015

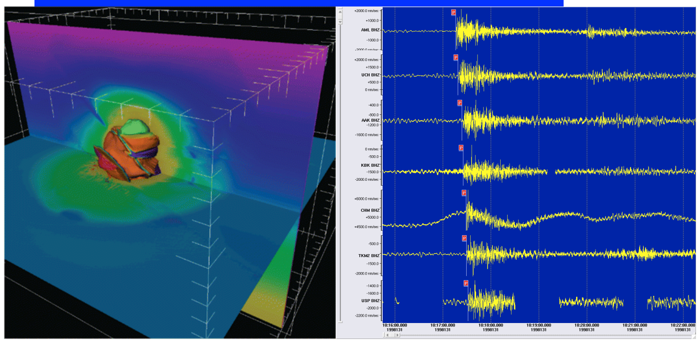
Acknowledgment: My team at BNU: Chen Wang, Wei Gong, Zhenhua Di, Yanjun Gan, Jianduo Li, Jiping Quan; my formal colleagues at LLNL: Charles Tong, Yunwei Sun

Simulation-based Science Provides A Powerful Complement to Experimental Science

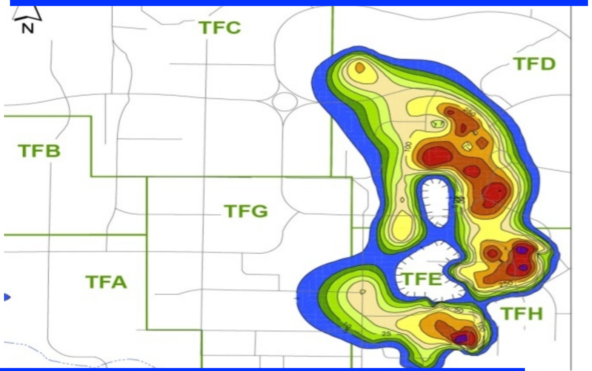
Weather and Climate Prediction



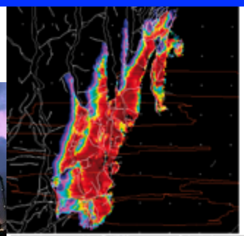
Simulation of Underground Nuclear Exploration



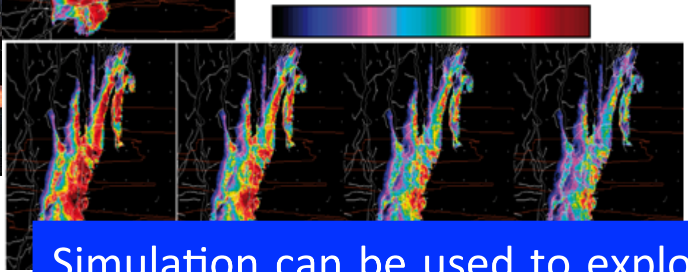
Environmental Management



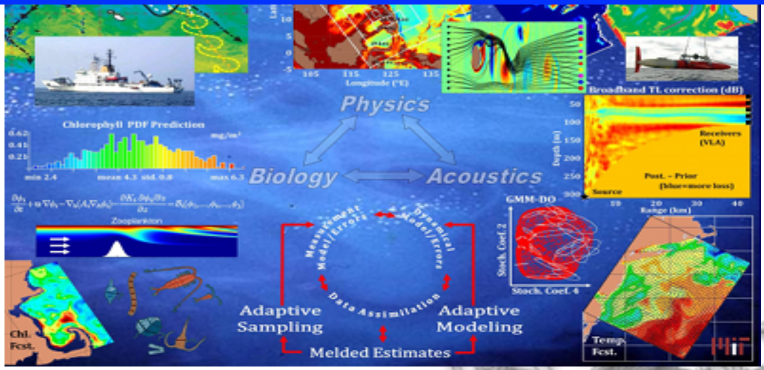
Deep Ocean Oil Reservoir Simulation



Upper: Oil saturation map within the Tarbert Fm from the Gullfaks reservoir model, 1985. Lower: Attribute maps from 4D inversion showing the averaged oil saturation value, extracted from 4D inversions from surveys in 1996, 1999, 2003 and 2005. Warm colours indicate high oil saturation.

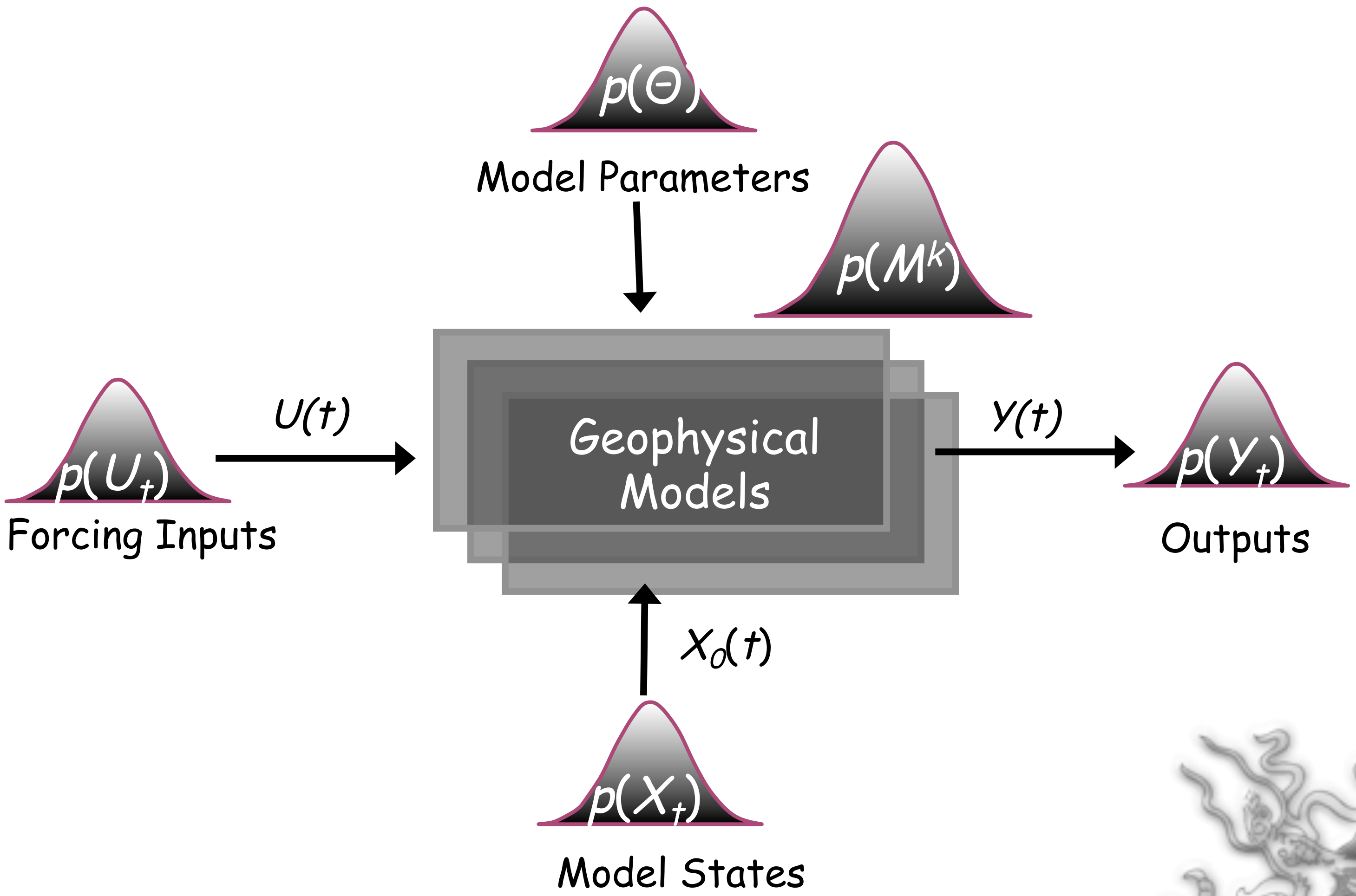


Environmental-acoustic Dynamics and Predictabilities



Simulation can be used to explore new theories and to design new experiments to test these theories. It also provides a powerful alternative to the experimental science when phenomena are not observable or measurements are impractical or too expensive

Uncertainty in Geophysical Model Simulation



Example: Uncertainty in Land Surface Modeling

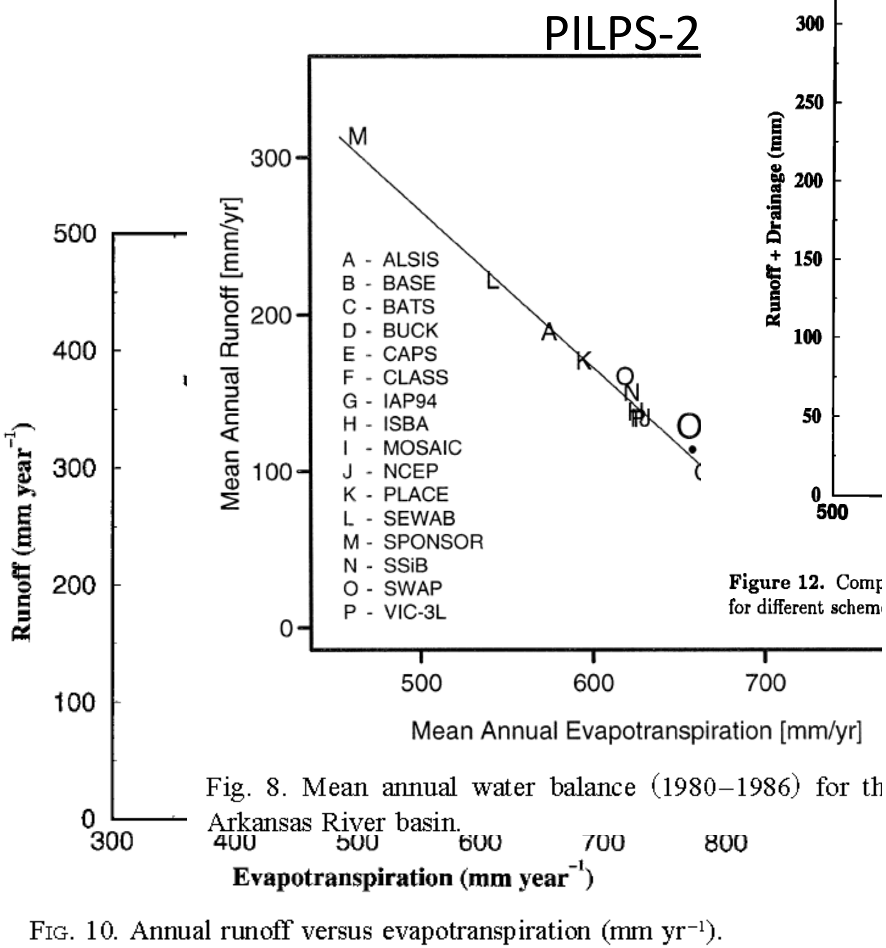
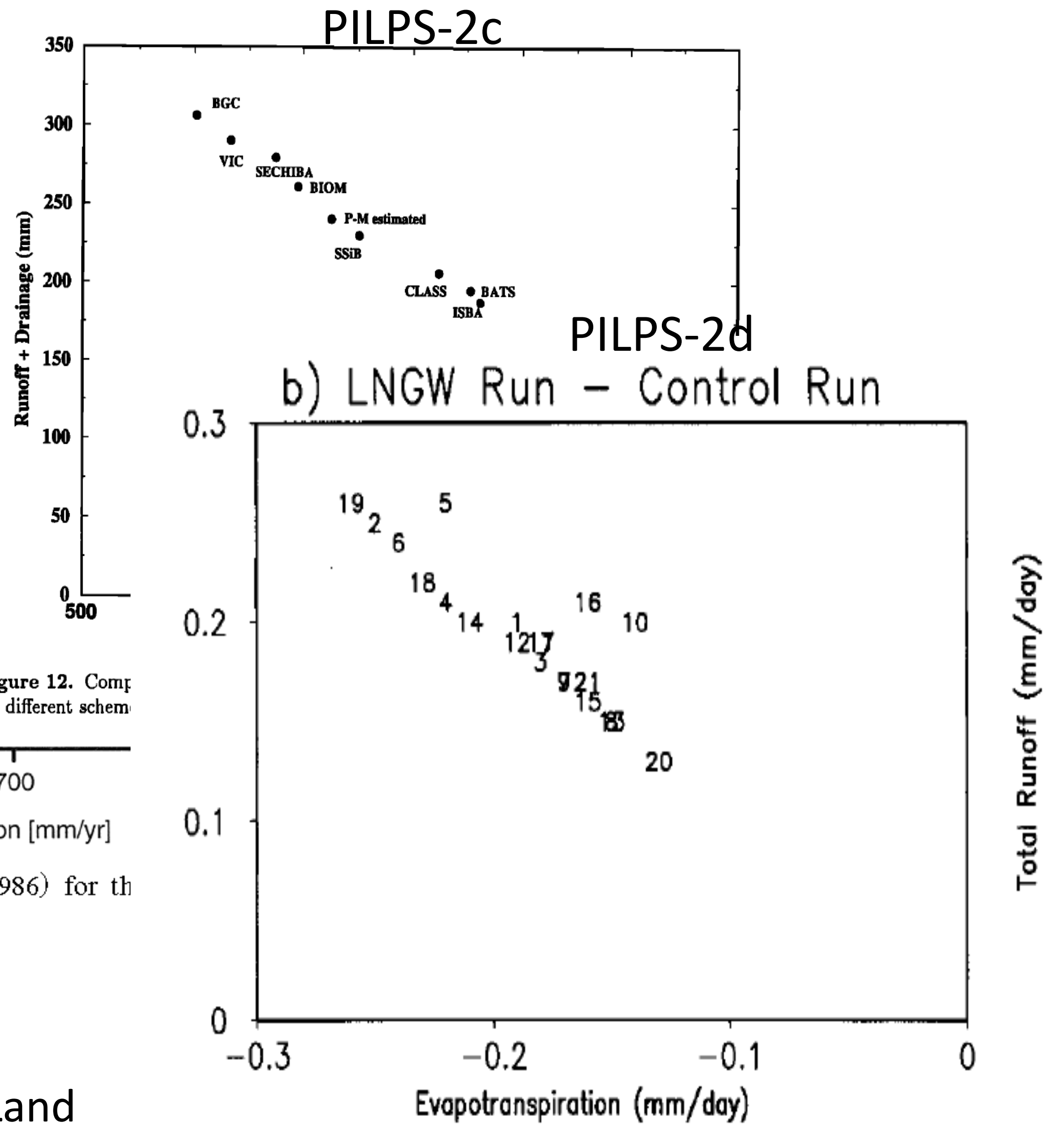


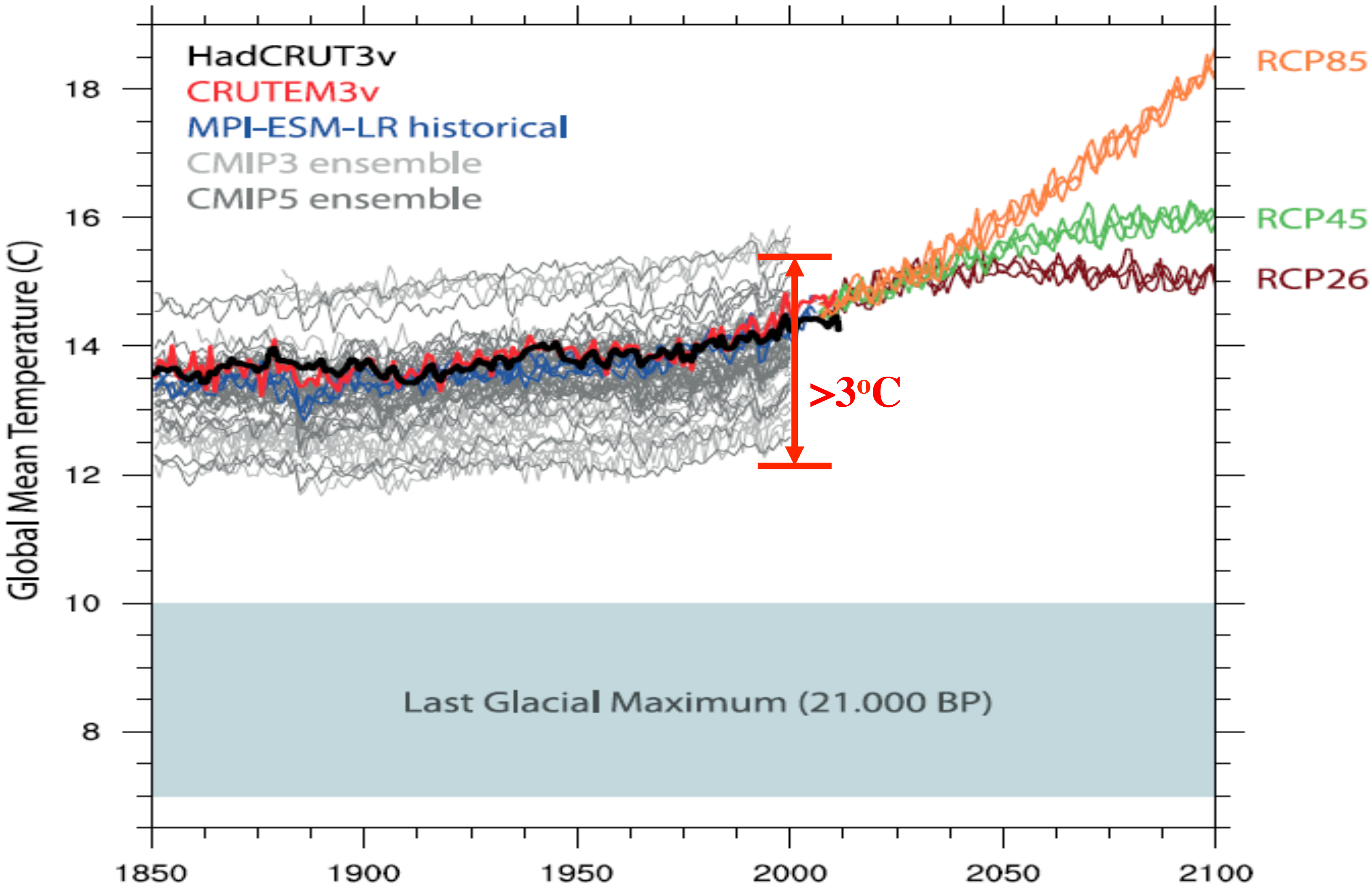
Fig. 8. Mean annual water balance (1980–1986) for th Arkansas River basin.

FIG. 10. Annual runoff versus evapotranspiration (mm yr^{-1}).



PILPS – Project for Intercomparison of Land Surface Parameterization Schemes

Example: Global Mean Temperature Simulation and Projection in CMIP3 & CMIP5



Mauritsen et al., 2012

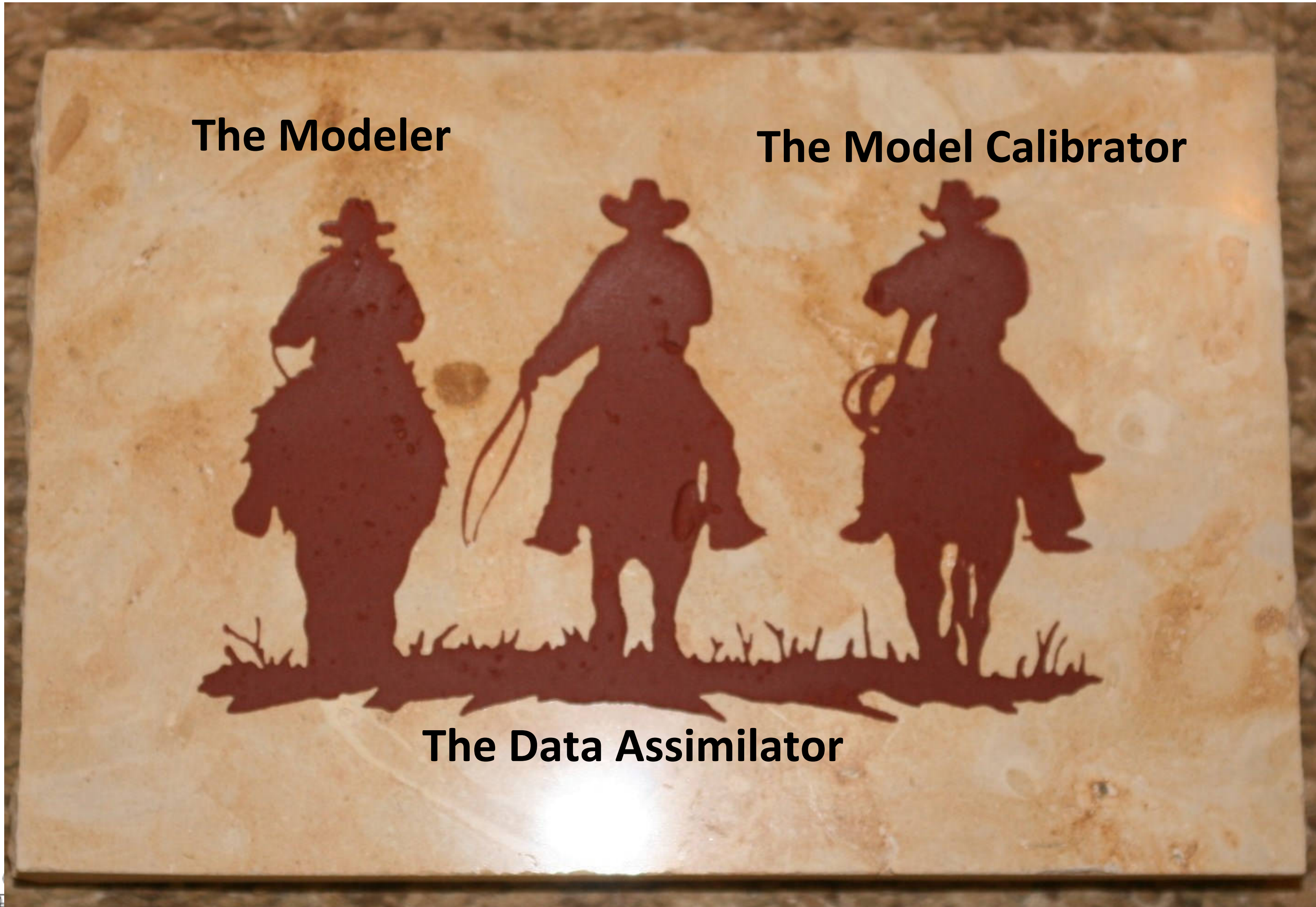
Large Complex Geophysical Model Simulations?

- Enhance the Model Physical Representation
  - Better models
  - Higher space/time resolution
  - Better numerical schemes

- Enhance the Representation of External Forcings and Boundary Conditions
  - Better Observations and Assimilation of Observations

- Enhance the Estimation of Model Parameters
  - Tuning model parameters so simulations match observations

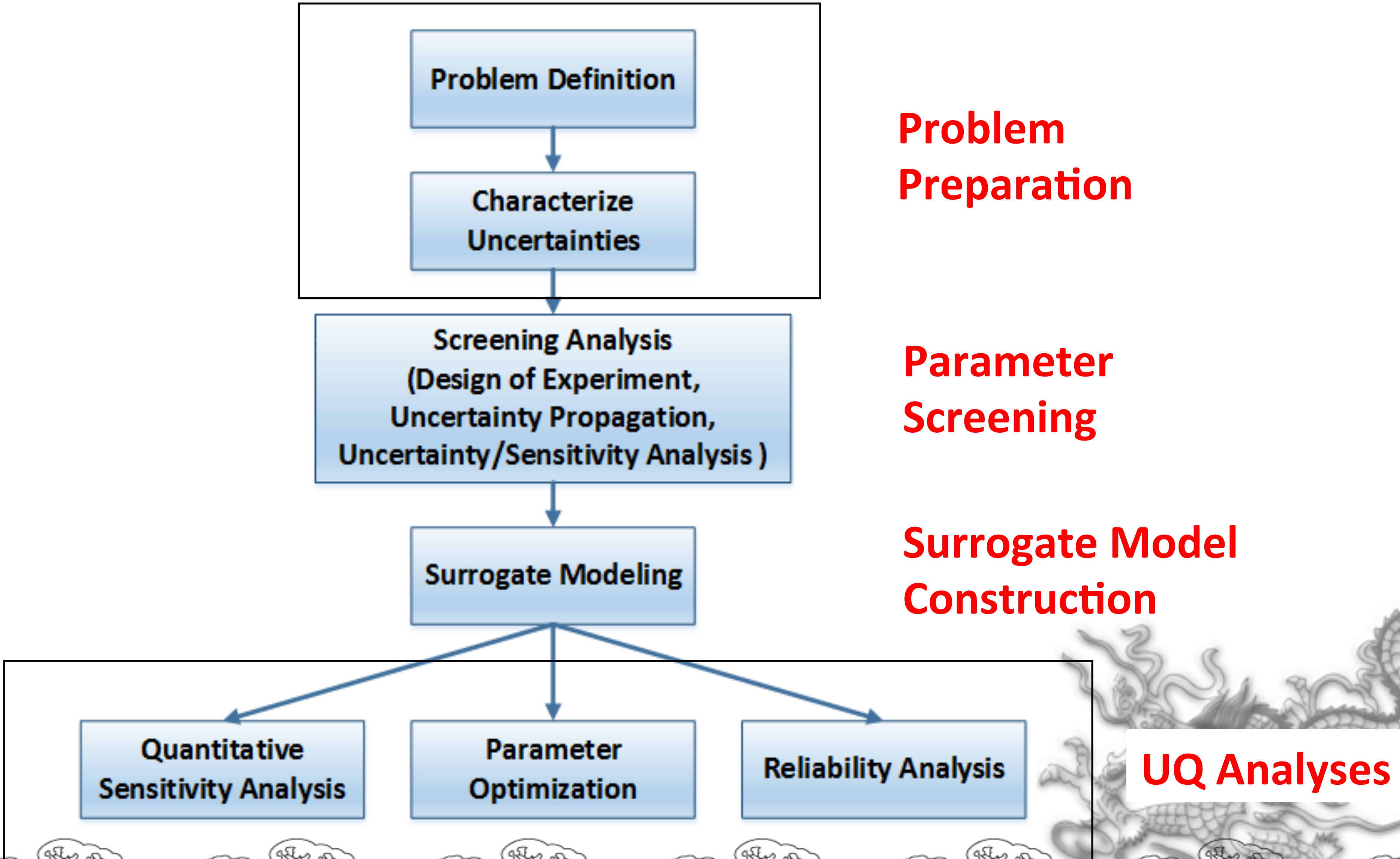
The Three Horsemen of Model Improvement



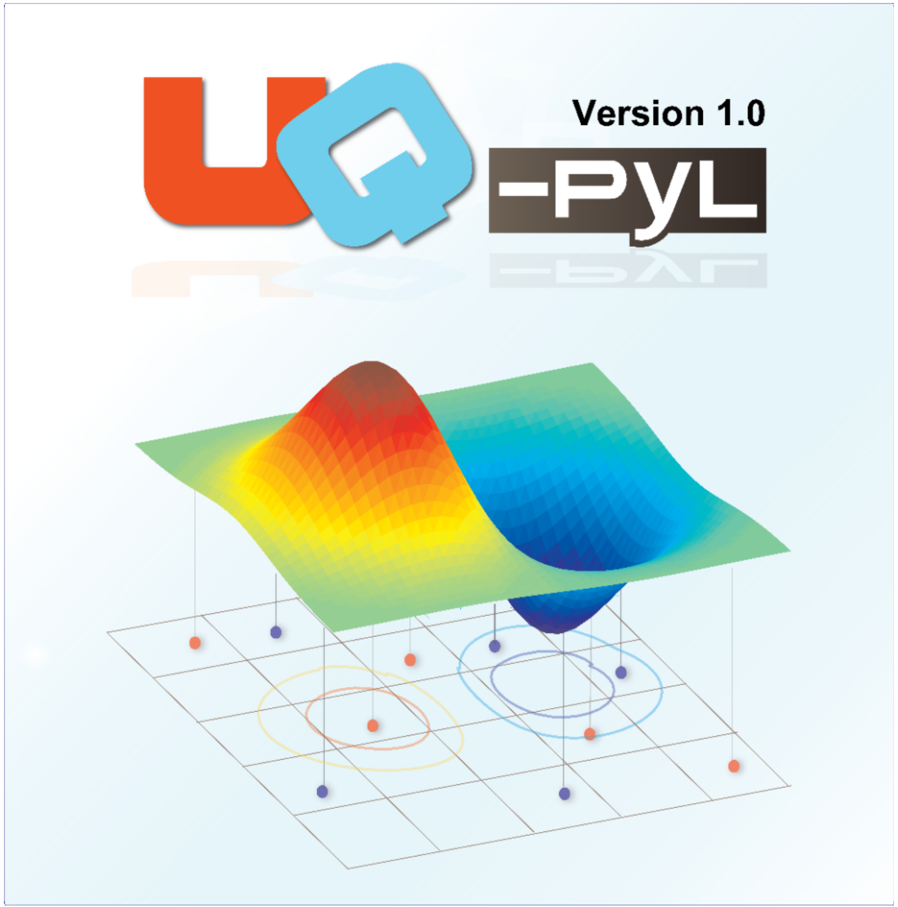
Challenges in Uncertainty Analysis for Large Complex Geophysical Models

- High-dimensionality of the uncertain parameters (10's -100's)
- High-dimensionality of the model outputs (can be millions)
- Difficult to prescribe parameter uncertainties (the priors)
- Models may be expensive to evaluate (many CPU-hours)
- Complex models show highly nonlinear (may be discontinuous) input-output relationships
- Data scarcity for the full system (difficult to calibrate)
- Models are often created by data far from operating conditions
  - extrapolation may be needed
- "Unknown unknowns" can greatly complicate the UQ process.

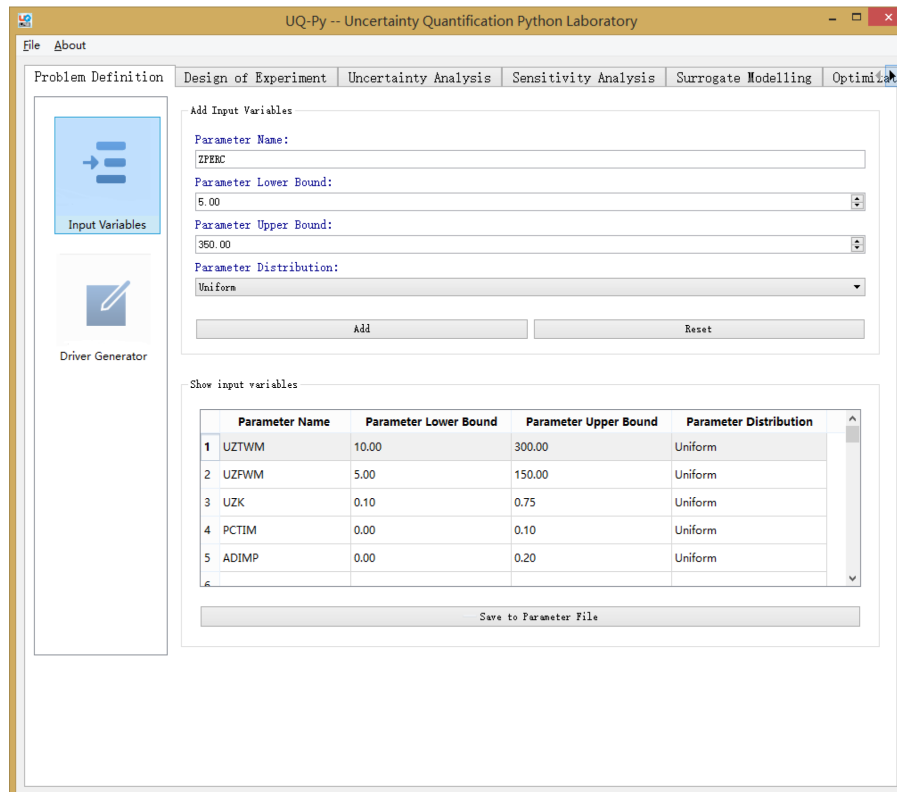
A UQ Methodology For Large Complex Geophysical Models



The Uncertainty Quantification Python Laboratory developed at Beijing Norma University



Splash page of UQ-PyL



Main page of UQ-PyL

What is UQ-PyL?

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science
GCESS

- A new, general-purpose, cross-platform UQ framework;
- Made of several components that perform various UQ functions, including
  - *Design of Experiments*
  - *Uncertainty Analysis*
  - *Sensitivity Analysis*
  - *Surrogate Modeling*
  - *Parameter Optimization;*
- Suitable for parametric uncertainty analysis of any computer simulation models;
- Has a Graphic User Interface.

A Review of Some UQ Methodologies

- Design experiment (DOE)
- Parameter screening (Dimension reduction)
- Surrogate modeling (Response surface analysis, meta-modeling, statistical emulation)
- Model calibration (Parameter optimization)

Design of Experiment (DOE)

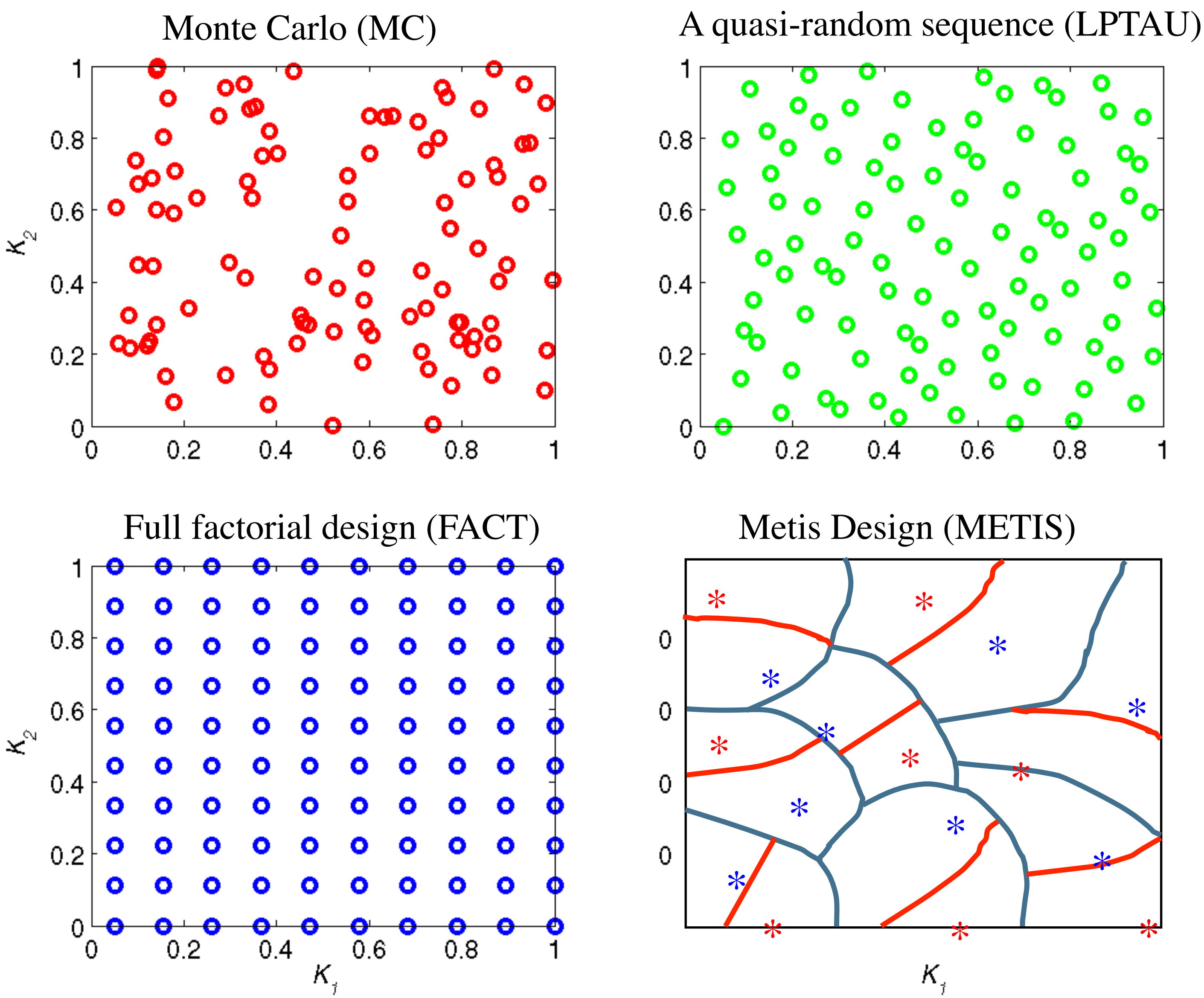
Design of Experiments (DOE)

- Statistical methodology for systematically investigating a system's input-output relationship to achieve one of several goals:
  - Identify important design variables (screening)
  - Optimize product or process design
  - Achieve robust performance

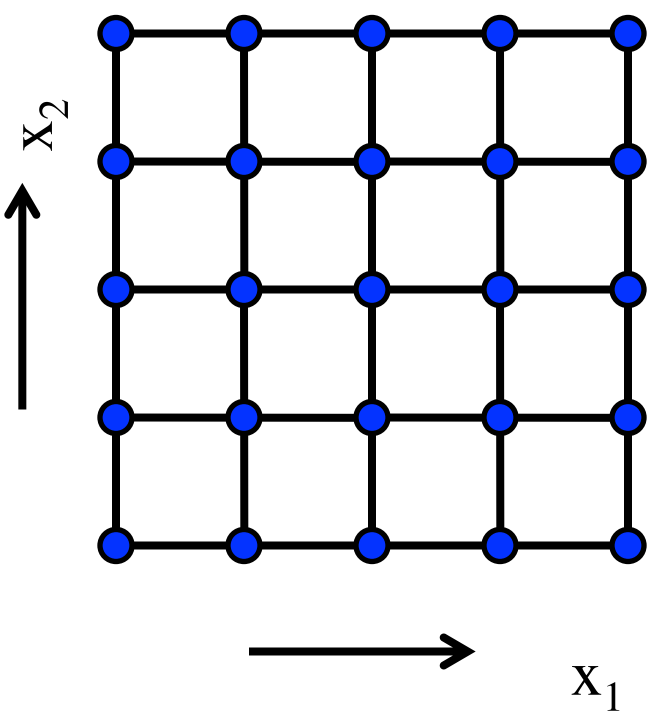
Type of Design of Experiment Methods

- Deterministic Design
  - Full Factorial design, Fractional Factorial design, Plackett-Burman design, Box-Behnken design, Central-Composite design, …
- Random Design
  - Monte Carlo design, Latin Hypercube design, Symmetric Latin Hypercube design, …
- Quasi-random Design
  - QMC Sobol' sequence, QMC Halton sequence, QMC Hammersley sequence, QMC faure sequence, …

Examples of Different DoEs Methods

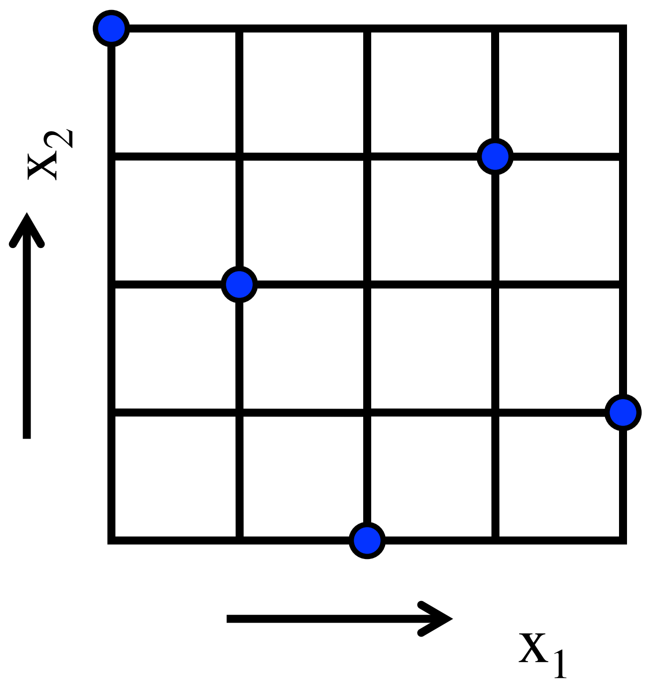


What Is A Full Factorial Design?



- Space-filling in all dimensions
- Sample size = s^m
- s: number of levels
- m: number of inputs
- Can be randomized by small perturbations
- Can resolve m-way interactions
- Only suitable for small number of inputs (expensive)

What Is Latin Hypercube Design?

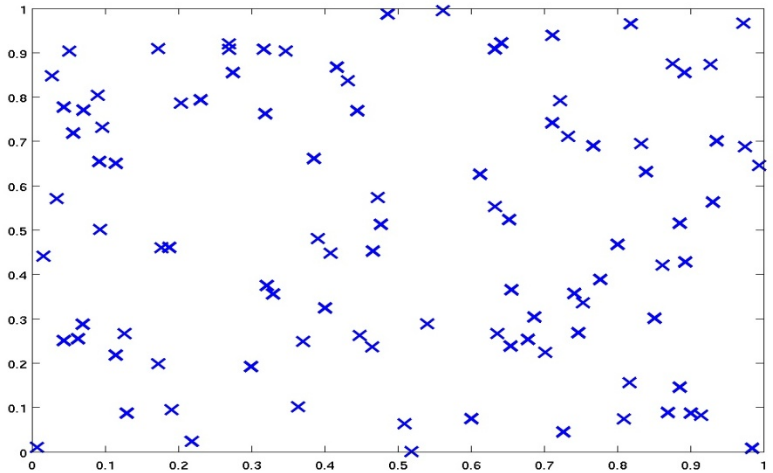


Latin hypercube
(stratified in each dimension)

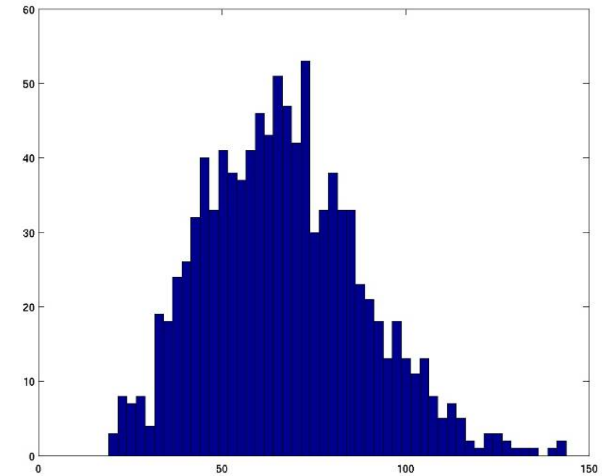
- A highly fractional factorial design
- Space-filling in any one dimension
- Faster convergence than MC
- Esp. for monotonic functions
- LHS(N, m, s) + noise
- N: sample size (5 here)
- m: number of parameters
- s: number of symbols
- r = N/s: number of replications
- How to choose sample size?
- Sampling refinement

Example: Monte Carlo Sampling – A Classical UQ Exercise

- Create N random sample points in the uncertain parameter space
- Run the points through the function and gather the Y's
- Compute basic statistical quantities: mean, std. dev.
- Bin the Y's and create an output histogram



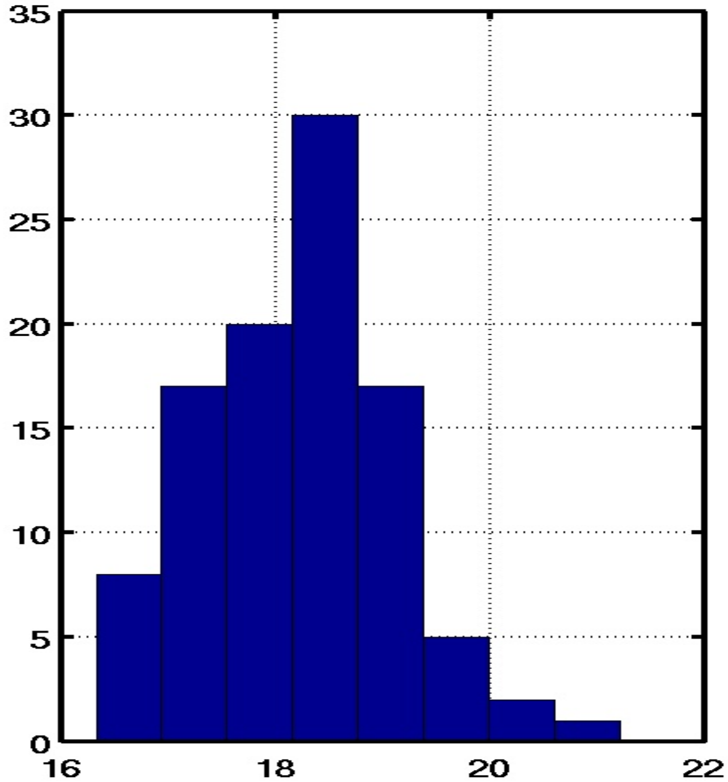
Sample points in parameter space

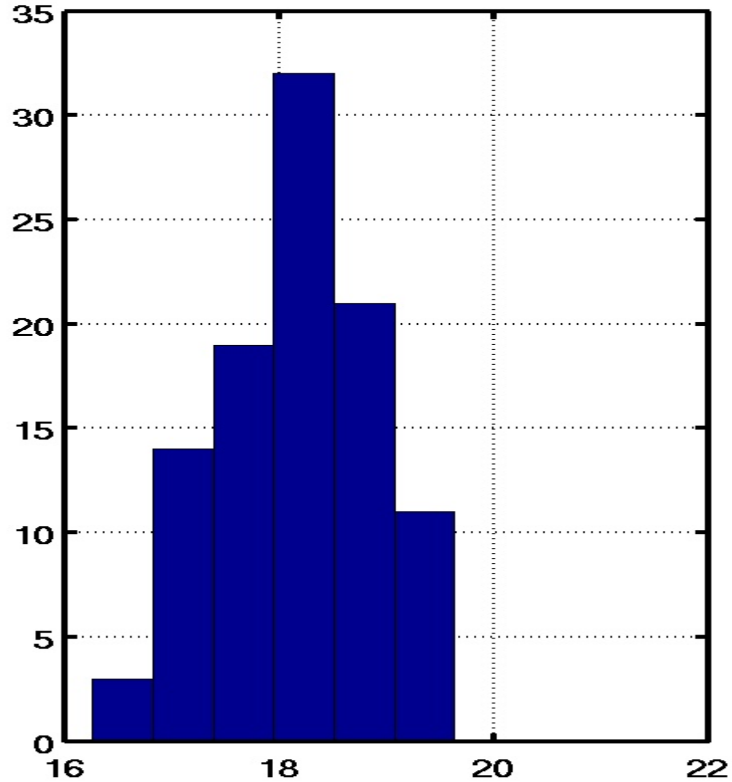


An example output distribution

Does Sampling Size Matter?

Distribution of the sample means



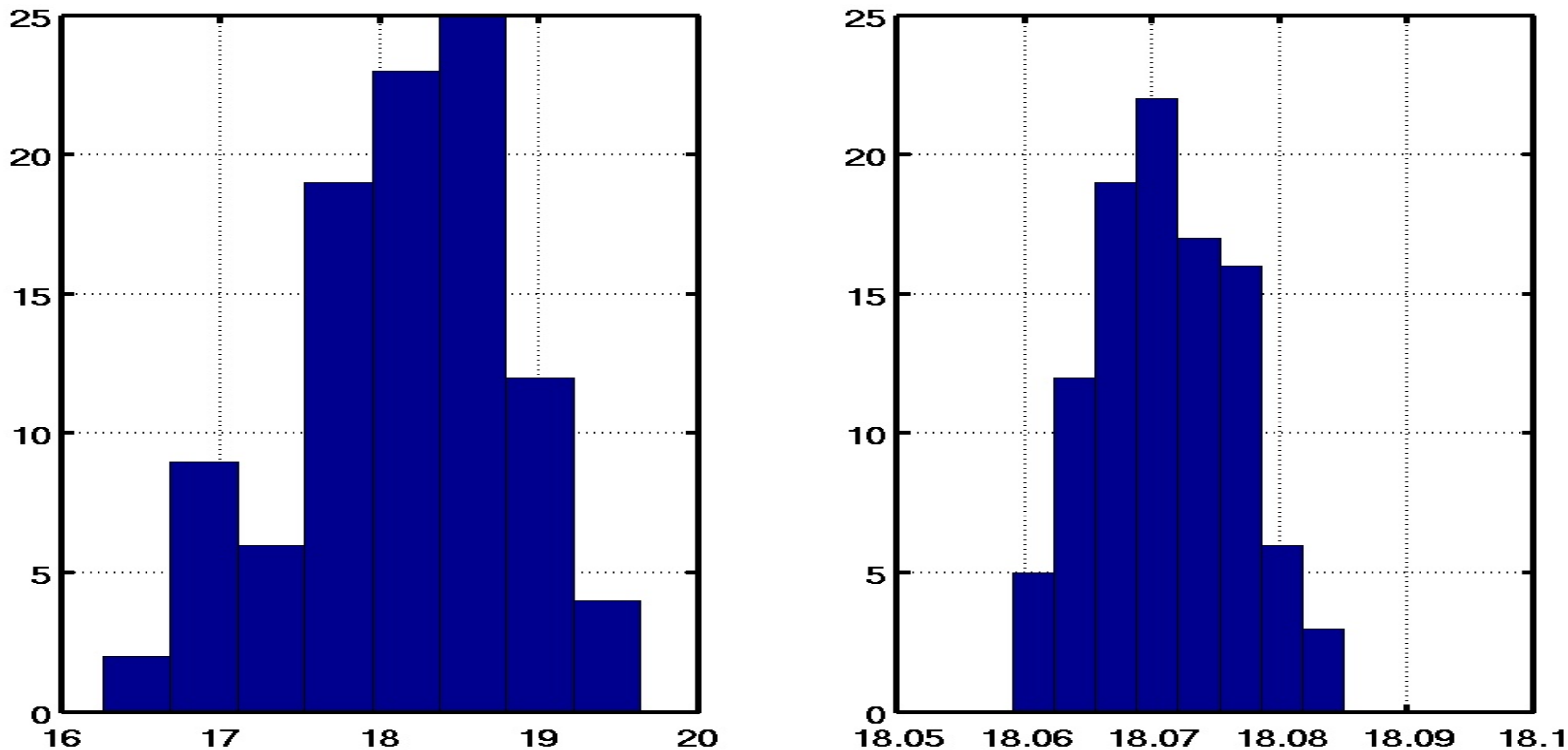


100 Monte Carlo runs (N=100)   100 Monte Carlo runs (N=1000)

Does Sampling Method Matter?

北京師範大學
Beijing Normal University
GCESS
全球变化与地球系统科学研究院
College of Global Change and Earth System Science

Distribution of the sample mean



100 Monte Carlo runs (N=1000)

100 Latin hypercube runs (N=100)

Parameter Screening Methods

Parameter Screening Methods

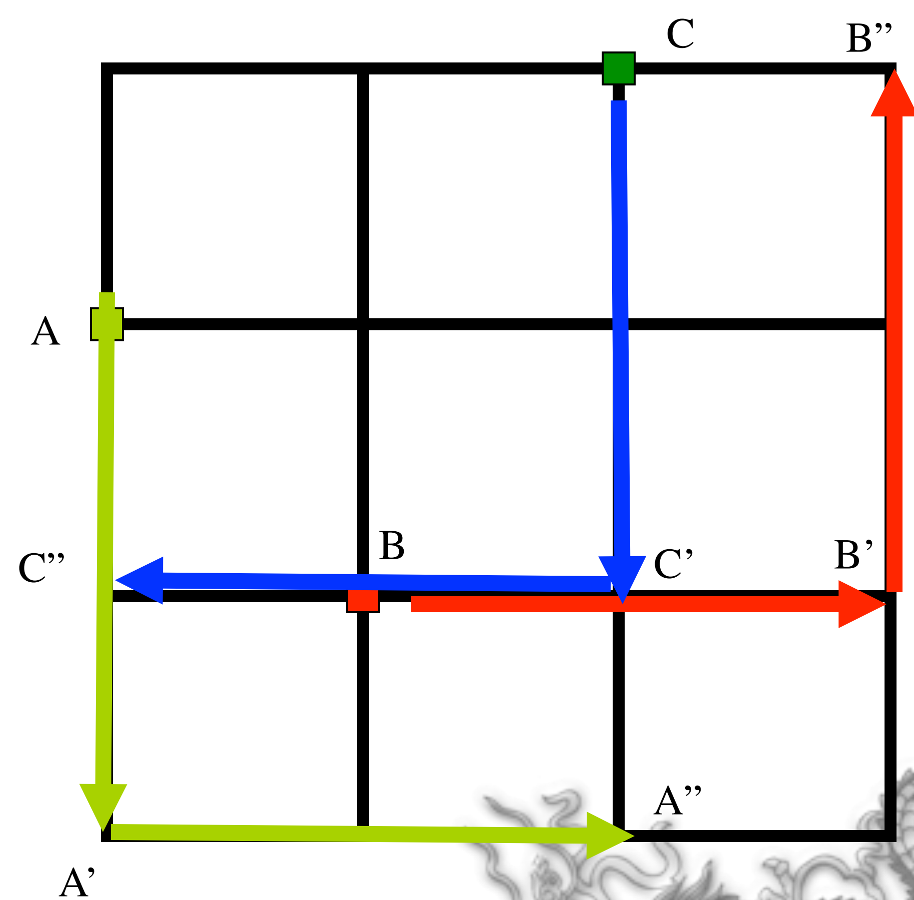
- Parameter screening, based on various sensitive metrics, is designed to separate important model parameters from unimportant ones:
  - The number of model parameters is very large
  - The model is computationally expensive and one can only afford run the model for a limited number of times
  - They usually provide only qualitative sensitivity measurements

Example: The Morris Screening Method

1. Start at a random point (A)

2. Create the next point by perturbing one input (A’)

3. Create the next point by perturbing another input (A’’)
   - Repeat step 1-3 r times (B,C..)
   - Form r gradients for each input and compute modified means and standard deviations
   - Plot mean vs standard dev. for each input → screening diagram



How Does the Morris Screening Method Work?

Gradient of response w.r.t the j-th input

$$z_j = \frac{y\left(x_1, x_2, \quad , x_j + \Delta x_j, \quad x_m\right) - y\left(x_1, x_2, \quad , x_j, \quad x_m\right)}{\Delta x_j}$$

Vector of gradients: with m input parameters

$$Z_r = \left(z_1, z_2, \quad , z_m\right)$$

Collection of gradient vectors (*R* paths or replications):

$$\Omega = \left\{Z_1, Z_2, \quad , Z_R\right\}$$

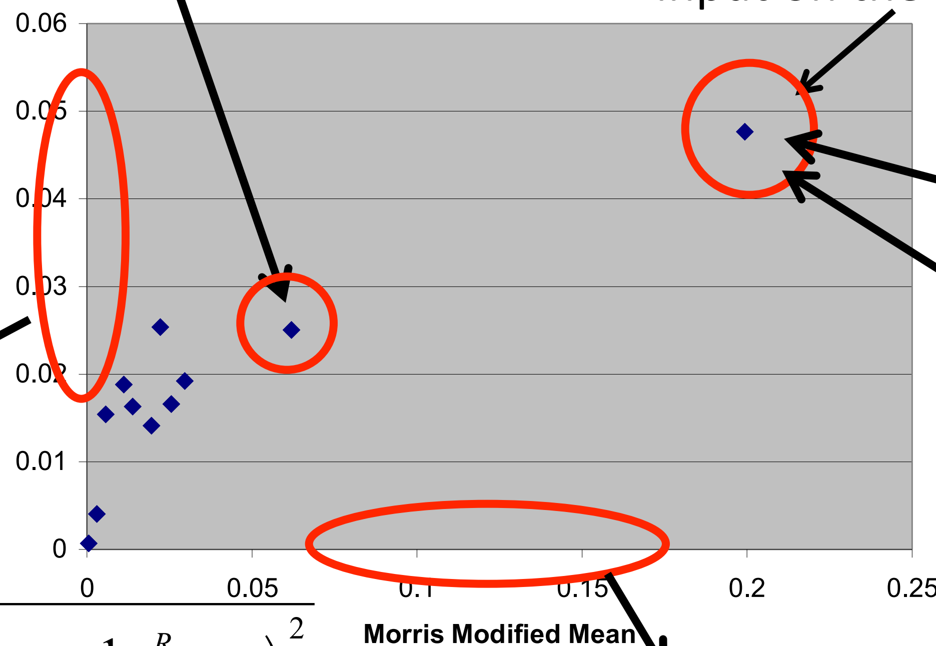
$$\overline{Z}_j = \frac{1}{R}\sum_{i=1}^{R}\left|Z_{ij}\right|$$

Study the statistics (mean and standard deviation) of Ω

Interpretation: Screening Diagram Is a Distillation of the Morris Screening Data

Each point refers to one particular input parameter

Each point represents the average "effect" of that particular input on the outputs



based on *R* points (R = # replicates)

$$\sigma_j = \sqrt{\frac{1}{R-1}\sum_{i=1}^{R}\left(Z_{ji} - \frac{1}{R}\sum_{i=1}^{R} Z_{ji}\right)^2}$$

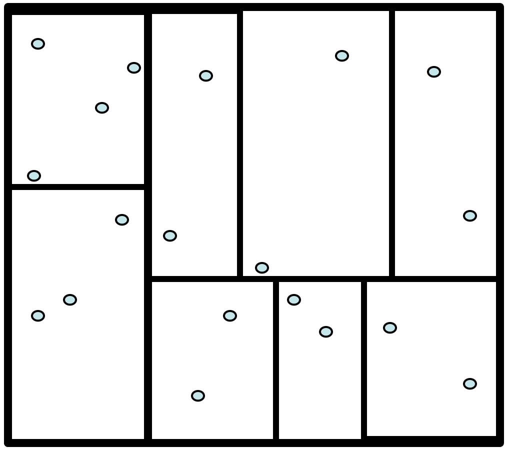
$$\bar{Z}_j = \frac{1}{R}\sum_{i=1}^{R}\left|Z_{ji}\right|$$

Note: mean is based on absolute value of the output

Large σ = non-linear relationship or inter-parameter interactions

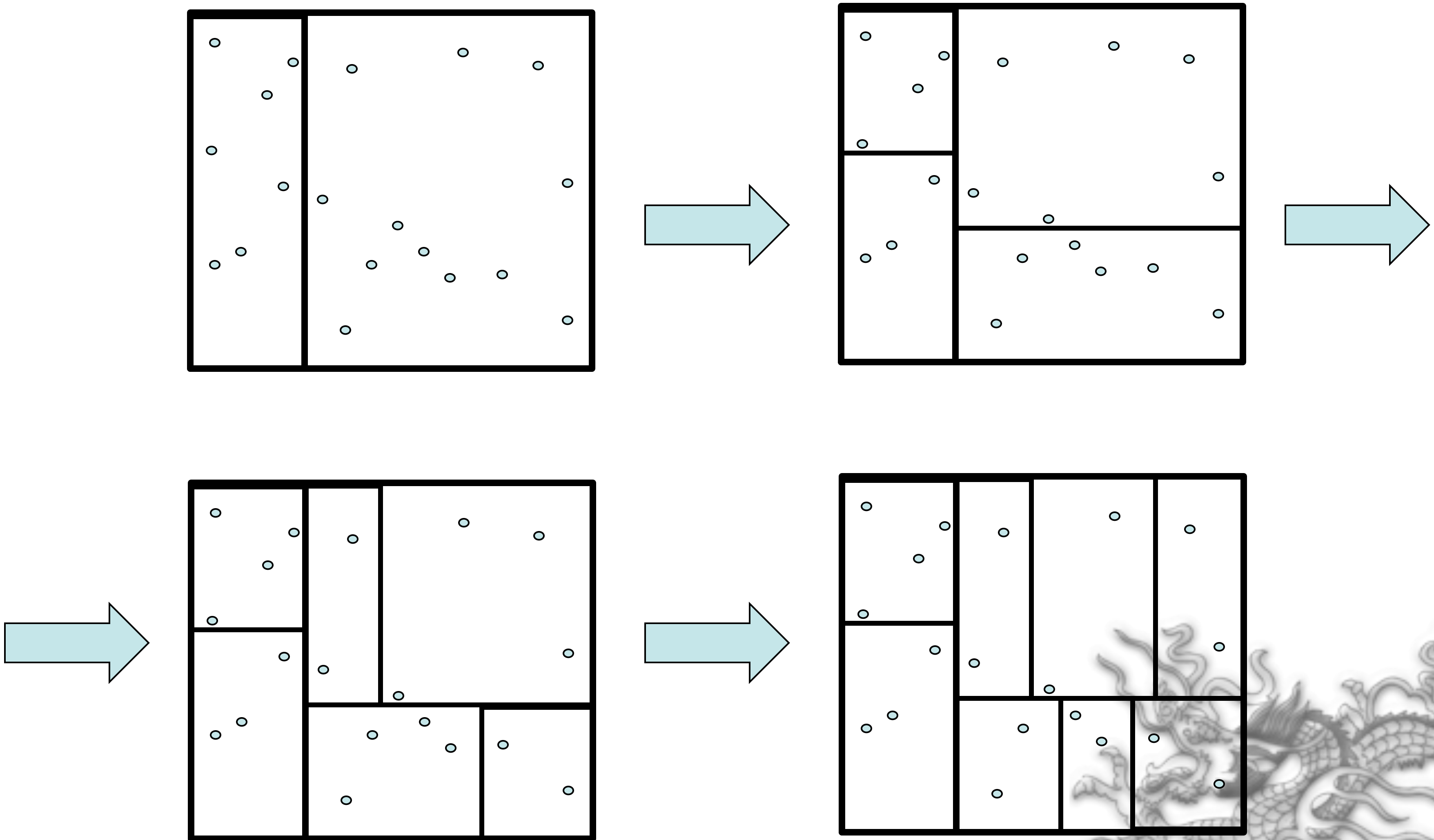
Large mean = "sensitive" parameter

Example: Tree-based Methods



- Sample design: random or quasi-random
- Based on creating a binary tree (unbalanced)
- Criteria for splitting: use impurity function
  - residual sum of squares
  - ratios of means and variances of sub-trees
- Splitting criterion: maximum decrease in impurity
- Stopping criteria:
  - minimum number of data points per terminal nodes
  - residual sum of squares falls below a threshold
- Sum-of-trees
  - use 100 bootstrapped samples and average (*)
  - use boosting and average
- Ranking criterion: (information metric)
  - total number of splittings (with scaling at each level) for each input
  - use standard deviations of the number of splittings as error bars

Tree-based Methods (cont)



More Screening Methods

- Delta Test (DT)
- Plackett-Burman (screening design for linear problems)
- Box-Behnken (3 level, fit quadratic)
- Gaussian Process Regression (GPR)
- Multivariate Adaptive Regression Splines (MARS)
- …

Quantitative Global SA Methods

- Sobol' total variance decomposition
- Mckey's main effect analysis
- Fourier Amplitude Sensitivity Test (FAST)
- Saltelli's modified total variance decomposition analysis

The Foundation of Variance Decomposition Is the Sobol' Property

- Any function can be decomposed into terms of increasing dimensionality, i.e. (such that the mean of each term is 0.)

$$F(x_{1,} \ldots, x_k) = \sum_{i=1}^{k} F_i(x_i) + \sum_{i=1}^{k} \sum_{j>i}^{k} F_{ij}(x_i, x_j) + \ldots + F_{1\ldots k}(x_1, \ldots, x_k)$$

- Then, the total variance is the sum of the variances of the individual terms.

$$V = \sum_{i=1}^{k} V_i + \sum_{i=1}^{k} \sum_{j>i}^{k} V_{ij} + \ldots + V_{1\ldots k}$$

- This holds true only for functions with uncorrelated inputs (the joint probability distribution function is 0)

We Need to Define a Few Sensitivity Measures

- Sensitivity index for input I (main effect or 1st order)

$$S_i = \frac{V_i}{V}$$

- Sensitivity index for input i and j (second order)

$$S_{ij} = \frac{V_{ij}}{V}$$

- Total sensitivity index for input i

$$S_{T_i} = \sum \text{all } V\text{'s involving i}$$

Another Useful Property From Statistics

- Variance decomposition based on conditioning input i

$$V = V[E(Y \mid X_i)] + E[V(Y \mid X_i)]$$

Variance of conditional expectation (conditioned on input i)

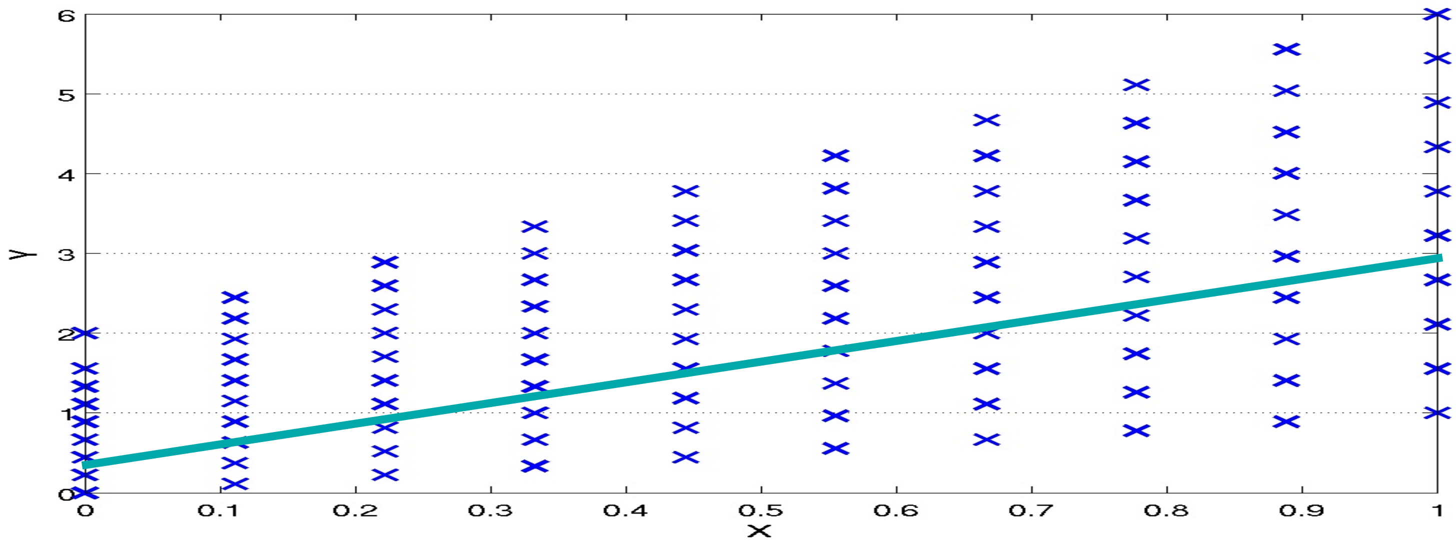
Remaining variability due to other inputs

- Sensitivity index for input i

$$S_i = \frac{V_i}{V} = \frac{V[E(Y \mid X_i)]}{V}$$

A Pictorial View of Variance Decomposition

- Given a scatter plot of output with respect to input i



$$V[E(Y \mid X_i)]$$

$$E[V(Y \mid X)]$$

Variance of the means (the red line) The variance of the trend shows the importance of X.

Each column shows the distribution of Y given a fixed X. Calculate the variances and take the mean of all X's

Similarly, We Can Derive Interaction and Total Sensitivity Indices

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science

- Interaction study (need different sampling methods)
  - use replicated orthogonal array design

$$V = V[E(Y \mid X_i, X_j)] + E[V(Y \mid X_i, X_j)]$$

- Total sensitivity indices
  - with correlated inputs, these are better measures
  - can use Fourier Amplitude Sampling Test (FAST) design

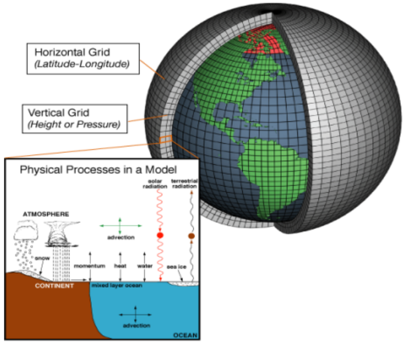
$$V = V[E(Y \mid X_{-i})] + E[V(Y \mid X_{-i})]$$

$$S_{T_i} = E[V(Y \mid X_{-i})] / V(Y)$$

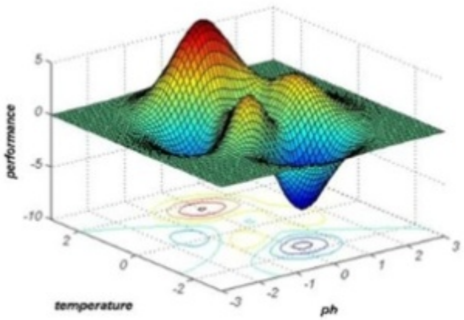
Surrogate Modeling Methods

What Is A Surrogate Model?

- The real world
  - A system
  - stimulus / response



- The dynamical simulation model
  - Abstraction of the real world
  - Based on physical processes, high computational complexity



- The surrogate model
  - A model of the model
  - Response surface, meta-model, emulator
  - Based on statistical theory, low computational complexity

Surrogate Modeling Basics

- Response surfaces are representations of the model output everywhere the parameter space

$$Y = F(X) \approx \hat{F}(X) \text{ in } \Omega$$

- Other names
  - response surface method
  - (stochastic or statistical) emulator
  - meta-model

- Basic ingredients of a response surface analysis
  - a sample (input-output pairs, space-filling)
  - a response surface fitting method

Surrogate Models Are Very Useful for the Uncertainty Quantification of Multi-physics Models

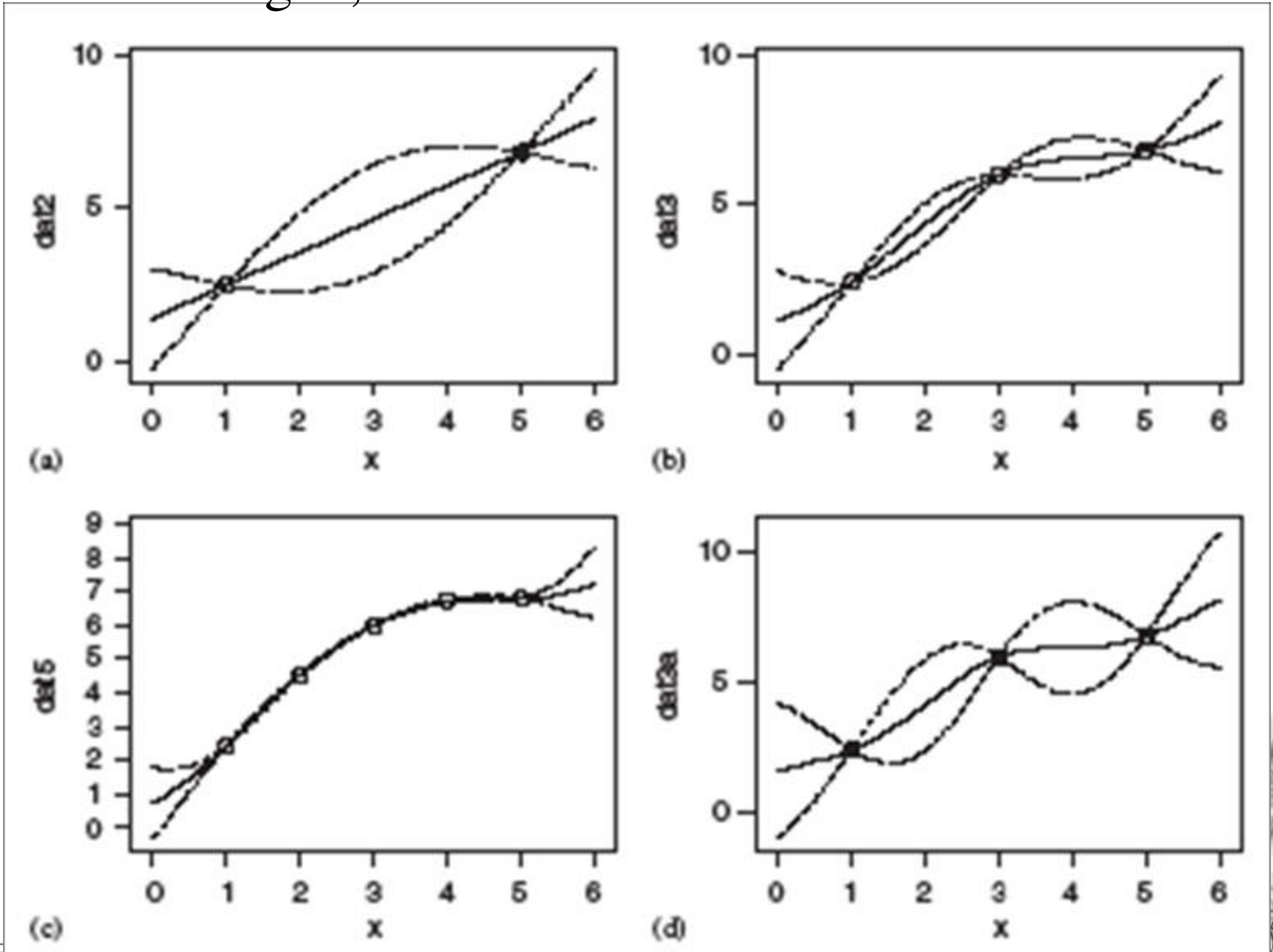
- Multi-physics models are generally expensive to evaluate (many CPU hours)
- Robust uncertainty quantification (forward/inverse uncertainty assessment, sensitivity analysis) needs many sample points
- Idea: use sampling and assumptions about the function *f* to construct an approximate mapping
- Challenges
  - Parameter space large (>10)
  - Near-singularities/discontinuities/noise

Definition:

Evaluate $S=\{(X^i,Y^i),i=1,\dots N\ ,X_i\in R^m,Y^i\in R\}$
Find $f \subset F$ (hypothesis function space) such that
$V(S,f)$ (some error measure) is minimized.

Illustration of a Surrogate Model

From O'Hagan, 2006



How to Create Surrogate Models?

1. Choose a sampling method (LP-tau, Metis, LH, etc.)
2. Run the simulator with the sample
3. Use response surface check to see goodness of fit
   - examine training errors
   - examine cross validation errors
4. If errors are not acceptable, add more points
5. Create a FF IV design to sample some corners
   - to test the robustness against extrapolation
6. Use 'rstest' to examine extrapolation errors
7. If good, add FF design and create new response surface

Choice of Sampling Strategies

Definition:

Evaluate $S = \{(X^i, Y^i), i = 1, \ldots N\ , X_i \in R^m, Y^i \in R\}$
Find $f \subset F$ (hypothesis function space) such that
$V(S, f)$(some error measure) is minimized.

- If nothing is known about the mapping, use space-filling samples
  - e.g. factorial, quasi-MC, max-min LH, orthogonal arrays, Metis
  - together with refinement to achieve sufficient accuracy
  - can use active learning (adaptive sampling)
- When the function space is partially known, use special sampling
  - e.g. function is linear or has bounded k-th order derivatives
  - e.g. fractional factorial, sparse grids
- If there are any discontinuities, should sample more near the discontinuities (adaptive sampling)

Choice of Error Measures

Definition:

Evaluate $S=\{(X^i,Y^i),i=1,\ldots N\ ,X_i\in R^m,Y^i\in R\}$
Find $f\subset F$ (hypothesis function space) such that
$V(S,f)$ (some error measure) is minimized.

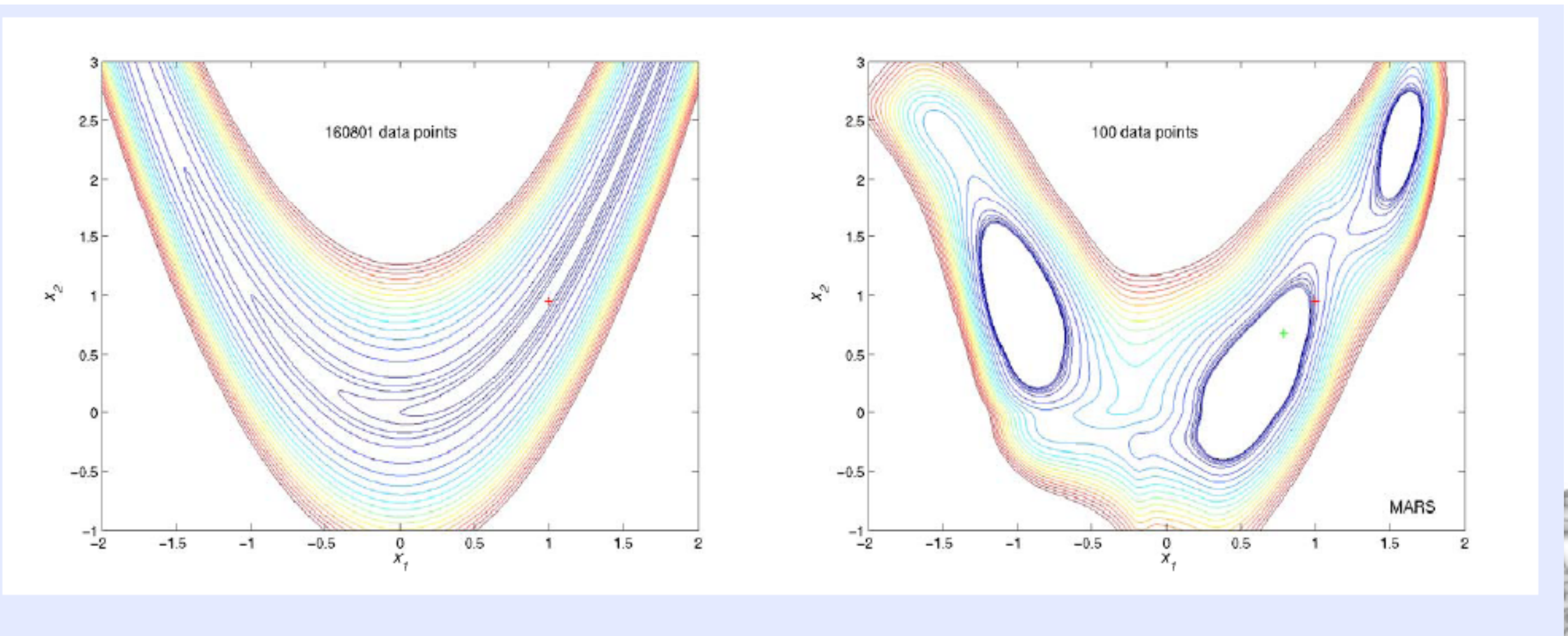
- R-square or adjusted R-squares (polynomial regression)
- Taylor expansion (truncation error)
- Convergence of the function mean (classical learning)
- Chi-square (training error, cannot account for generalization error)
- Holdout data set (training and test set)
- k-fold cross validation (check generalization error)
- Statistics on point-wise standard deviation for Gaussian process
- Extrapolation analysis: Gower distance

K-fold Cross Validation

- Given a sample of N points $S=\{(X^i,Y^i), i=1,\ldots N\ , X_i \in R^m, Y^i \in R\}$
- Divide the sample into k roughly same size groups
- For i = 1 to k
  - take out group i and use the rest to build a response surface
  - use the response surface to predict the outputs of group i
  - compute the sum of squares of the output discrepancies
- Add up all k sum of squares, divide by N and assess sufficiency
- Advantage: all N sample points are used in the response surfaces
- Provide some checking for extrapolation accuracy
- Exhaustive cross validation: using k = N, N/2, N/3, …
- Ideal error statistics: approximately Gaussian with zero mean and small standard deviation

Rosenbrock Function Example

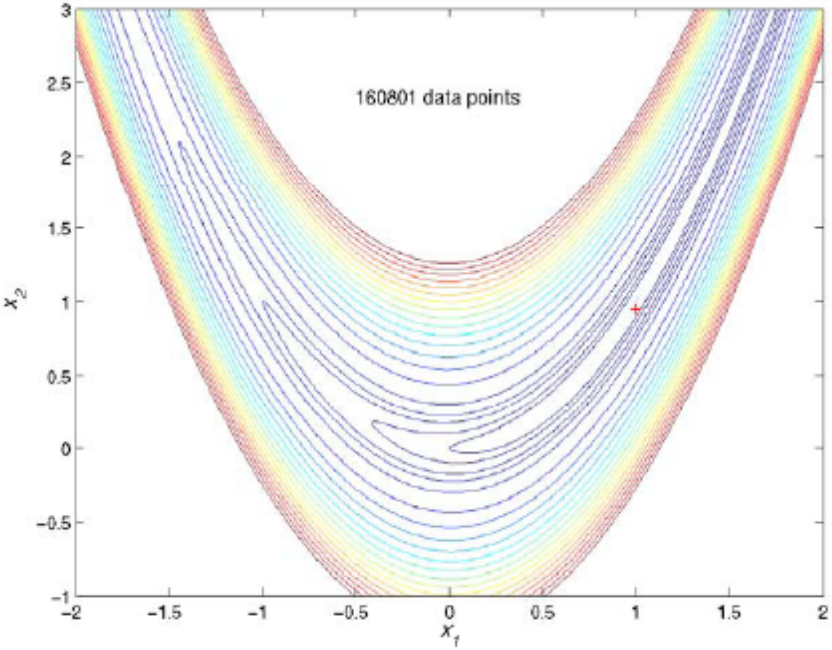
$$f(x_1, x_2) = 100\left(x_2 - x_1^2\right)^2 - \left(1 - x_1\right)^2$$

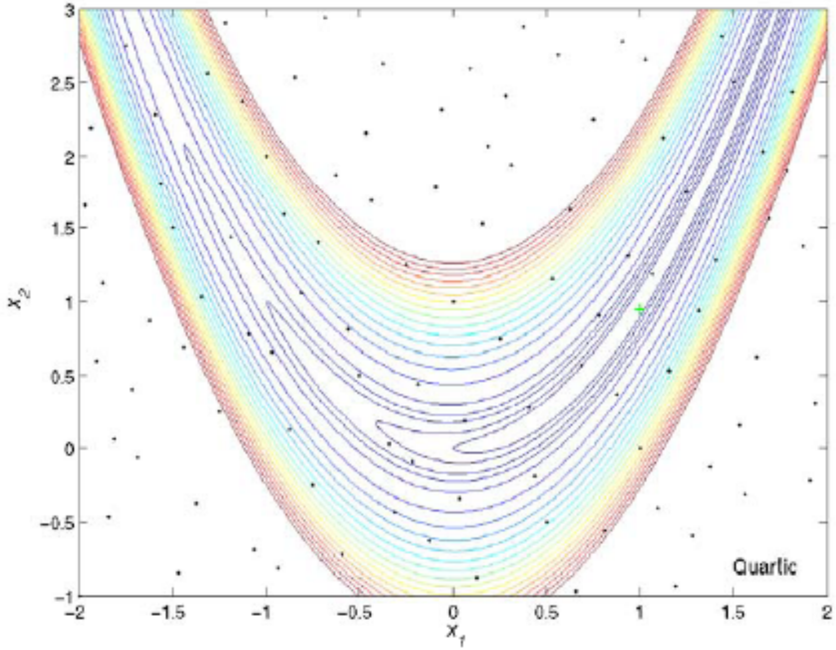


Monte Carlo (>100000 samples) MARS (100 samples)

Rosenbrock Function Example

$$f(x_1, x_2) = 100\left(x_2 - x_1^2\right)^2 - \left(1 - x_1\right)^2$$





Monte Carlo (>100000 samples)

Quartic (100 samples)

A Word About Surrogate Models

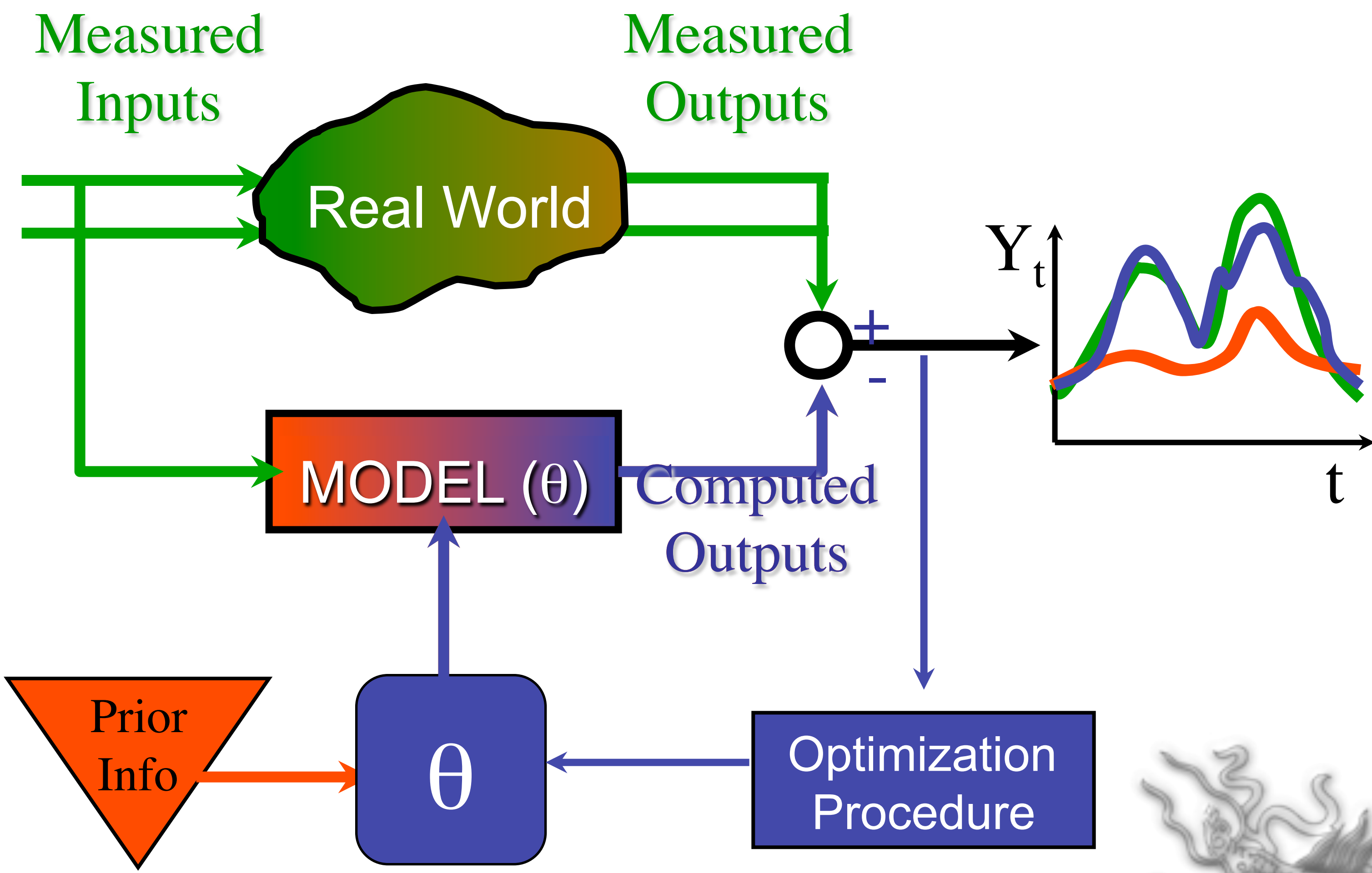
- Conclusions from the analysis are valid only if the surrogate model approximates the output response well (some smoothness assumptions)

- Need surrogate model validation
  - response surface design: adequate resolution
  - response surface design: true space filling
  - response surface design: avoid extrapolation
  - validation via training set and test set
  - cross validation (e.g. bootstrap, jackknife), k-fold CV
  - R-square in regression

Optimzation and Model Calibration

Model calibration

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science
GCESS



"Calibration: constraining the model to be consistent with observations"

Model calibration – 2 Types of Approaches

- Based on deterministic optimization
  - formulate an objective function (e.g. least-squares)
  - define independent variables and bounds
  - define any inequality constraints
  - run optimization algorithms
- Stochastic optimization (e.g. Bayesian)
  - given data and standard deviation (assume normal)
  - define a likelihood function
  - define independent variables and distributions
  - run Markov Chain Monte Carlo algorithm

* For efficiency reason, response surface is preferred.

Model Calibration – Deterministic

- Formulate an objective function (e.g. least-squares)
- Define independent variables and bounds

$$G(X) = \min_{X} \sum_{i=1}^{n} [(Y_i^s(X) - Y_i^e)/\sigma_i]^2$$

$$\text{subject to} \quad l_i \le X_i \le u_i$$

- Run optimization algorithm to identify candidates
- If outputs have uncertainties, perform sensitivity analysis in the neighborhood of the candidates

Bayesian Calibration

- Formulate an objective function (e.g. least-squares)

$$Y(x) = M(x,\theta) + \delta(x) + \varepsilon$$

- Formulate likelihood function:

$$\pi(\theta \mid \{Y\}) \propto P(\theta)L(\{Y\} \mid \theta)$$

- Run MCMC to get the posteriors of parameter set θ (not just the optimum):
- Different MCMC algorithms:
  - Metropolis
  - Gibbs sampler
- You may first create a response surface for the Y's

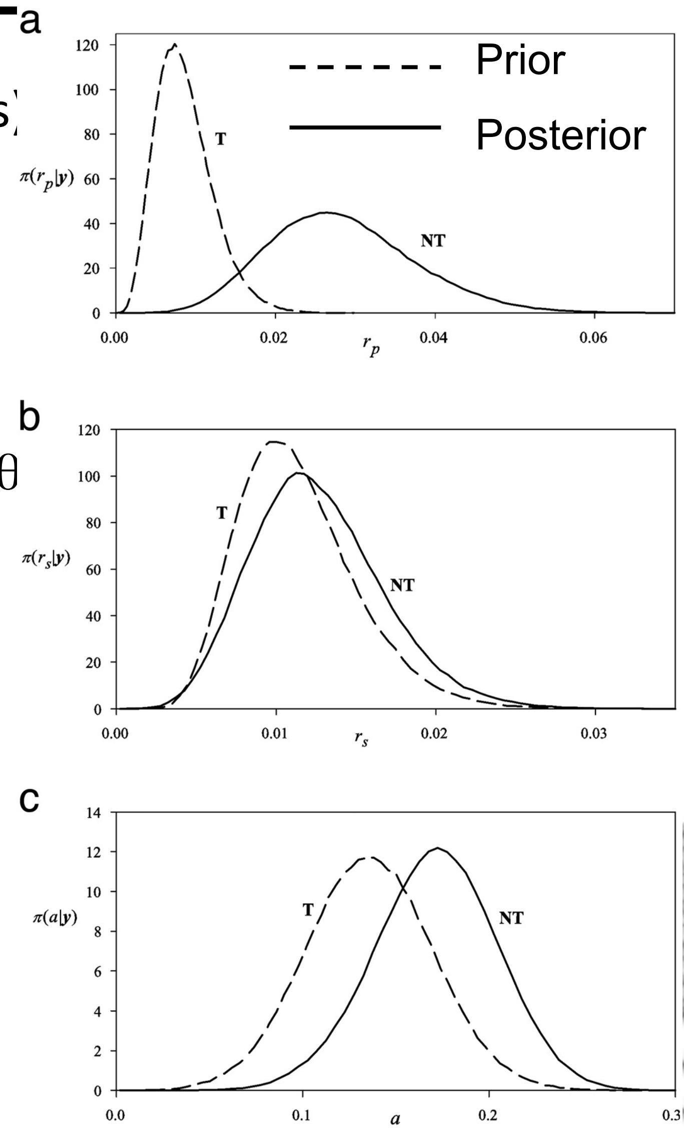
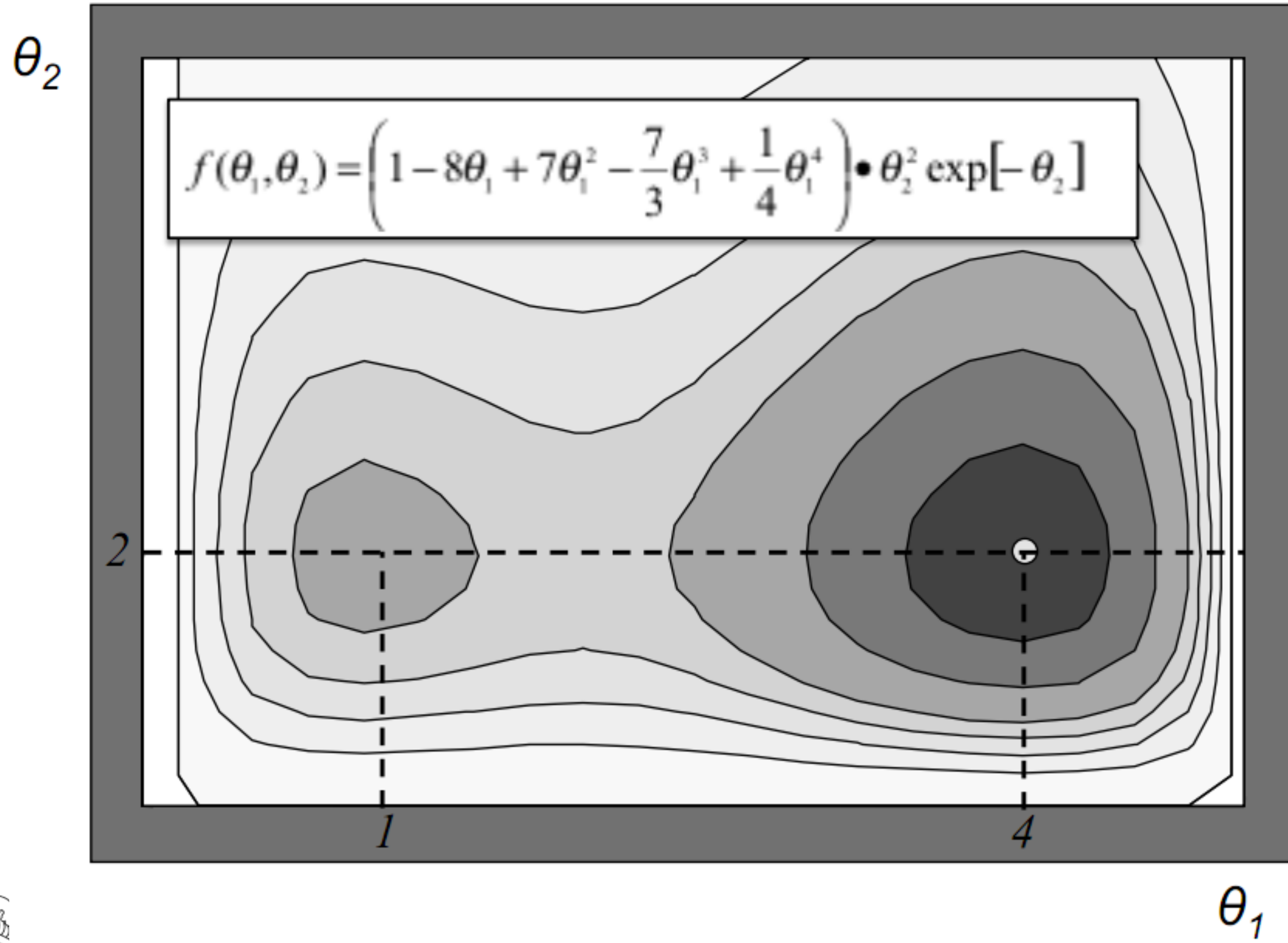


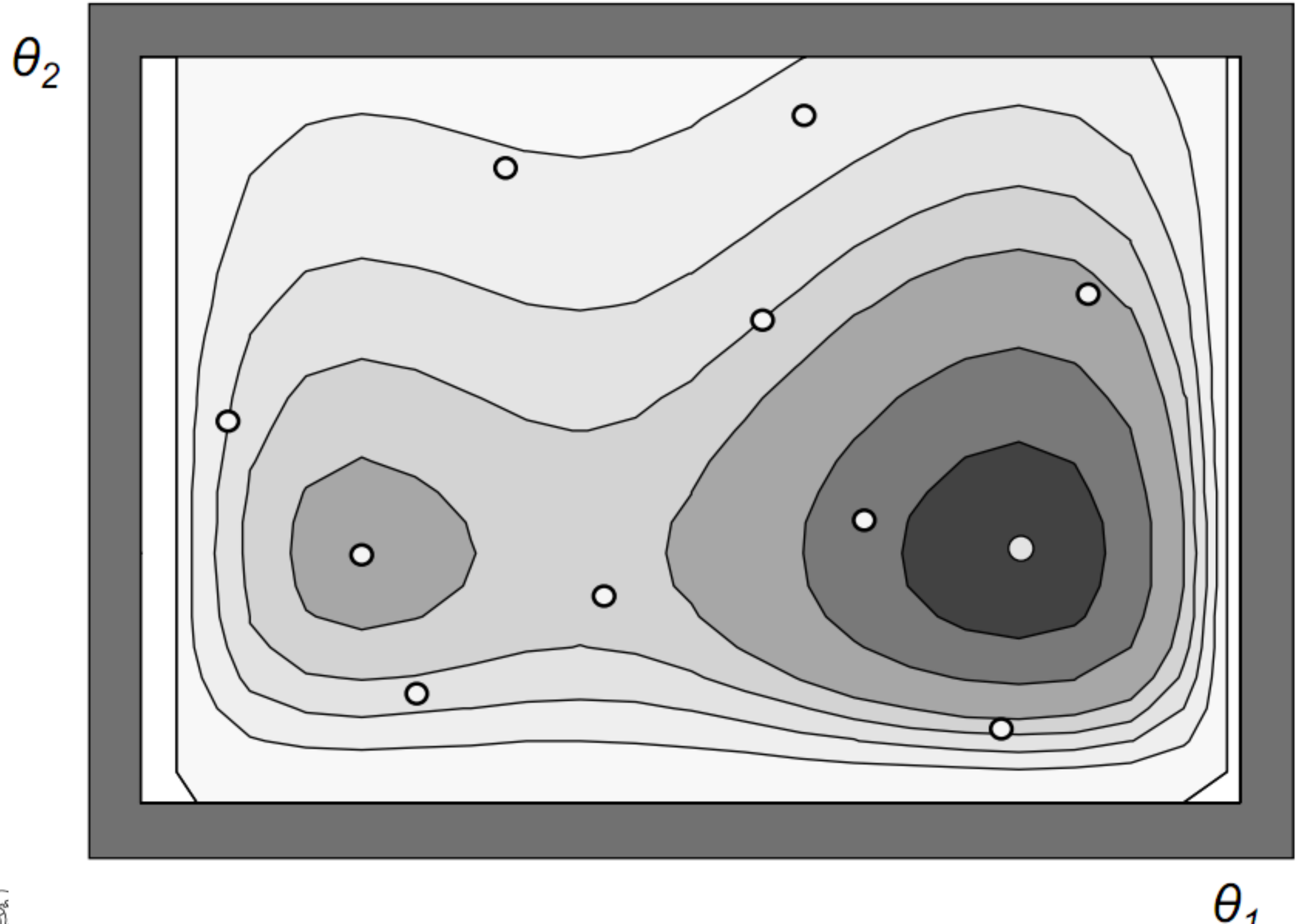
Illustration of Shuffled Complex Evolution Method

(based on Duan, Gupta, and Sorooshian, 1992, WRR)

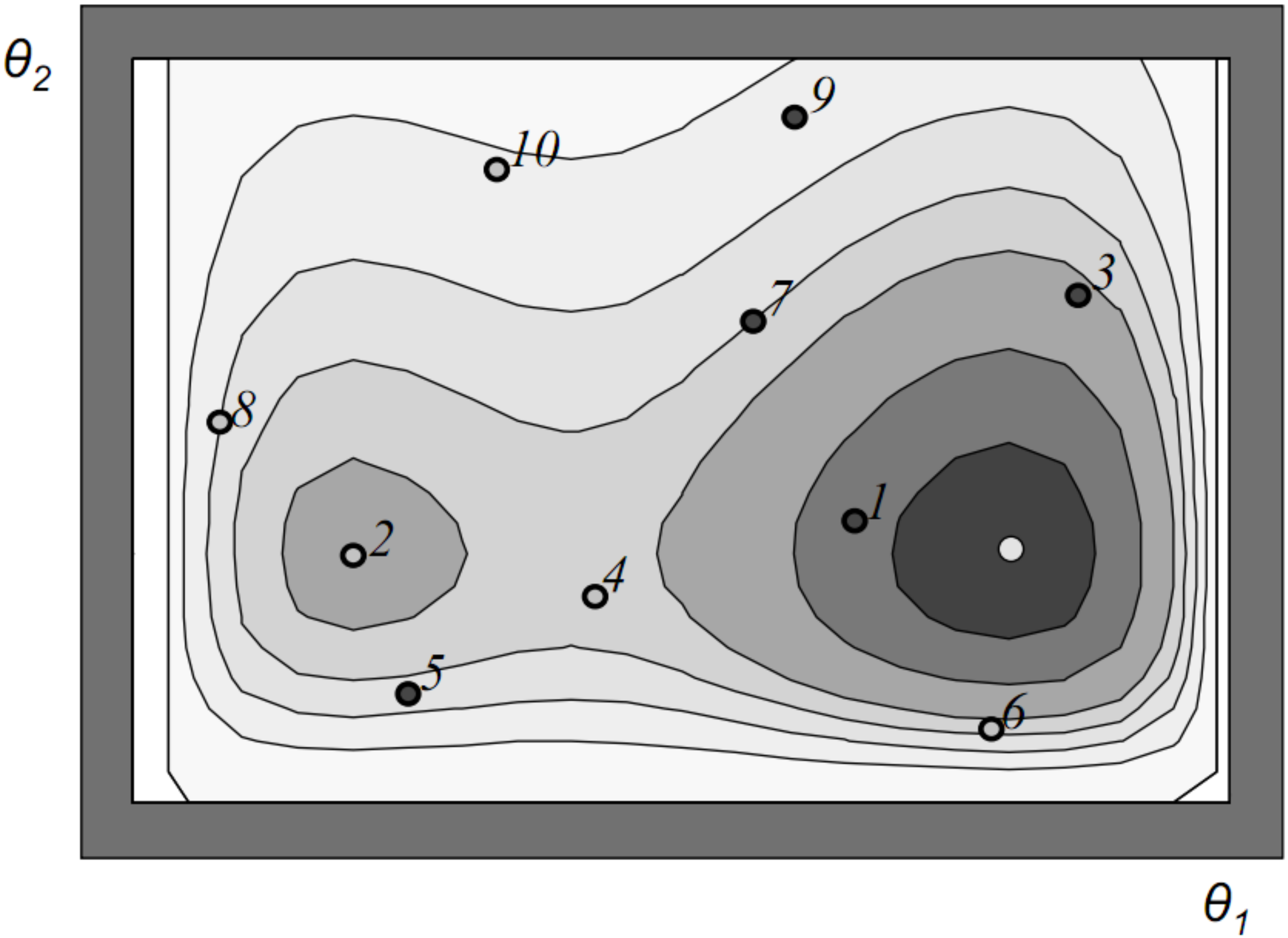
2-D problem (Hosaki Function)



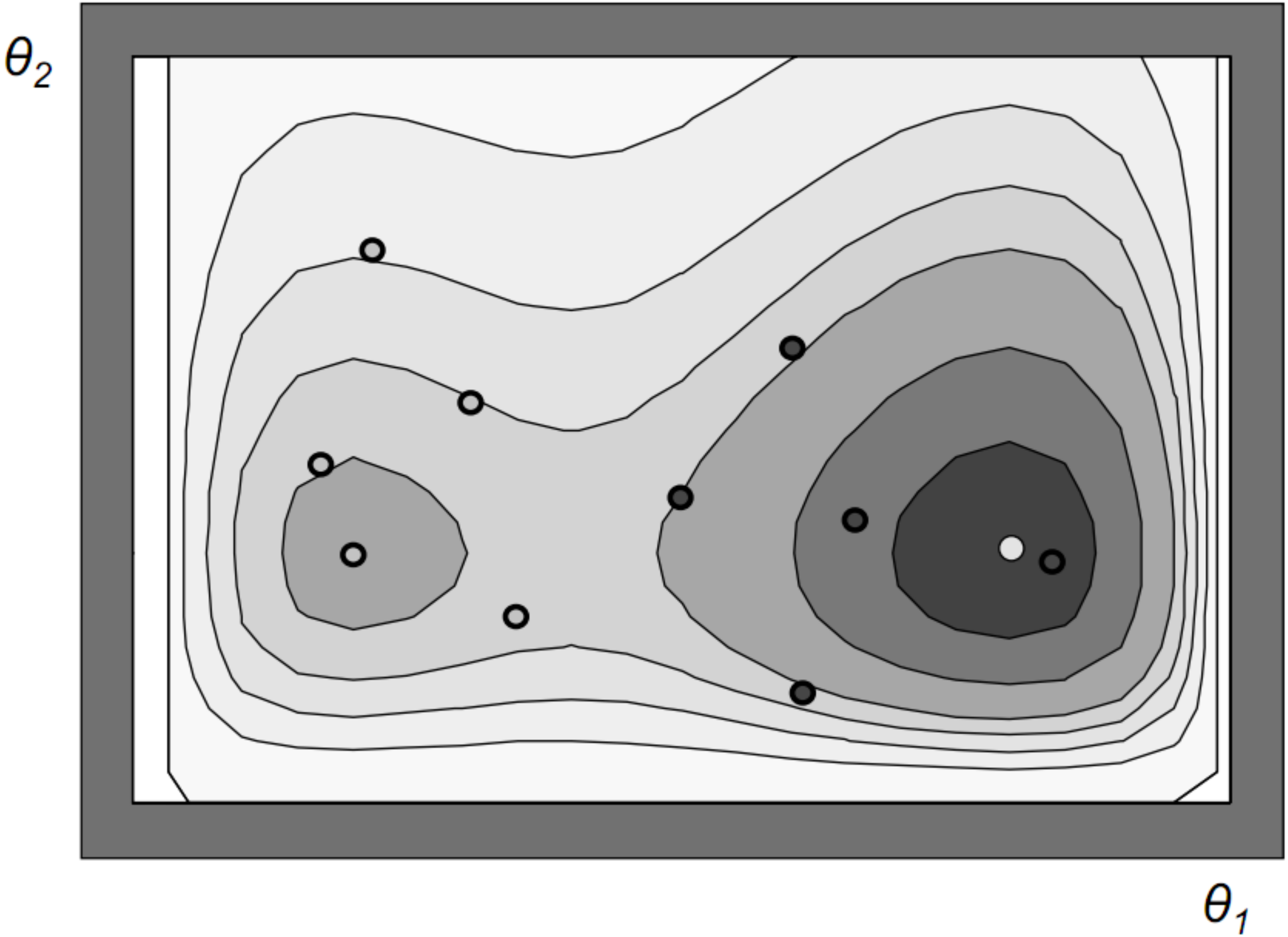
1. N = 2 so select P = 2 (or larger)
2. S = P(2N+1) = 10
3. Randomly generate initial population & evaluate function



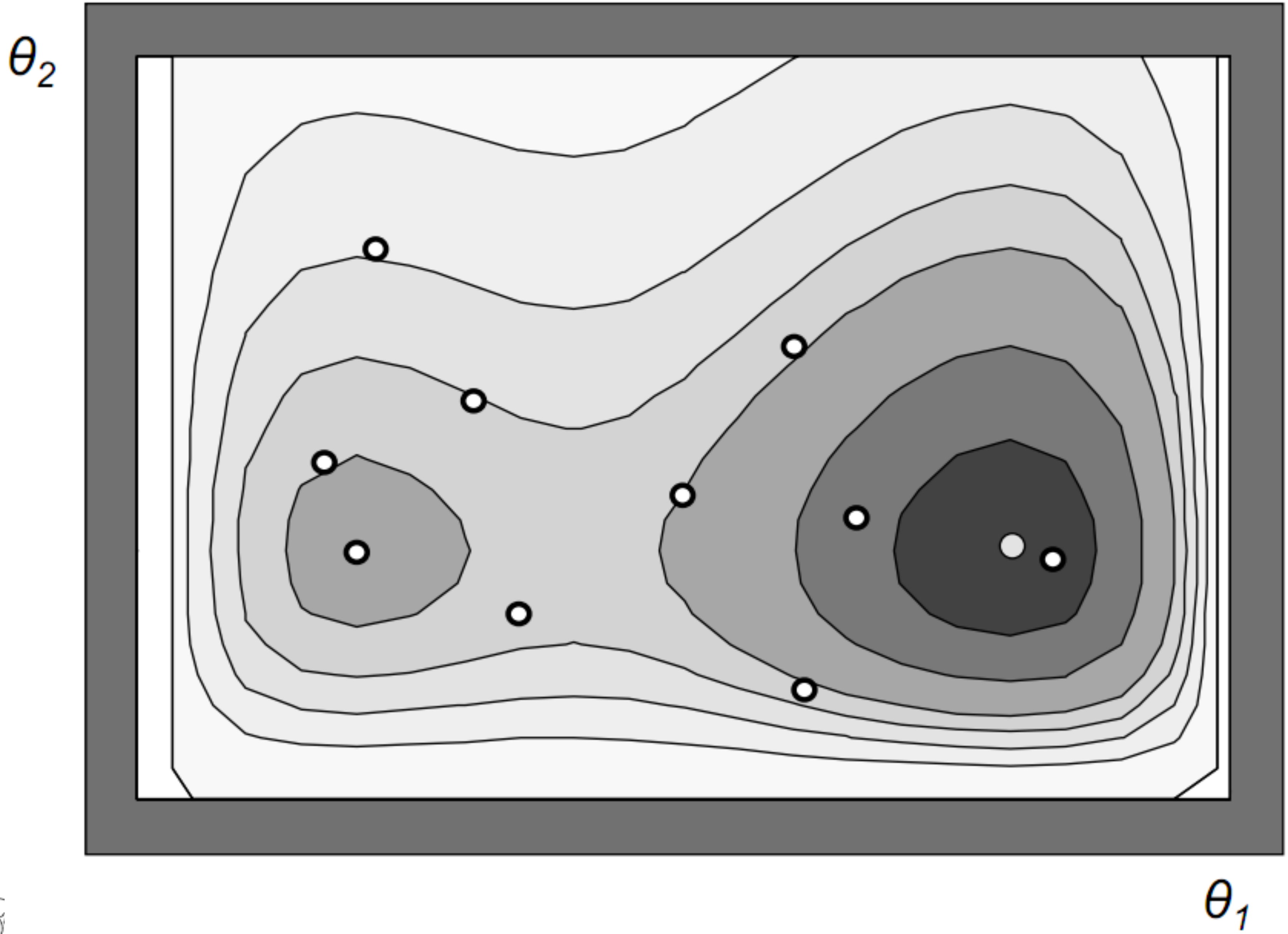
4. Sort population S according to increasing function value F(.) and distribute into complexes (like dealing cards)



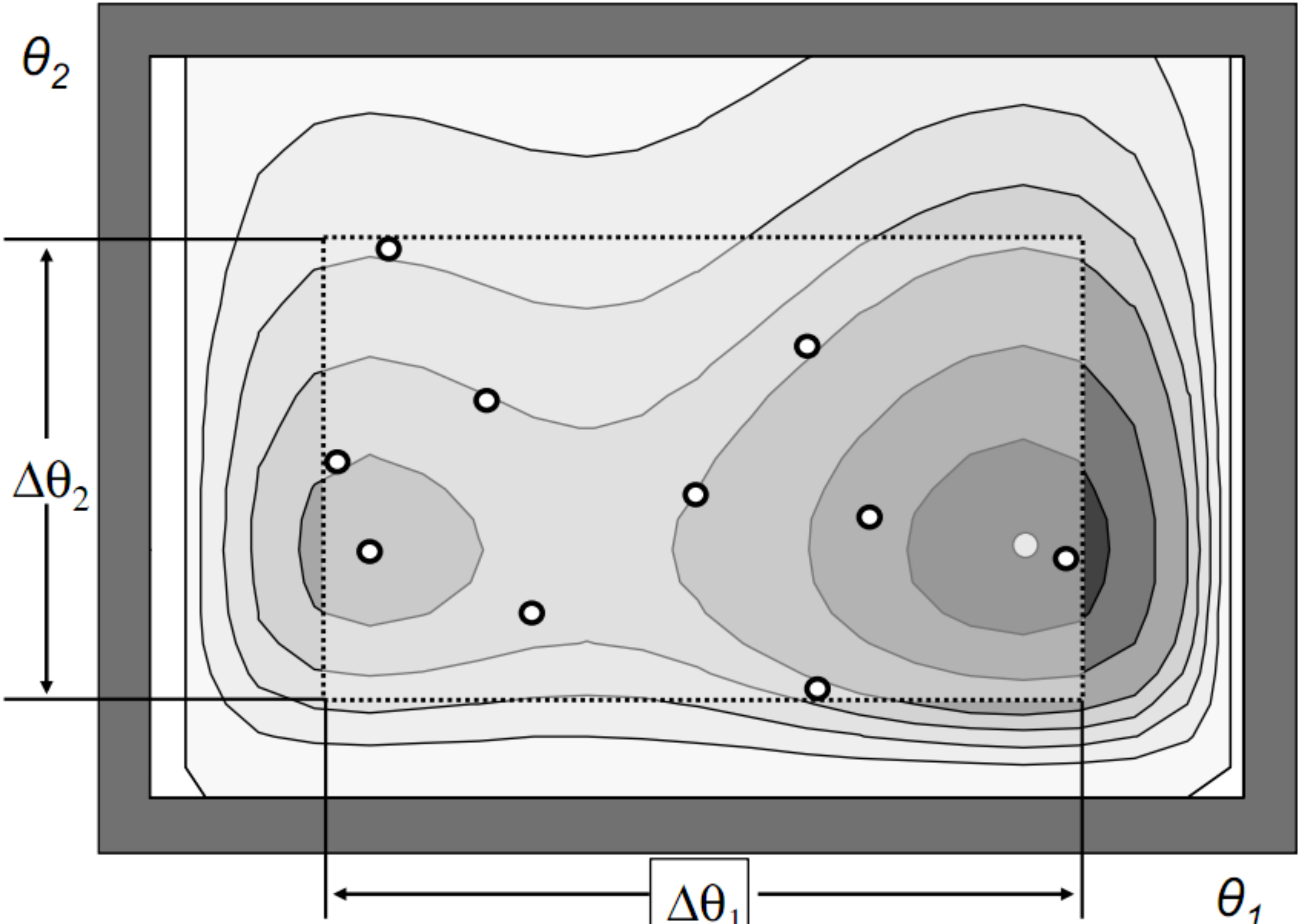
5. End of evolution Step (loop #1)



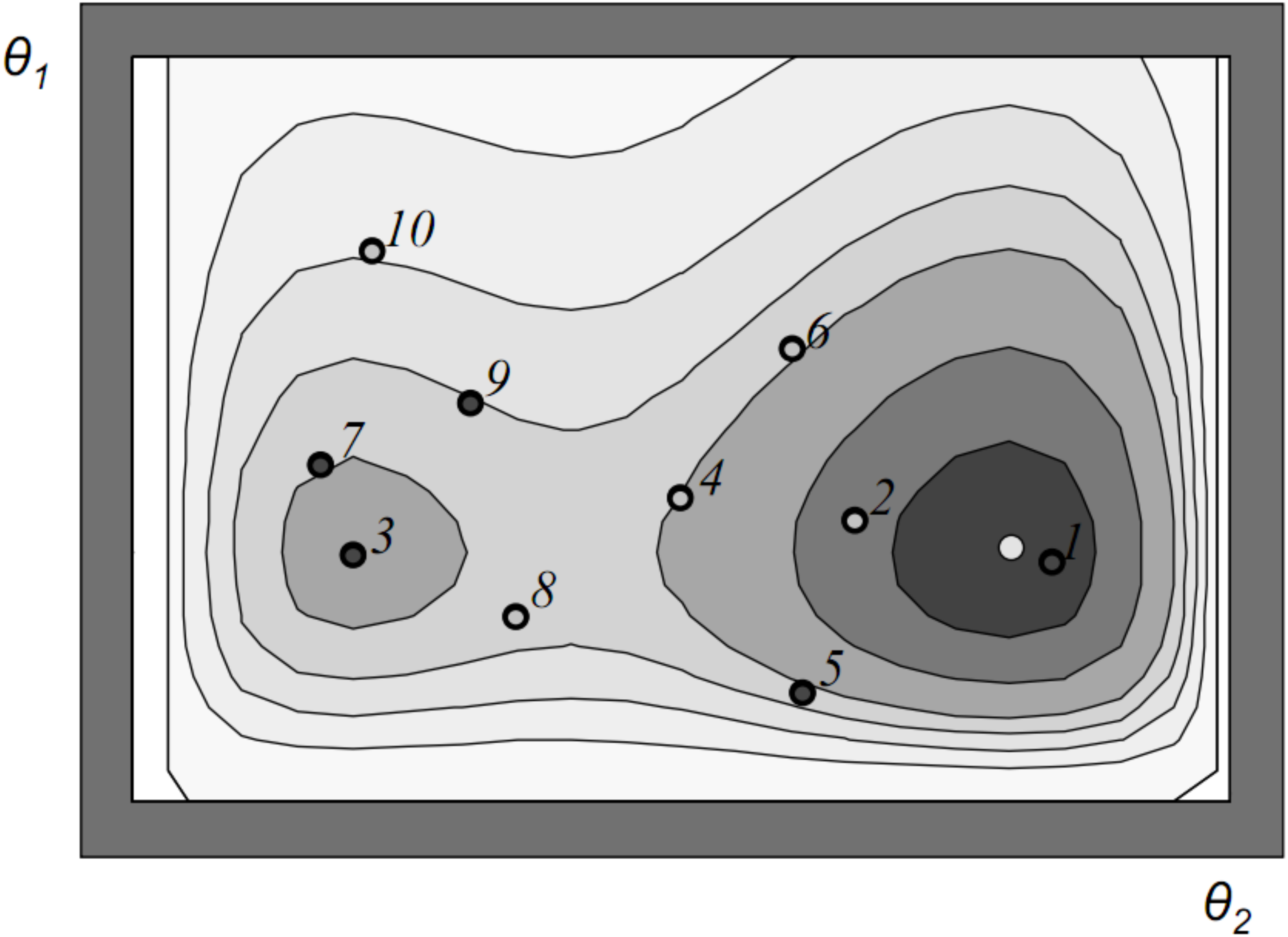
6. Shuffle the complexes together



7. Evaluate convergence criteria

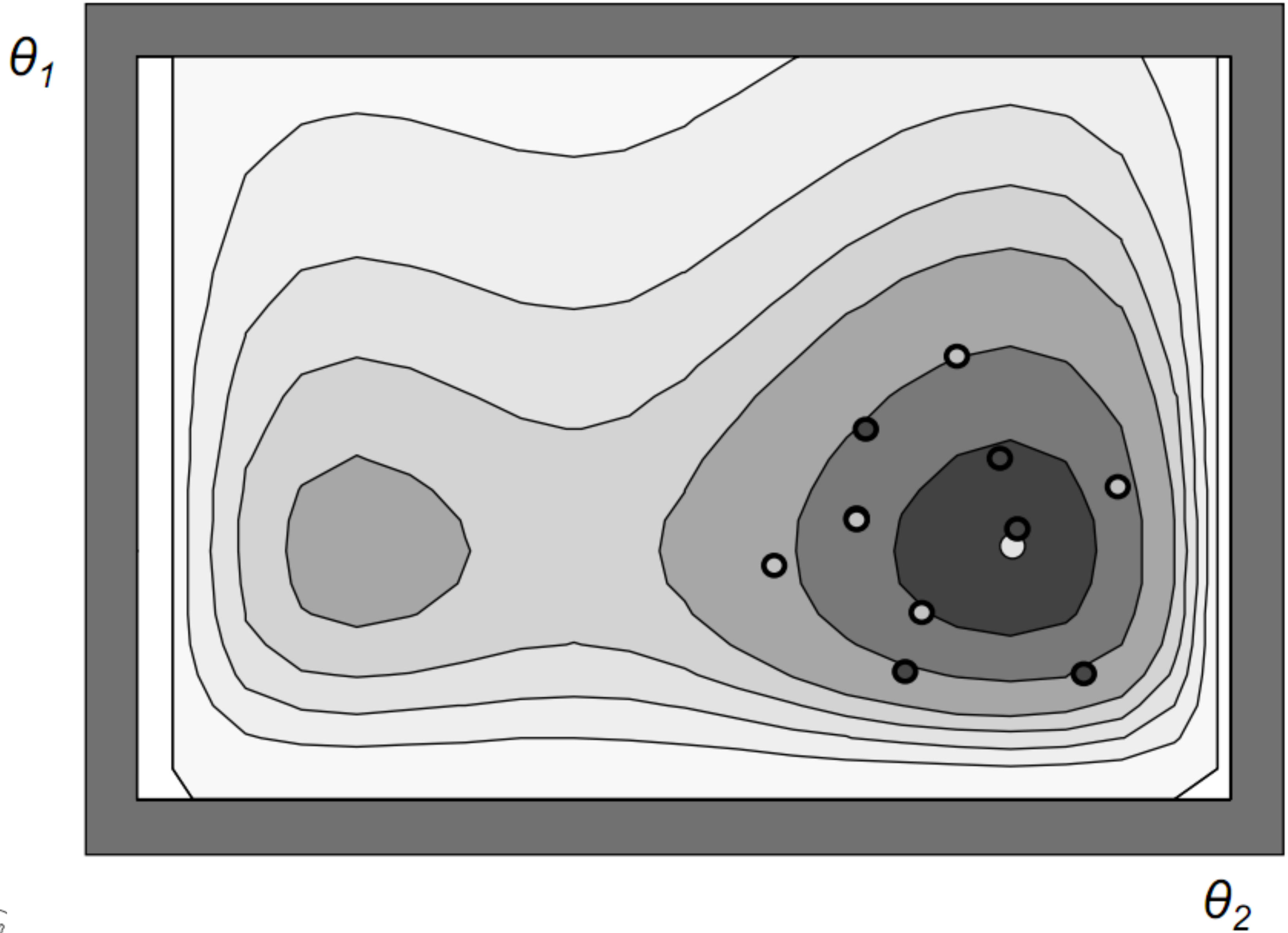


8. REPEAT – Redistribute points into complexes and continue

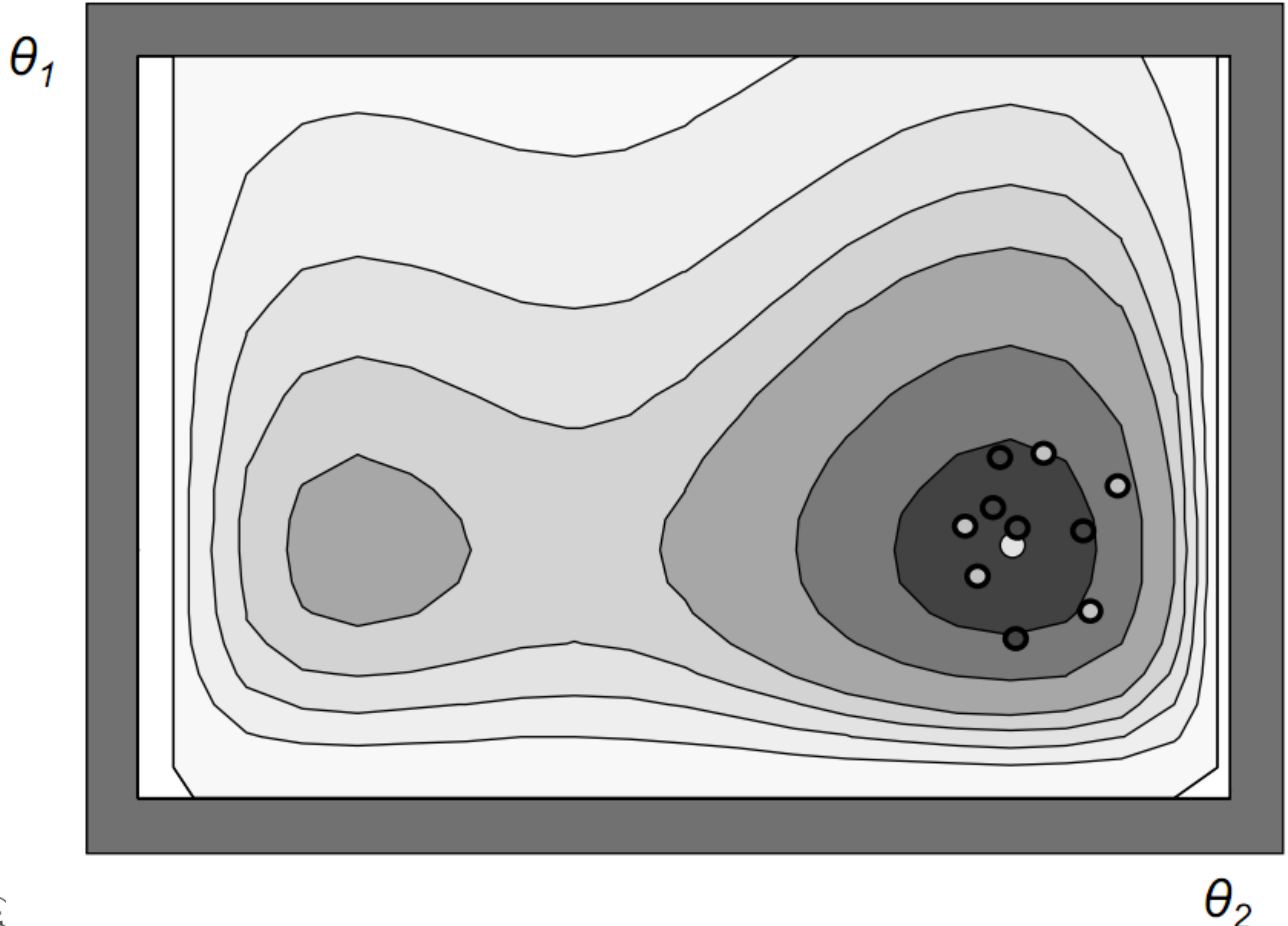


End of Evolution Step (Loop #2)

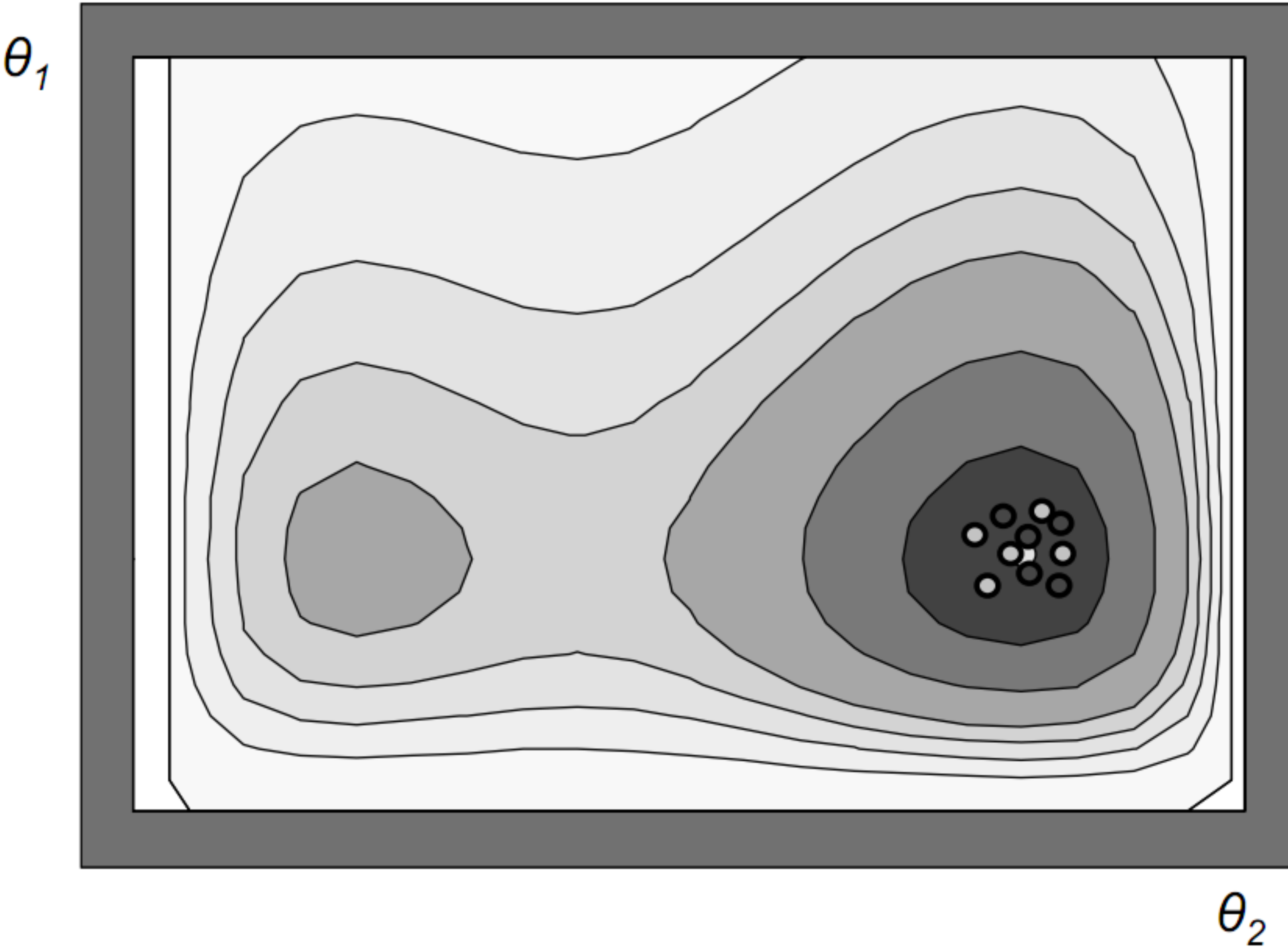
北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science



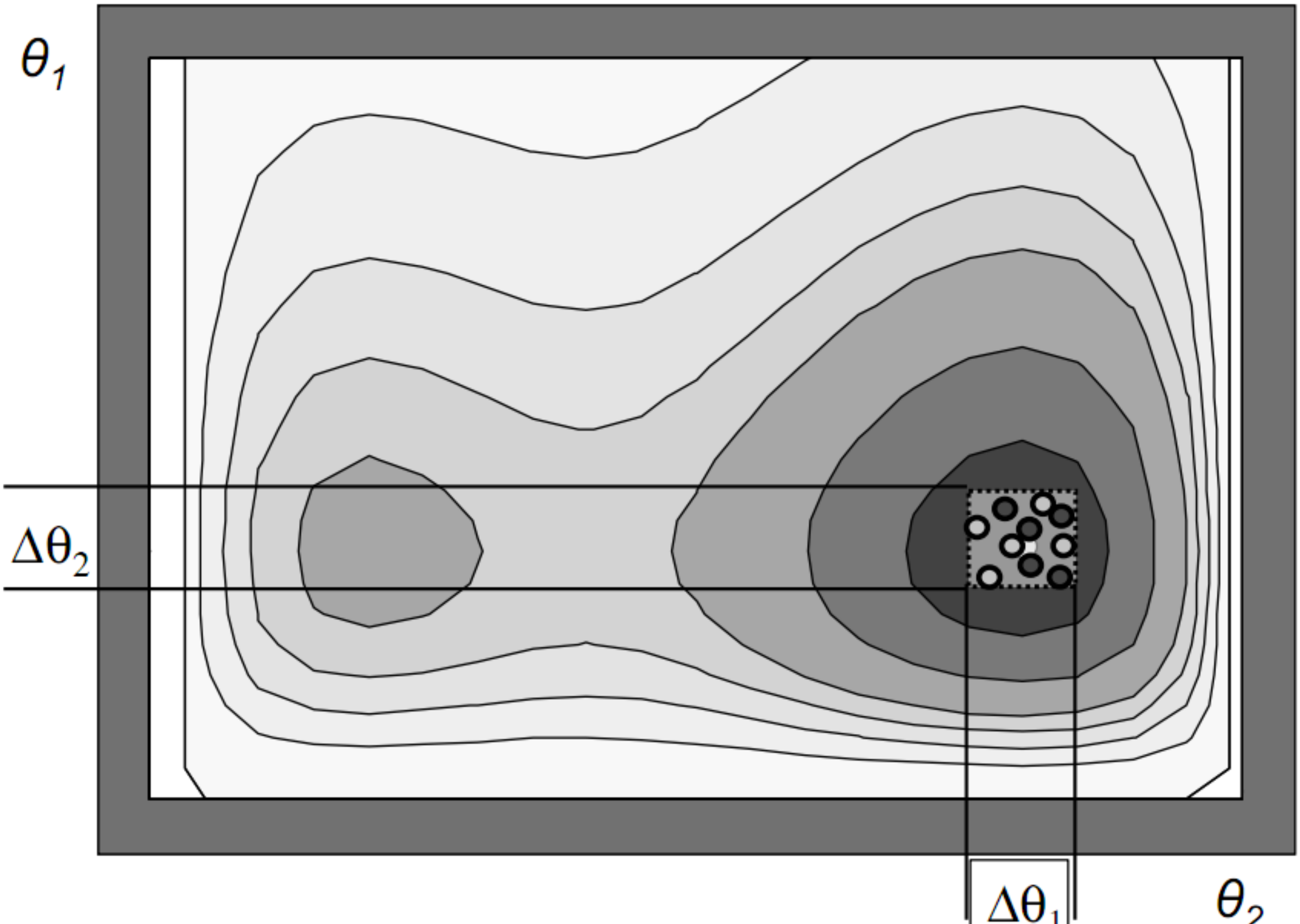
End of Evolution Step (Loop #3)



End of Evolution Step (Loop #4)



Termination – Parameter Convergence



ASMO: Adaptive Surrogate Modeling Based Optimization

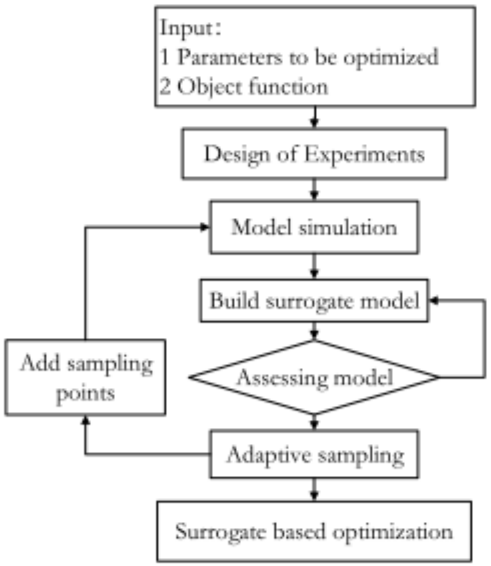


Figure 1. A schematic description of the ASMO scheme.

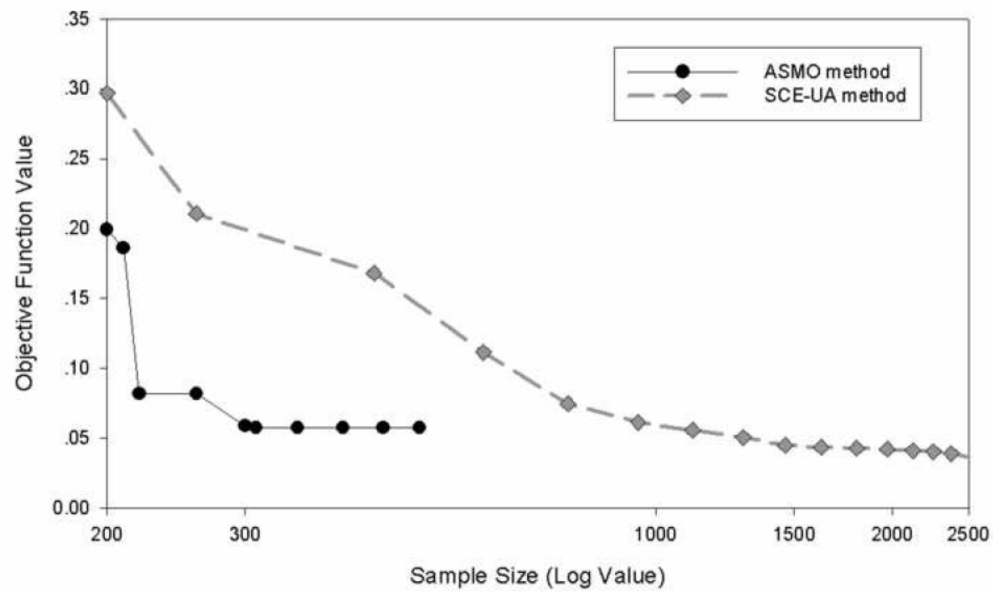
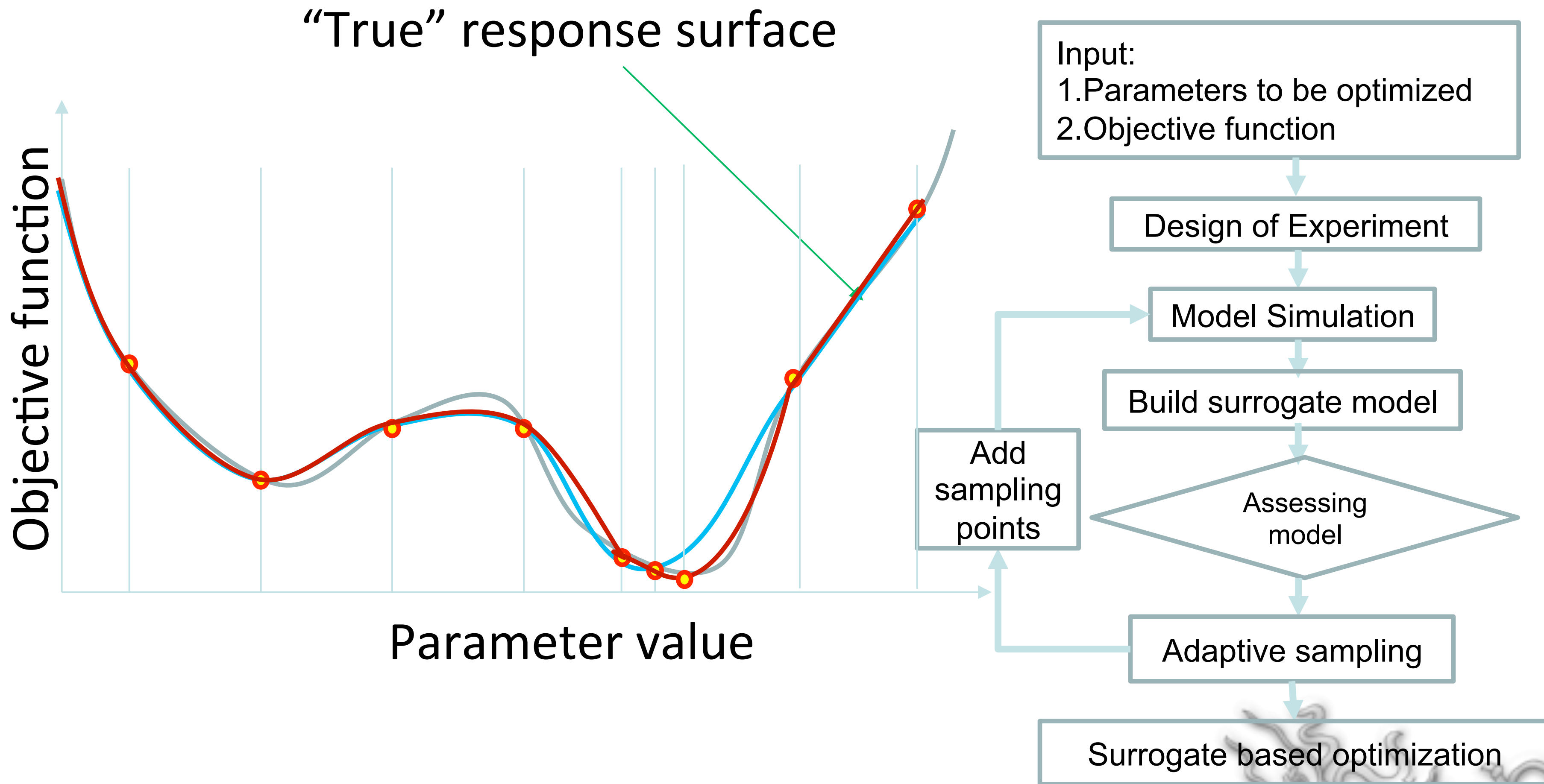


Figure 7. Optimized objective function values and sampling sizes of the SAC-SMA model by ASMO and SCE-UA methods.

ASMO: A wise way of doing optimization with surrogate model.

[Chen Wang et.al. 2014, EMS]

How Does ASMO Work?



[Chen Wang et.al. 2013, EMS]

Summary

- Concepts of UQ
- Sampling techniques
- Sensitivity analysis and parameter screening
- Surrogate model methods
- Model calibration and optimization methods

UQ Demonstration by Case Studies

Case Study 1:

PARAMETER SCREENING: A CASE STUDY WITH THE COMMON LAND MODEL

Common Land Model (CoLM)

[*Dai et al.*, 2003]

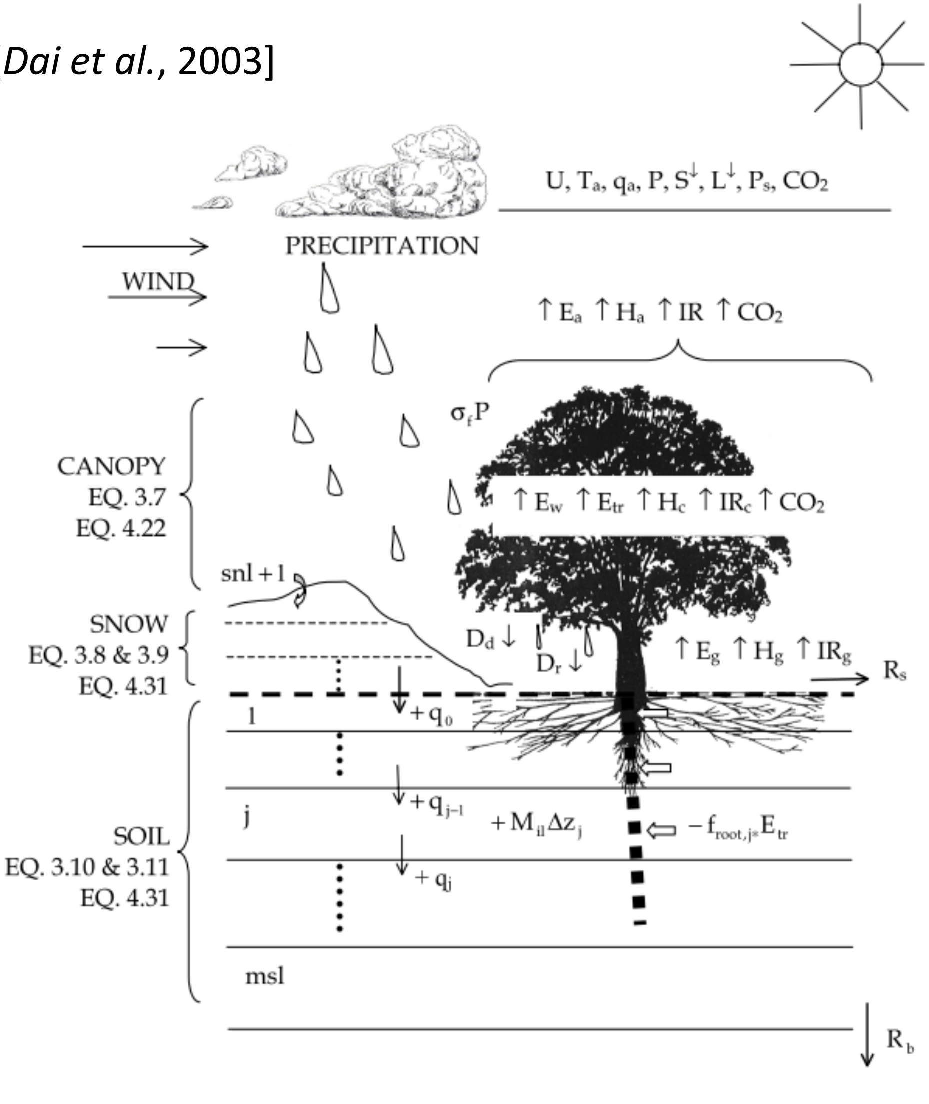


Figure 2.2. Structure of the CLM model. CLM has one vegetation layer, ten soil layers and up to 5 snow layers depending on the snow depth.

Six examined output fluxes:

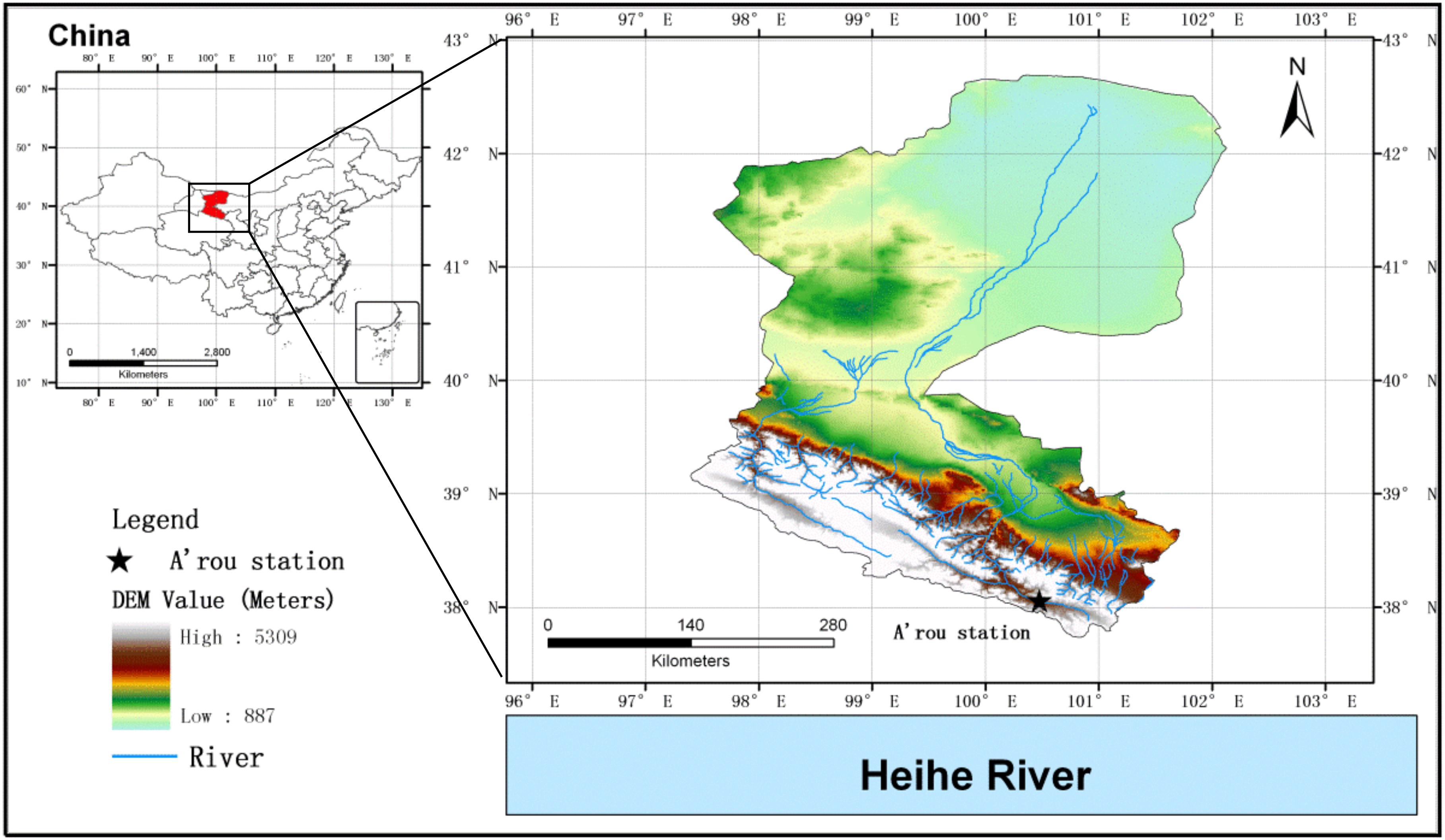
- *sensible heat*
- *latent heat*
- *upward longwave radiation*
- *net radiation*
- *soil temperature*
- *soil moisture*

Objective function:

Normalized RMSE

$$NRMSE_i = \frac{\sqrt{\sum_{j=1}^{N}\left(y_{i,j}^{sim} - y_{i,j}^{obs}\right)^2}}{\sum_{j=1}^{N} y_{i,j}^{obs}}$$

Location of The Study area.



Spin up: 2008-Jan-01~2008-Dec-31

Simulation: 2009-Jan-01~2009-Dec-31

Adjustable Parameters and Ranges of CoLM

Beijing Normal University

Index	Parameter	Physical meaning	Range
P1	dewmx	maximum ponding of leaf area	[0.05, 0.15]
P2	hksati	maximum hydraulic conductivity	[0.001, 1]
P3	porsl	porosity	[0.25, 0.75]
P4	phi0	minimum soil suction	[50, 500]
P5	wtfact	Fraction of shallow groundwater area	[0.15, 0.45]
P6	bsw	Clapp and Hornberger "b" parameter	[2.5, 7.5]
P7	wimp	infimum of porosity	[0.01, 0.1]
P8	zlnd	roughness length for soil surface	[0.005, 0.015]
P9	pondmx	maximum ponding depth for soil surface	[5, 15]
P10	csoilc	drag coefficient for soil under canopy	[0.002, 0.006]
P11	zsno	roughness length for snow	[0.0012, 0.0036]
P12	capr	tuning factor of soil surface temperature	[0.17, 0.51]
P13	cnfac	Crank Nicholson factor	[0.25, 0.5]
P14	slti	slope of low temperature inhibition function	[0.1, 0.3]
P15	hlti	1/2 point of low temperature inhibition function	[278, 288]
P16	shti	slope of high temperature inhibition function	[0.15, 0.45]
P17	sqrtdi	the inverse of square root of leaf dimension	[2.5, 7.5]
P18	effcon	quantum efficiency of vegetation photosynthesis	[0.035, 0.35]
P19	vmax25	maximum carboxylation rate at 25℃	[10e-06, 200e-06]
P20	hhti	1/2 point of high temperature inhibition function	[305, 315]
P21	trda	temperature coefficient of conductance-photosynthesis model	[0.65,1.95]
P22	trdm	temperature coefficient of conductance-photosynthesis model	[300, 350]
P23	trop	temperature coefficient of conductance-photosynthesis model	[250, 300]
P24	gradm	slope of conductance-photosynthesis model	[4, 9]
P25	binter	intercept of conductance-photosynthesis model	[0.01, 0.04]
P26	extkn	coefficient of leaf nitrogen allocation	[0.5, 0.75]
P27	chil	leaf angle distribution factor	[-0.3, 0.1]
P28	ref(1,1)	VIS reflectance of living leaf	[0.07, 0.105]
P29	ref(1,2)	VIS reflectance of dead leaf	[0.16, 0.36]
P30	ref(2,1)	NIR reflectance of living leaf	[0.35, 0.58]
P31	ref(2,2)	NIR reflectance of dead leaf	[0.39, 0.58]
P32	tran(1,1)	VIS transmittance of living leaf	[0.04, 0.08]
P33	tran(1,2)	VIS transmittance of dead leaf	[0.1, 0.3]
P34	tran(2,1)	NIR transmittance of living leaf	[0.1, 0.3]
P35	tran(2,2)	NIR transmittance of dead leaf	[0.3, 0.5]
P36	z0m	aerodynamic roughness length	[0.05, 0.3]
P37	ssi	irreducible water saturation of snow	[0.03, 0.04]
P38	smpmax	wilting point potential	[-2.e5, -1.e5]
P39	smpmin	restriction for min of soil potential	[-1.e8, -9.e7]
P40	trsmx0	maximum transpiration for vegetation	[1.e-4, 100. e-4]

Parameter Screening based on Qualitative & Quantitative Sensitivity Analysis (SA)

- **Qualitative methods:**
  1. Morris method: based on one-at-a-time sampling (MOAT)
  2. Delta test: Based on nearest-neighbor analysis
  3. Random Forest: Belongs to the class of tree-based methods
  4. MARS: Multivariate Adaptive Regression Spline

- **Quantitative method:**
  1. RSMSobol: (**R**esponse **S**urface **M**ethod based **Sobol**' method)
     Based on variance decomposition

[Jianduo Li et.al. 2013, HESS]

Finding 1: 200-400 Model Runs Are Enough For Parameter Screening Based on Qualitative SA Methods

Different Sample Sizes

Different DoE Methods

Monte Carlo

Latin Hypercube

LPtau

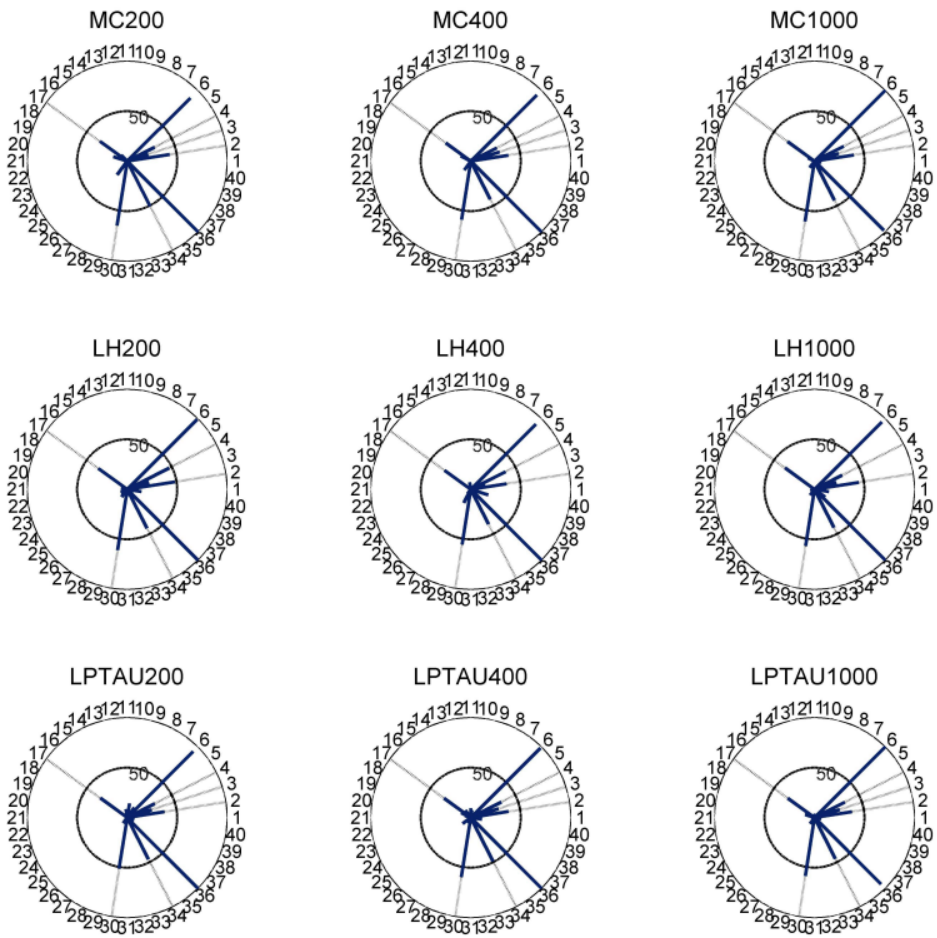
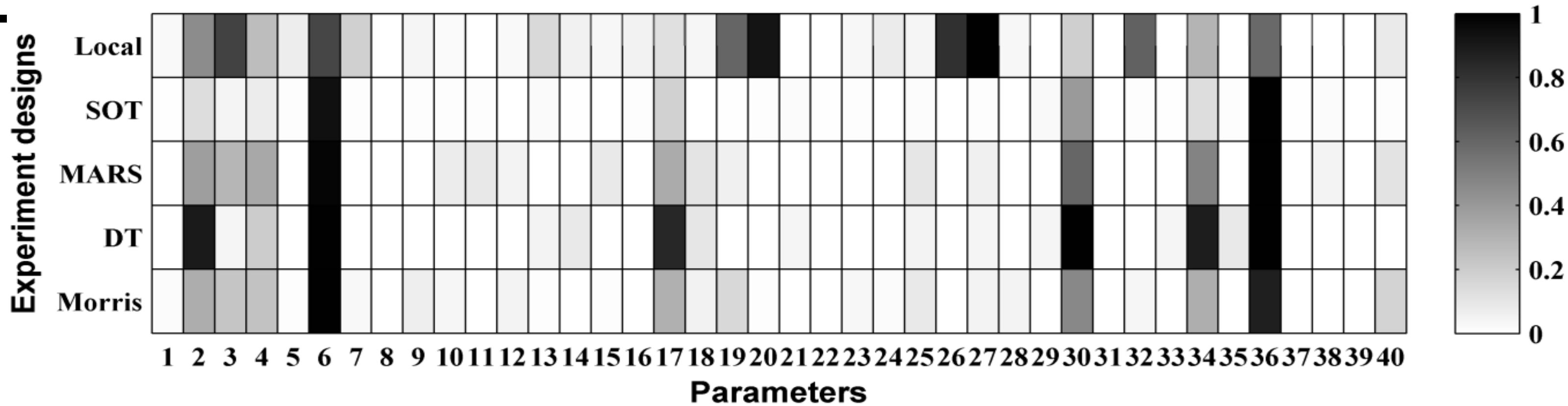


Fig. 3. The sensitivity score of sensible heat given by MARS. The length of needles represents the sensitivity score.

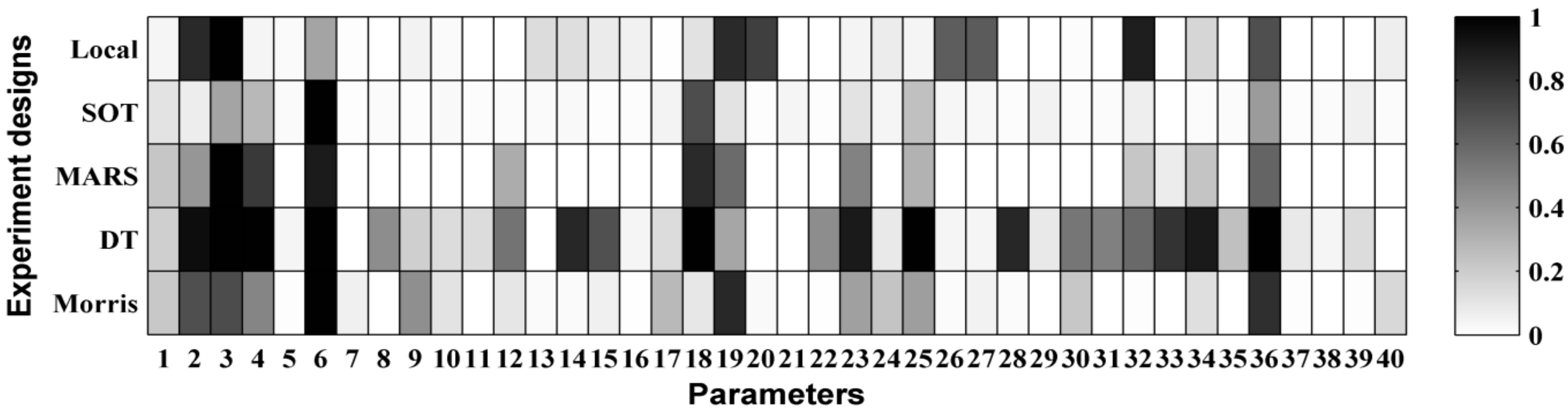
[Jianduo Li et.al. 2013, HESS]

Finding 2: Global Qualitative SA Methods Can Consistently Identify the Sensitive/Insensitive Parameters

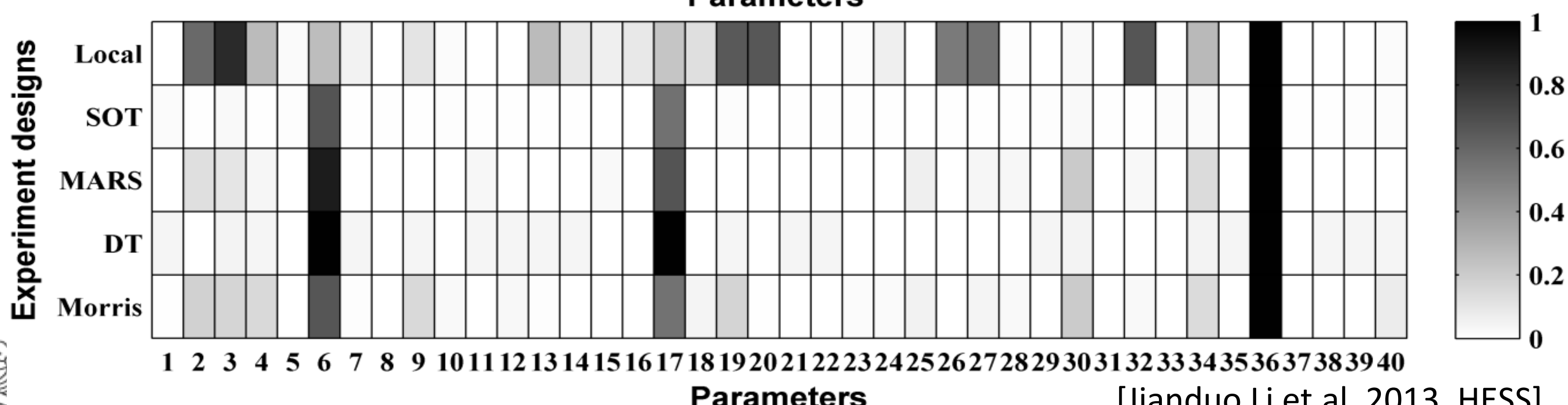
Sensible heat



Latent heat



Upward longwave radiation



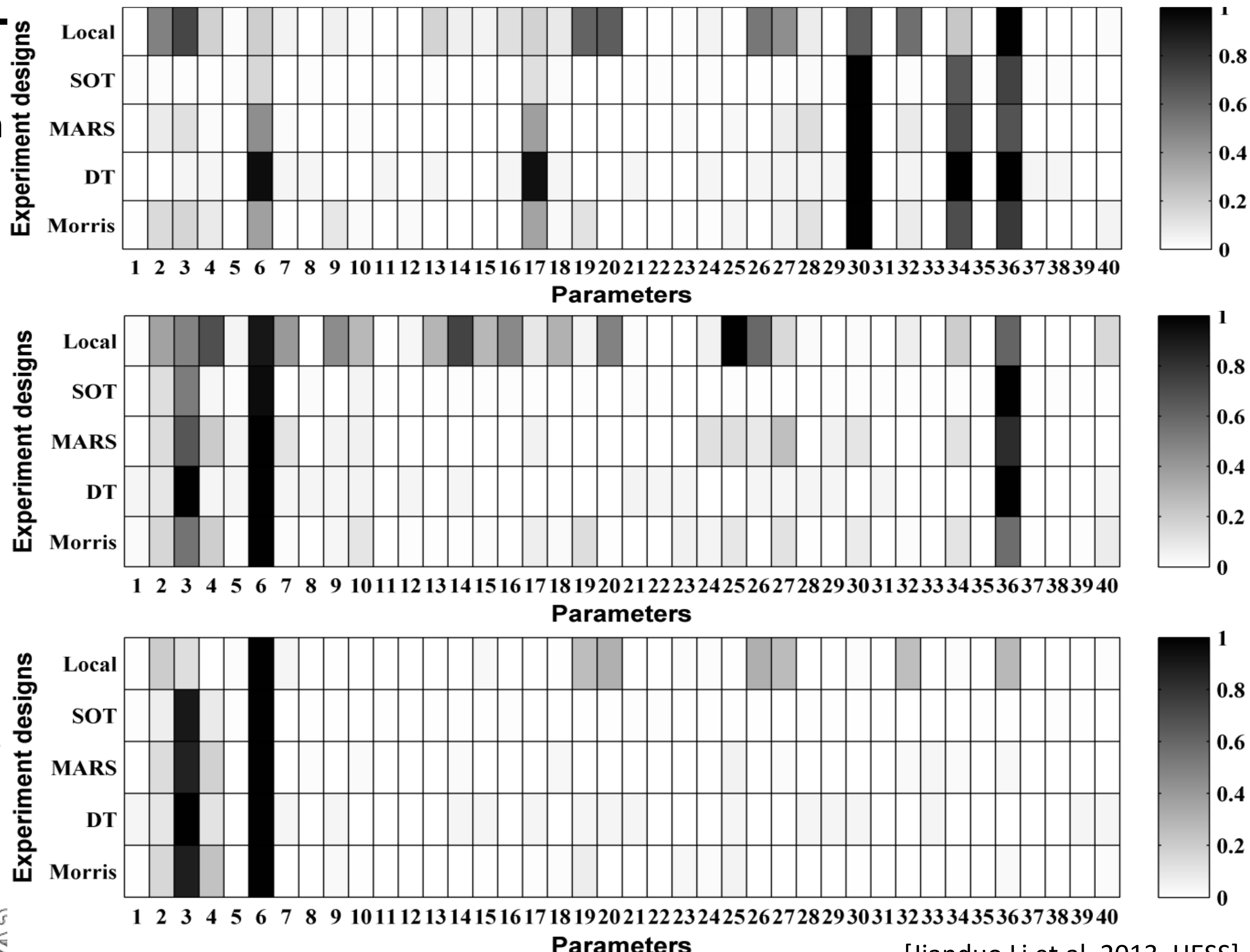
[Jianduo Li et.al. 2013, HESS]

Finding 2: Global Qualitative SA Methods Can Consistently Identify the Sensitive/Insensitive Parameters

Net radiation

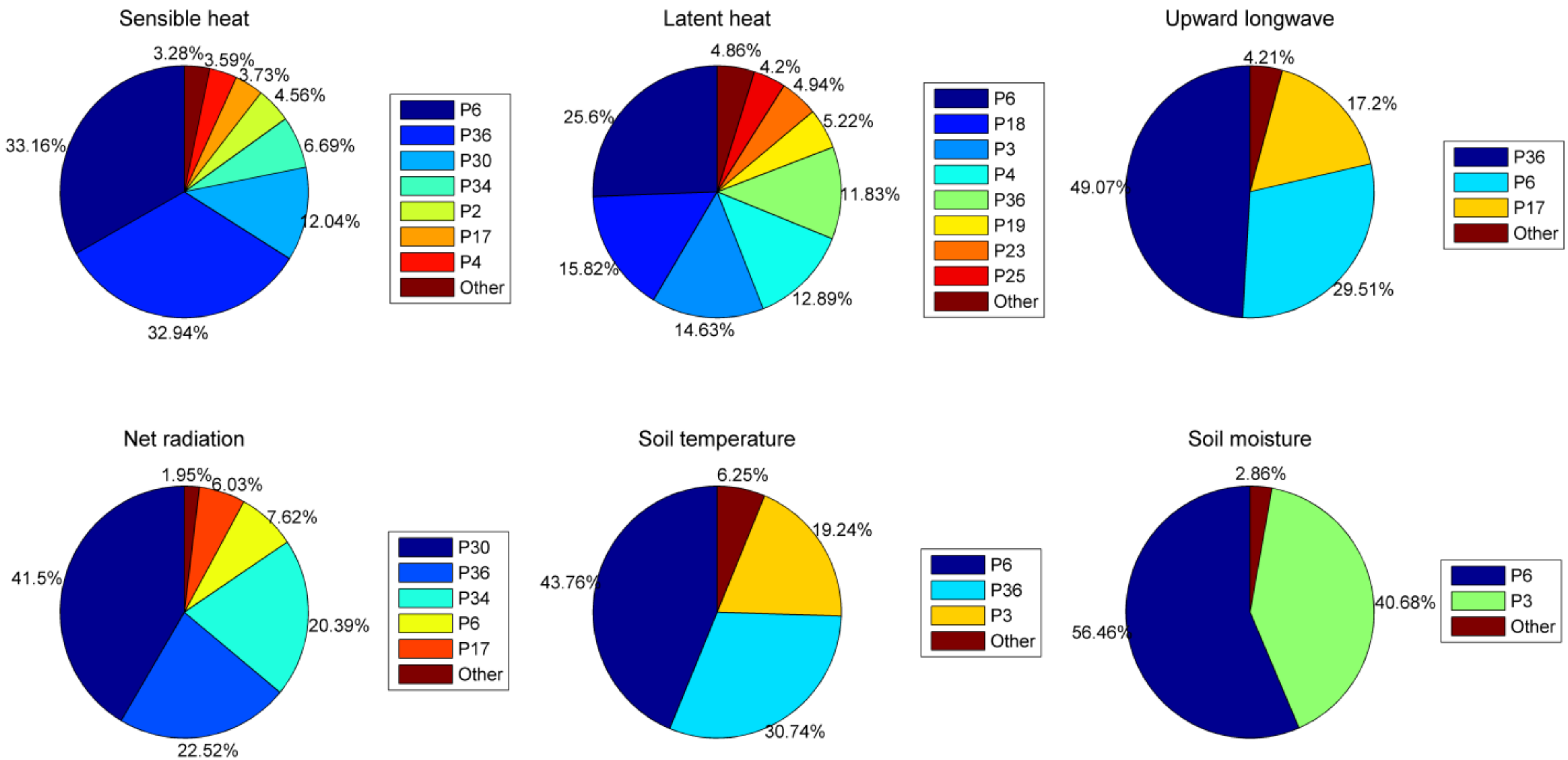
Soil temperature

Soil moisture



[Jianduo Li et.al. 2013, HESS]

Finding 3: The Effectiveness Is Confirmed by the Quantitative Sobol' Method*.



* 2000 Samples are used

[Jianduo Li et.al. 2013, HESS]

Case Study 2:

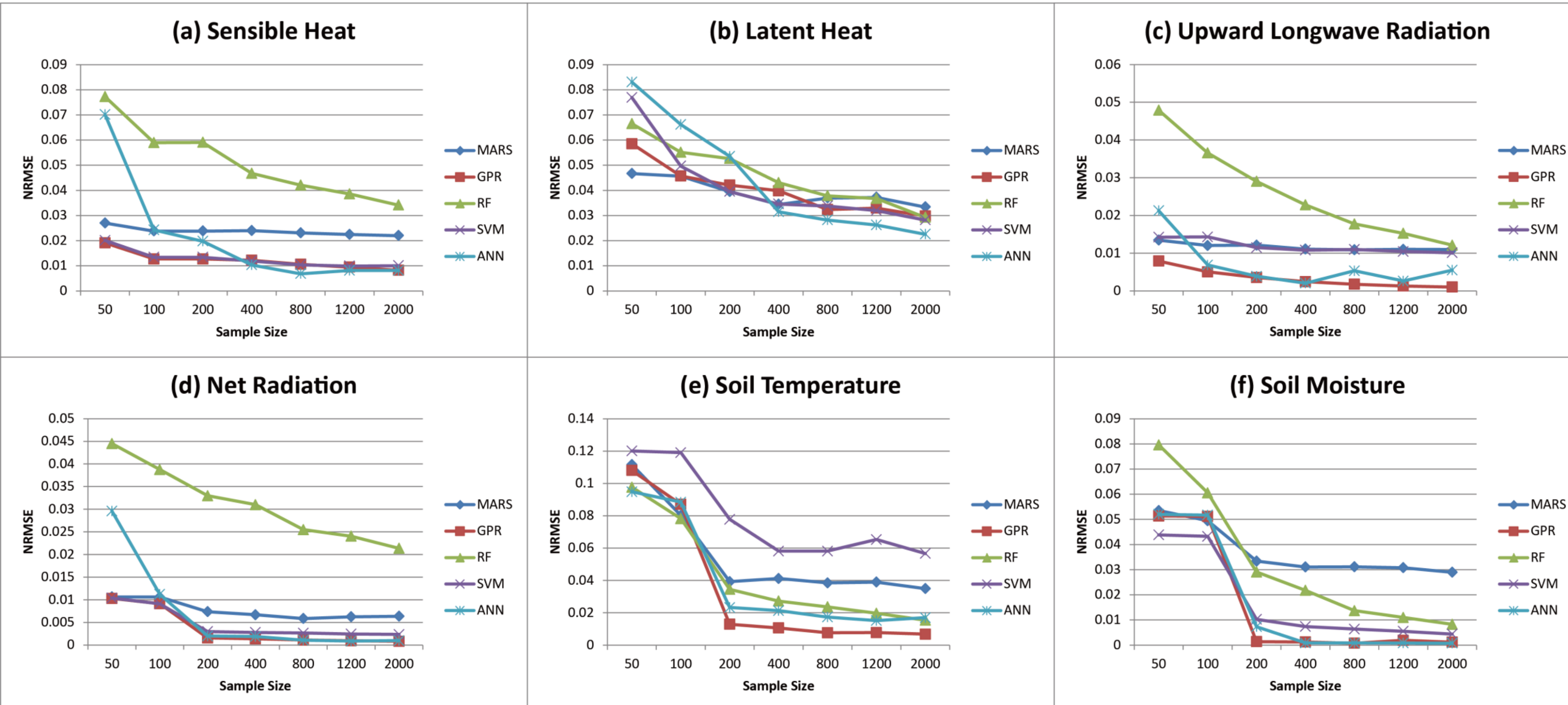
SURROGATE MODELING: SINGLE AND MULTI-OBJECTIVE OPTIMIZATION OF COLM PARAMETERS

An Inter-comparison of Surrogate Modeling Methods – Non-Adaptive

- MARS (**M**ultivariate **A**daptive **R**egression **S**plines):
- RF (**R**andom **F**orest)
- GPR (**G**aussian **P**rocesses **R**egression)
- SVM (**S**upport **V**ector **M**achine)
- ANN (**A**rtificial **N**eural **N**etwork)

An Inter-comparison of Surrogate Modeling Methods

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science
GCESS



Findings:

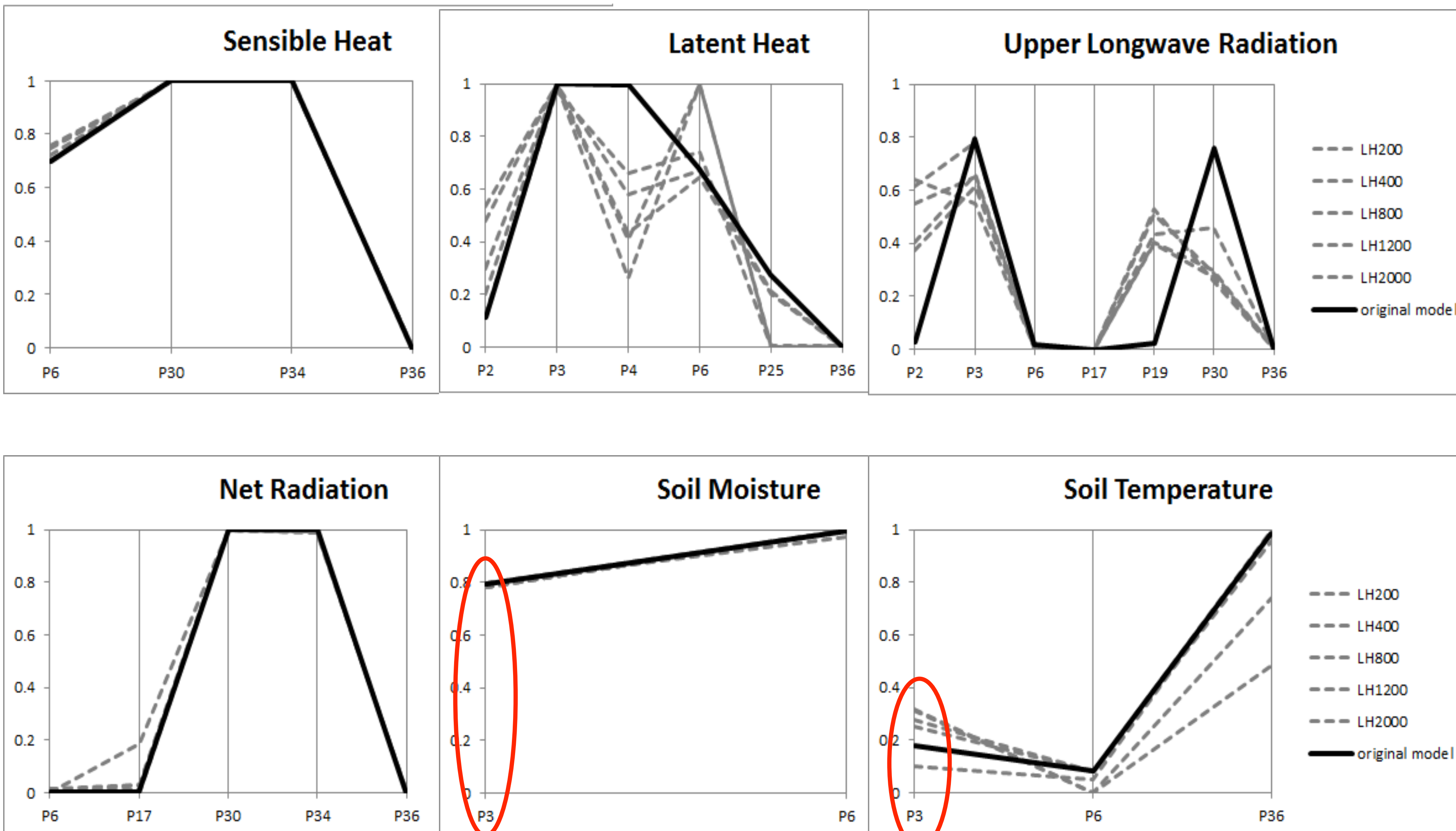
1) GPR is the best one among all 5 surrogate modeling methods.

2) 200-2000 sample points are enough for surrogate model construction.

[Wei Gong et.al. 2015, HESS]

Single Objective Parameter Optimization Results

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science
GCESS

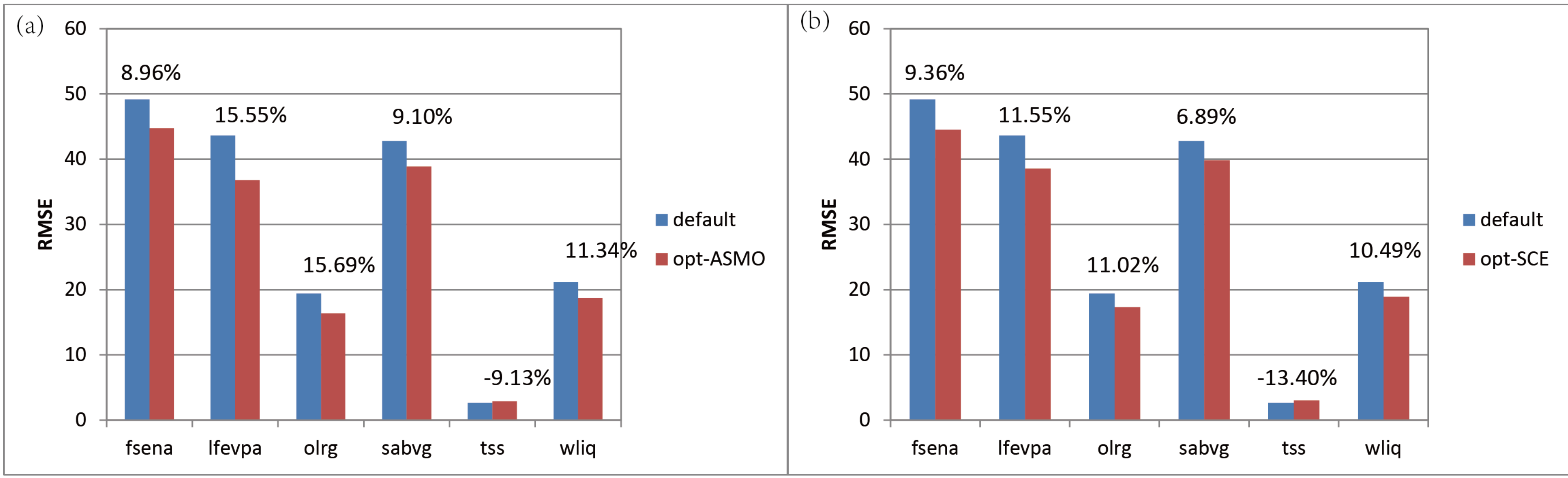


Conflicting optima for different objectives

[Wei Gong et.al. 2015, HESS]

Multi-Objective Optimization Results

北京师范大学
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science
GCESS



ASMO:

1) Use GPR as surrogate model
2) Initial sample size is 400
3) Optimization completed at 411 model runs

SCE:

1) Number of complex = 4
2) Maximum model runs = 1000
3) Optimization completed at 1089 model runs

Key Findings:

1) All fluxes are improved at least 8.96% except soil temperature.
2) ASMO performs better than SCE-UA with fewer model runs.

Summary of Sensitivity Analysis and Optimization Results

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science
GCESS

- Global SA methods are generally effective, but local method is not
- ~200-400 model runs are enough for screening of 40 parameters using qualitative SA methods
- 2-8 parameters found to be sensitive for different surface fluxes
- Five kinds of surrogate modeling methods (MARS, GPR, RF, SVM, ANN) are compared, and the fitting ability of GPR is the best.
- Single-objective optimization may lead to conflicting optimal parameter sets.
- Multi-objective optimization can improve simulation of almost all fluxes

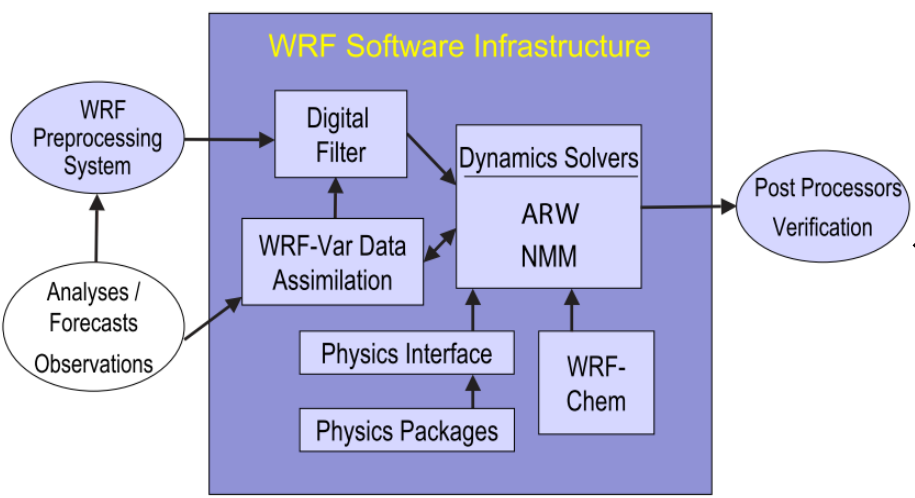
Case Study 3:

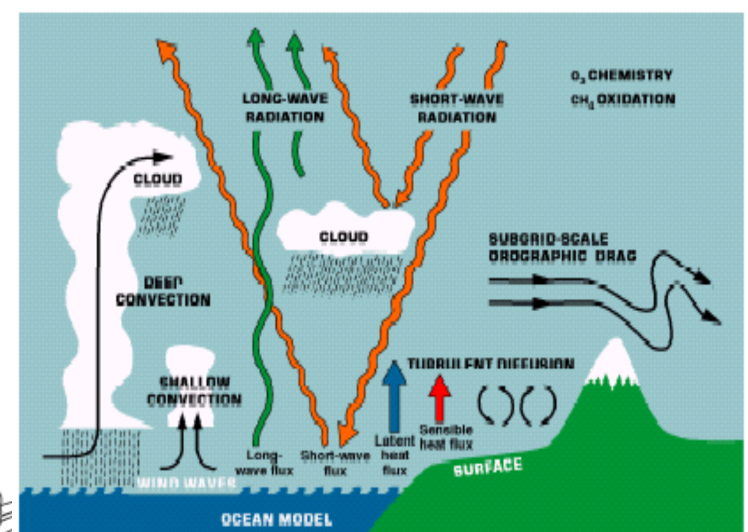
ANALYSIS AND OPTIMIZATION OF PARAMETRIC UNCERTAINTY OF WRF MODEL

Analysis of Parametric Uncertainty of WRF Model

北京师范大学 Beijing Normal University
GCESS 全球变化与地球系统科学研究院 College of Global Change and Earth System Science

- Weather and Research Forecast (WRF) is a widely used regional weather and climate modeling system. The model includes seven major physical processes:
    - Microphysics
    - Cumulus Cloud
    - Surface Layer
    - Land-Surface
    - Planetary Boundary Layer
    - Longwave Radiation
    - Shortwave Radiation





Definition of the Problem

北京師範大學
Beijing Normal University
GCESS
全球变化与地球系统科学研究院
College of Global Change and Earth System Science

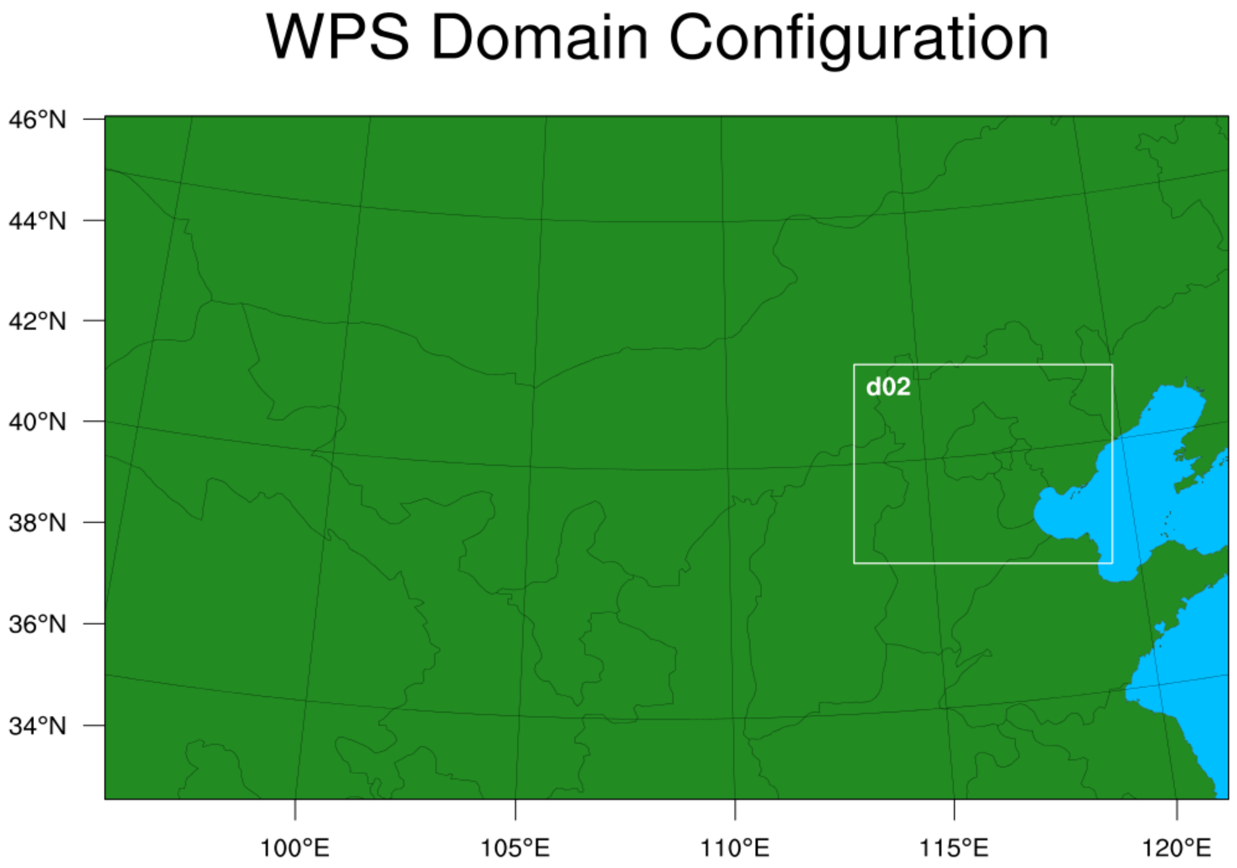
- Many processes and many choices:

Microphysics	Long-wave radiation	Short-wave radiation	Surface layer	Land surface	PBL	Cumulus
14	6	6	7	5	10	7

- **There are 14×6×6×7×5×10×7=1234800（combinations）**

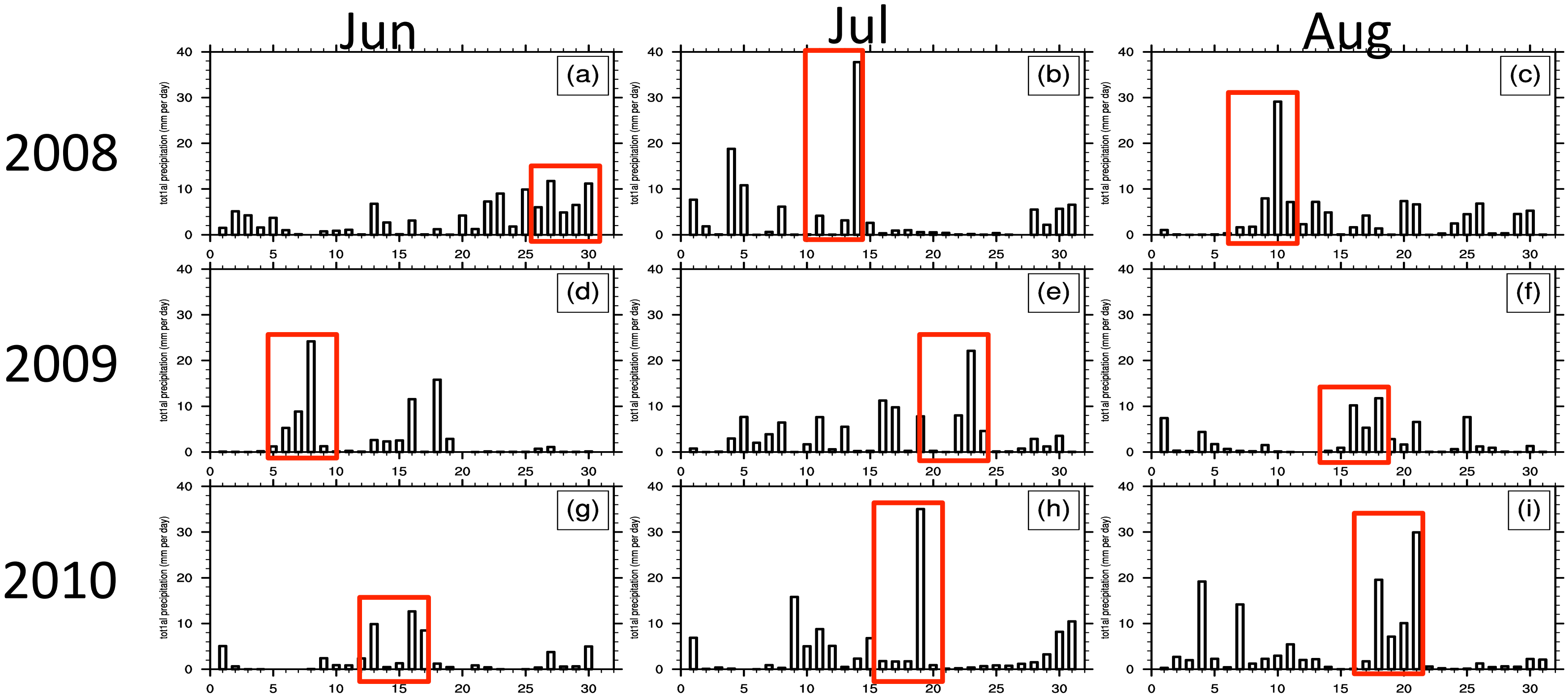
- Problems:
  - What physical processes are most sensitive to the meteorological variables of interest?
  - Which combinations of physical parameterization schemes are the best?
  - **Given a combination of physical parameterization scheme, how can we find the most sensitive parameters and their optimal values?**

The Study Domain



- 2-level nested grids
  - Level 1: 27km, 60×48 grids
  - Level 2: 9km, 87x55 grids

Forecasted Events



降雨事件	模拟日期	模拟日期	模拟日期
(a)---(c)	20080626-20080630	20080710-20080714	20080807-20080811
(d)---(f)	20090605-20090609	20090720-20090724	20090814-20090818
(g)---(i)	20100613-20100617	20100716-20100720	20100817-20100821

WRF Model Parameters To Be Examined

number	scheme	name	Default	range	description
1	Surface layer (module_sf_sfclay.F)	xka	0.000024	[0.000012 0.00005]	The parameter for heat/moisture exchange coefficient
2		CZO	0.0185	[0.01 0.037]	The coefficient for coverting wind speed to roughness length over water
3	Cumulus (module_cu_kfeta.F)	pd	0	[-1 1]	The coefficient related to downdraft mass flux rate
4		pe	0	[-1 1]	The coefficient related to entrainment mass flux rate
5		ph	150	[50 350]	Starting height of downdraft above USL
6		TIMEC	2700	[1800 3600]	Compute convective time scale for convection
7		TKEMAX	5	[3 12]	the maximum turbulent kinetic energy (TKE) value between the level of free convection (LFC)and lifting condensation level (LCL)
8	Microphysics (module_mp_wsm6.F)	ice_stokes_fac	14900	[8000 30000]	Scaling factor applied to ice fall velocity
9		n0r	8000000	[5000000 12000000]	Intercept parameter rain
10		dimax	0.0005	[0.0003 0.0008]	The limited maximum value for the cloud-ice diameter
11		peaut	0.55	[0.35 0.85]	Collection efficiency from cloud to rain auto conversion
12	short wave radiation (module_ra_sw.F)	cssca	0.00001	[0.000005 0.00002]	Scattering tuning parameter in clear sky
13		Beta_p	0.4	[0.2 0.8]	Aerosol scattering tuning parameter
14	Longwave (module_ra_rrtm.F)	Secang	1.66	[1.55 1.75]	Diffusivity angle
15	Land surface (module_sf_noahlsm.F)	hksati	0	[-1 1]	hydraulic conductivity at saturation
16		porsl	0	[-1 1]	fraction of soil that is voids
17		phi0	0	[-1 1]	minimum soil suction
18		bsw	0	[-1 1]	Clapp and hornbereger "b" parameter
19	Planetary Boundary Layer (module_bl_ysu.F)	Brcr_sbrob	0.3	[0.15 0.6]	Critical Richardson number for boundary layer of water
20		Brcr_sb	0.25	[0.125 0.5]	Critical Richardson number for boundary layer of land
21		pfac	2	[1 3]	Profile shape exponent for calculating the momentum diffusivity coefficient
22		bfac	6.8	[3.4 13.6]	Coefficient for prandtl number at the top of the surface laer
23		sm	15.9	[12 20]	Countergradient proportional coefficient of non-local flux of momentum moh 2002

The Experimental Setup (1): Model Setup

系统科学研究院
d Earth System Science

- 2-Level nested grids
  - Level 1: 27 km, with 60x48 grids
  - Level 2: 9 km, with 87x55 grids
- Nine 5-day forecasts during Jun-Aug from 2008-2010
  - 1st day as spin-up, last 4 day results analyzed
- NCEP reanalysis data used to initiate the forecasts
- 23 WRF model parameters examined for study their sensitivity with respect to precipitation forecast
- Computational cost
  - 4.5 CPUs for one 5-day forecast
  - Nine 5-day forecasts require 180 CPUs

The Experimental Setup (2) – Validation Datasets

Table 1 Ground observation data products

Variable name	Horizontal resolution	Temporal resolution	time range	source
Precipitation, Temperature, Wind speed, Wind direction, Humidity, Pressure, Downward shortwave radiation, upward shortwave radiation	0.05^{o}	3 hours	2008-2010	BNU Zheng Group
	$(1/16)^{o}$	1hour	2011-2013	CMA

Table 2 Other observation data products

Variable name	Product name	Horizontal resolution	Temporal resolution	time range	Source
Cloud Fraction	MOD06_L2-Level 2 Cloud Product; MYD06_L2-Level 2 Cloud Product;	5km×5km	Time-varying	MOD: 1999-2014 MYD: 2002-2014	http://ladsweb.nascom.nasa.gov/data/search.html
Total Precipitable Water	Aqua AIRS Level 2 Standard Physical Retrieval (AIRS+AMSU) (AIRX2RET.006)	50km×50km	Time-varying	2002-2014	http://disc.sci.gsfc.nasa.gov/
Boundary Layer Height	MERRA Chem 2D IAU Diagnostics, Fluxes and Meteorology, Time Average 3-hourly (MAT3FXCHM.5.2.0)	1.25°×1°	3 hours	1979-2014	http://disc.sci.gsfc.nasa.gov/
Upward long wave radiation at top of the atmosphere	FY-2D卫星9210格式日平均射出长波辐射产品	0.1°×0.1°	1 day	2007-2014	http://satellite.cma.gov.cn/PortalSite/Data/Satellite.aspx

The Experimental Setup (3) – Analysis Method

- Sensitivity Analysis method used:
  - Morris One-At-a-Time (MOAT)

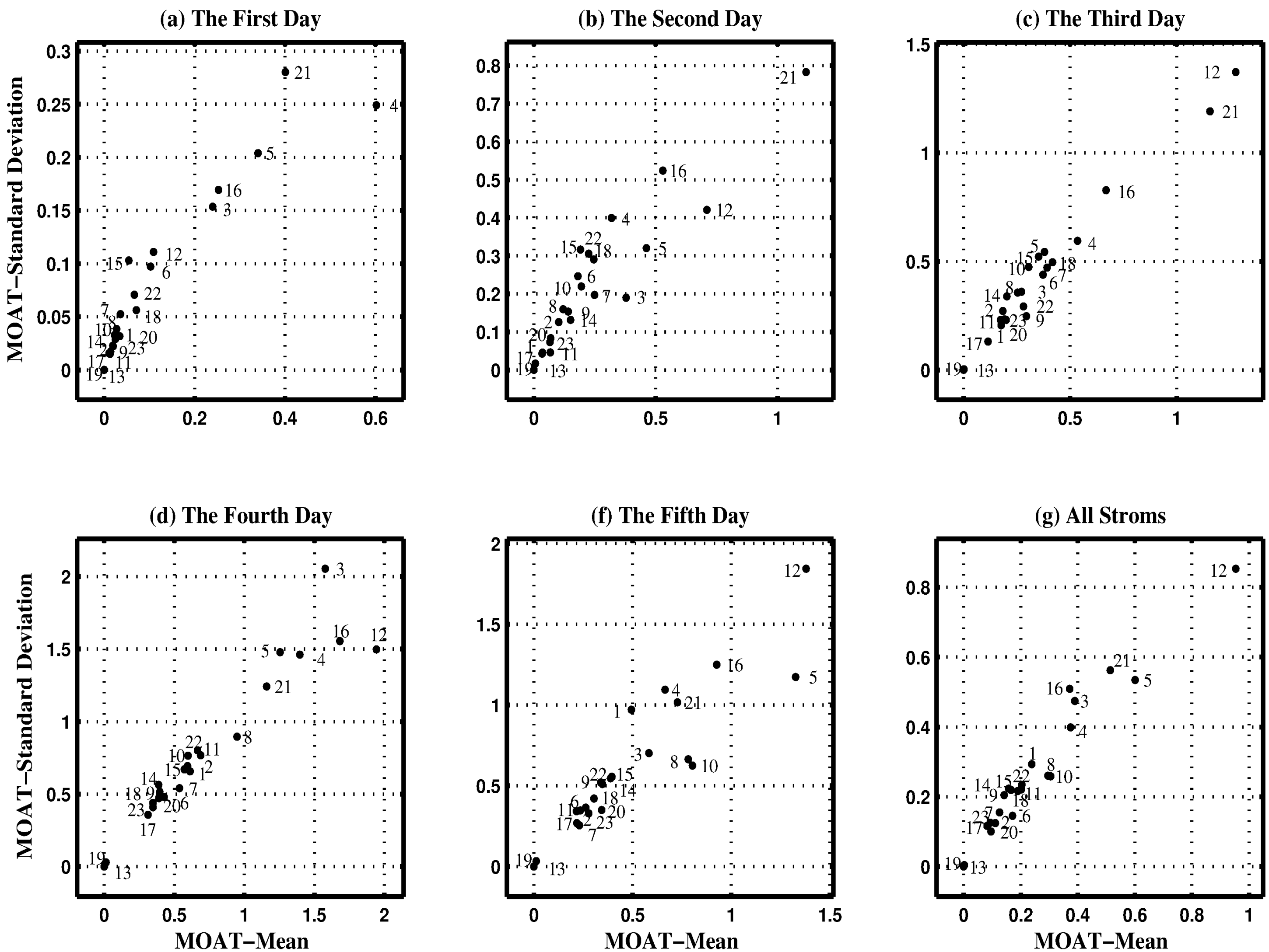
- Objective function used:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | Sim_i - Obs_i |$$

sim_i and obs_i are the forecasted and observed daily precipitation at i^{th} grid

- Number of parameter samples used: 240
  - Total CPU hours: 240 x 180 = 43,000 hours

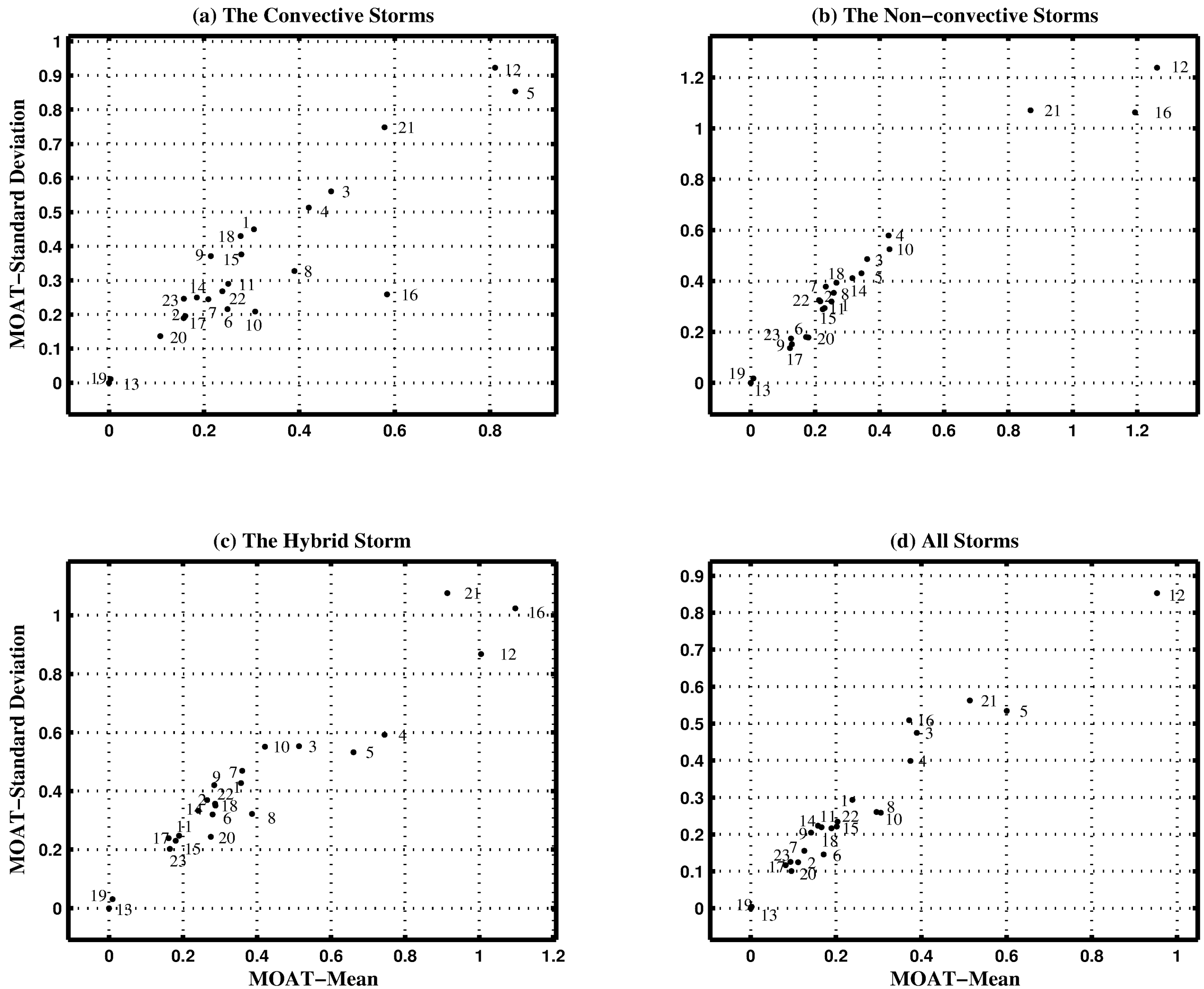
MOAT Results – Precipitation based on lead times:



Sensitive parameters for precipitation: P3, P4, P5, P12, P16, P21

[Zhenhua Di et.al. 2015, GRL]

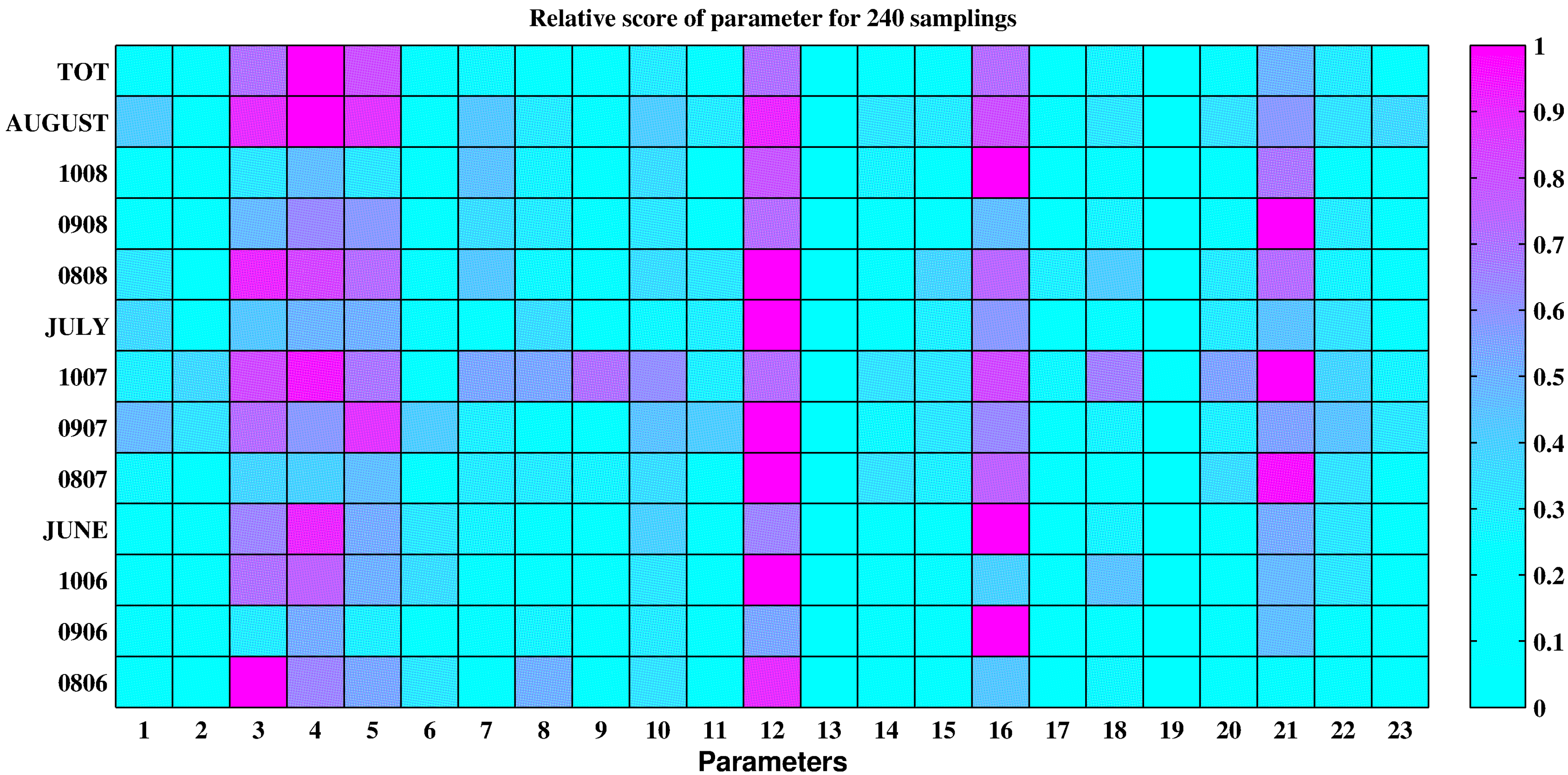
MOAT Results – Precipitation based on storm types:



Sensitive parameters for precipitation: P3, P4, P5, P12, P16, P21

[Zhenhua Di et.al. 2015, GRL]

MOAT Normalized Results - Precipitation:



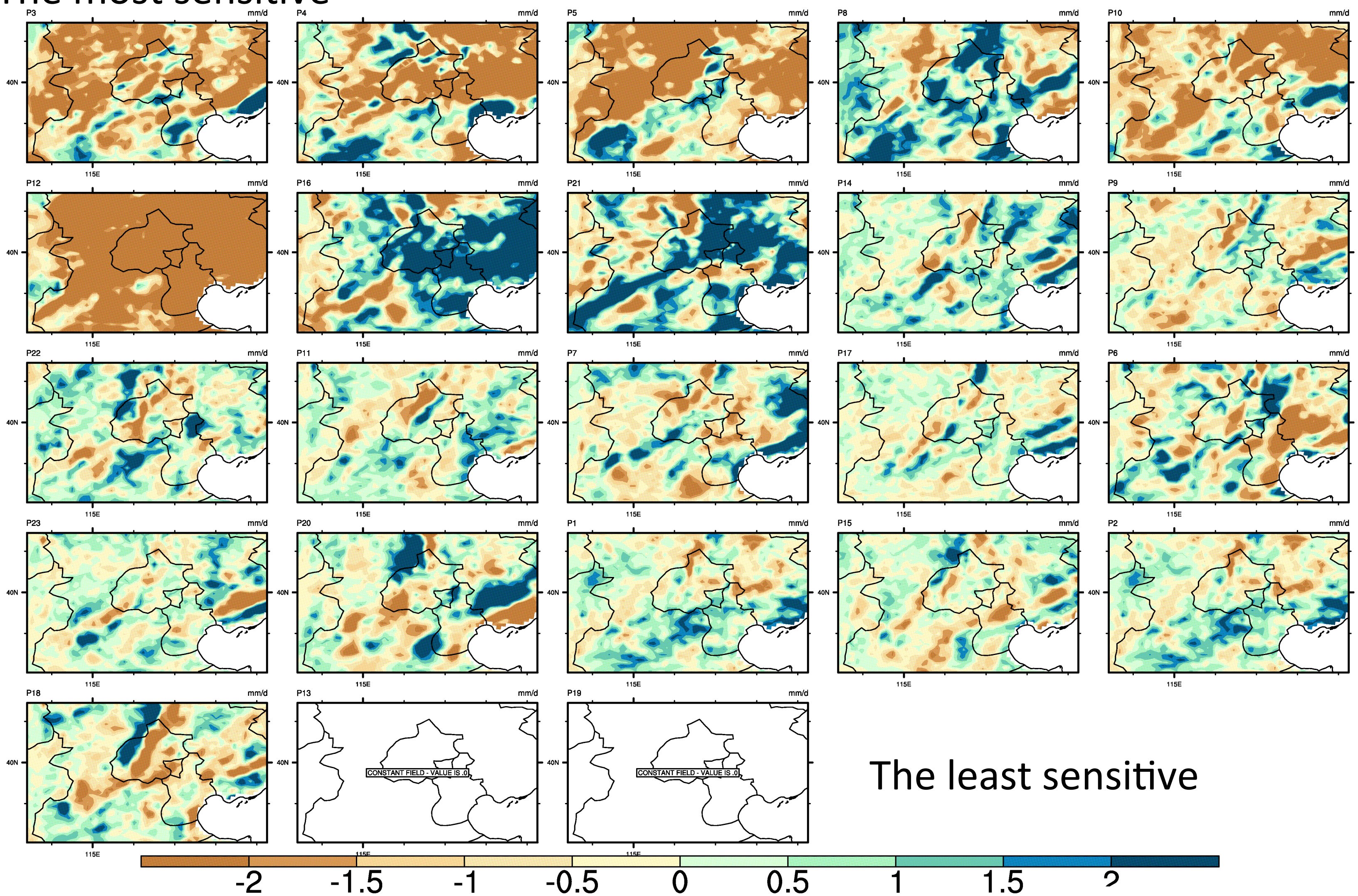
All parameters normalized to [0 1] range, with purple red indicating sensitive, cyan indicating insensitive. Sensitive parameters found: P3、P4、P5、P12、P16、P21

Cumulus:	**P3、P4、P5;**
Shortwave radiation:	**P12;**
Land surface:	**P16;**
Planatary BL:	**P21;**

[Zhenhua Di et.al. 2015, GRL]

Comparison of Sensitivities to Precipitation Forecasts $(P_{\{max\}}-P_{\{min\}})$

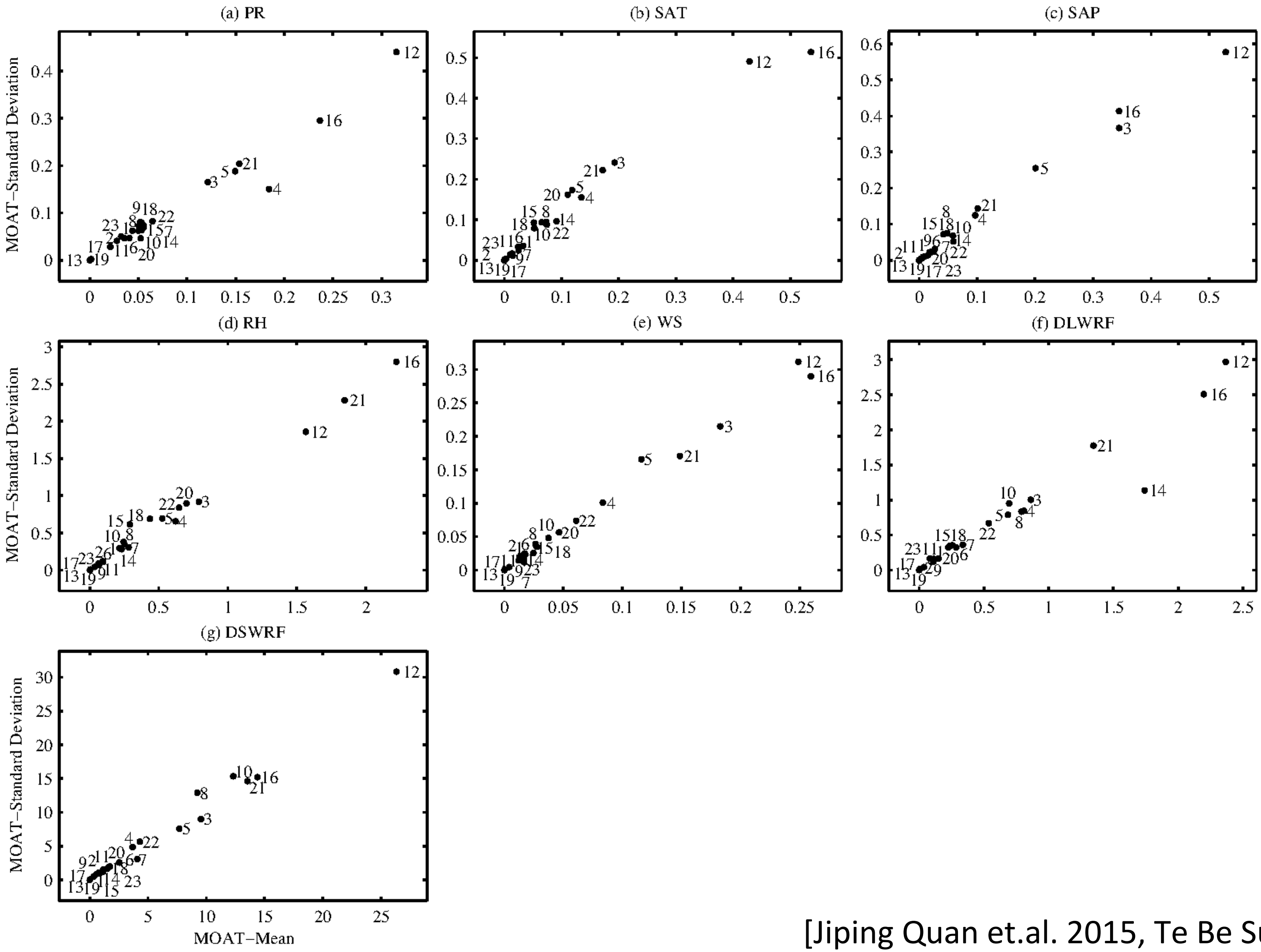
The most sensitive



The least sensitive

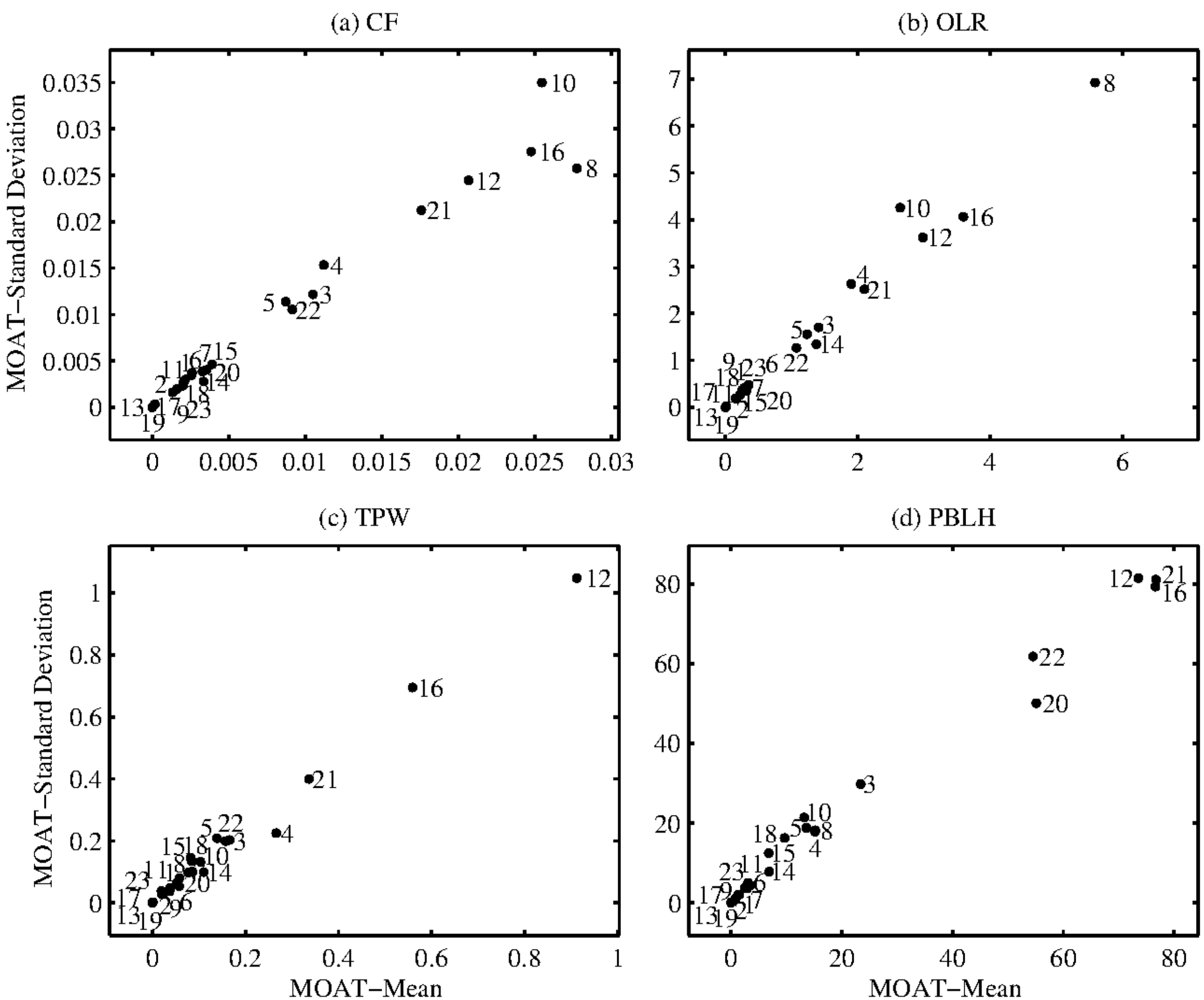
[Zhenhua Di et.al. 2015, GRL]

MOAT Results – Surface Meteorological Variables



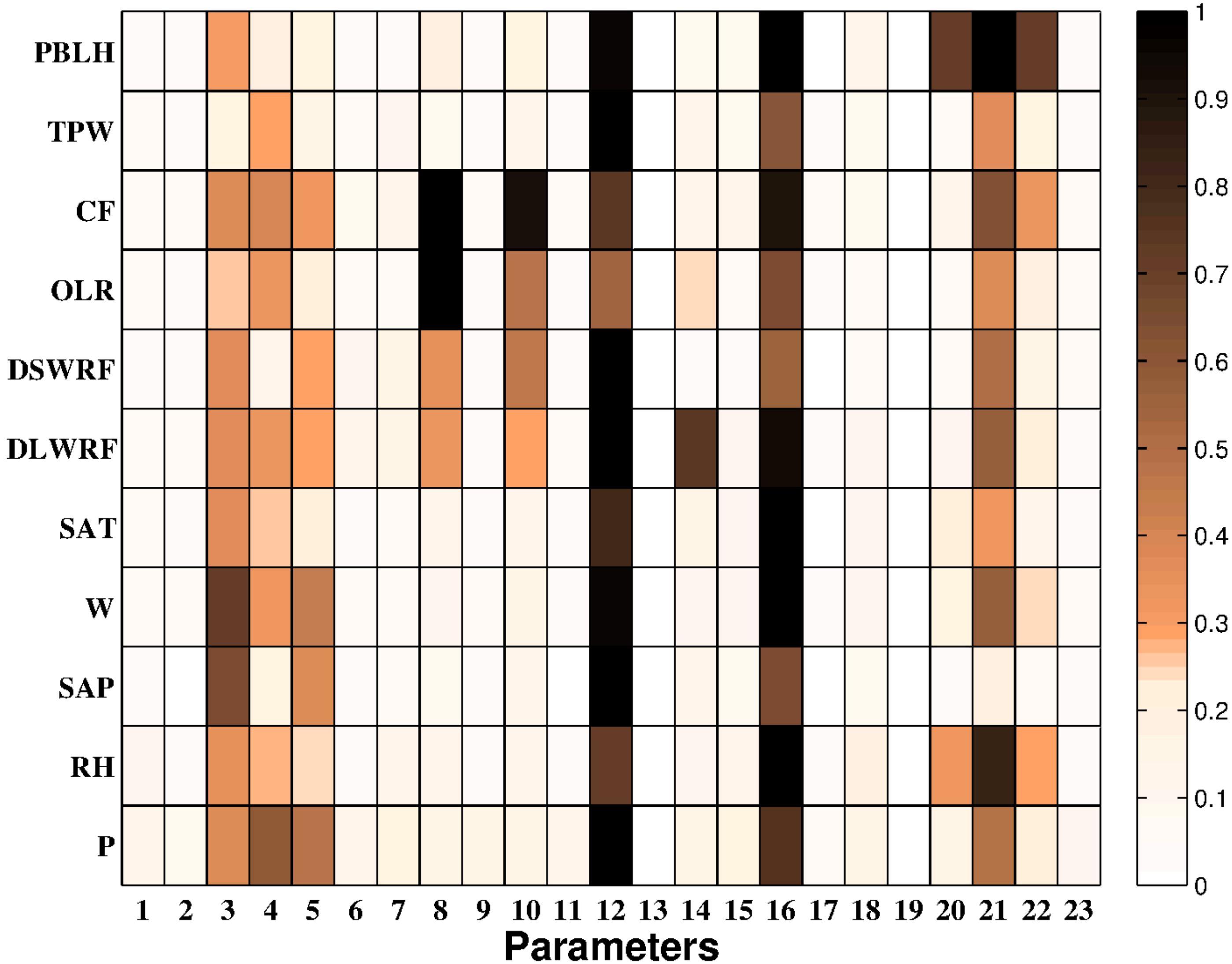
[Jiping Quan et.al. 2015, Te Be Submitted]

MOAT Results – Atmospheric Variables



[Jiping Quan et.al. 2015, Te Be Submitted]

Summary of Parameter Sensitivities to Different Model Outputs



[Jiping Quan et.al. 2015, Te Be Submitted]

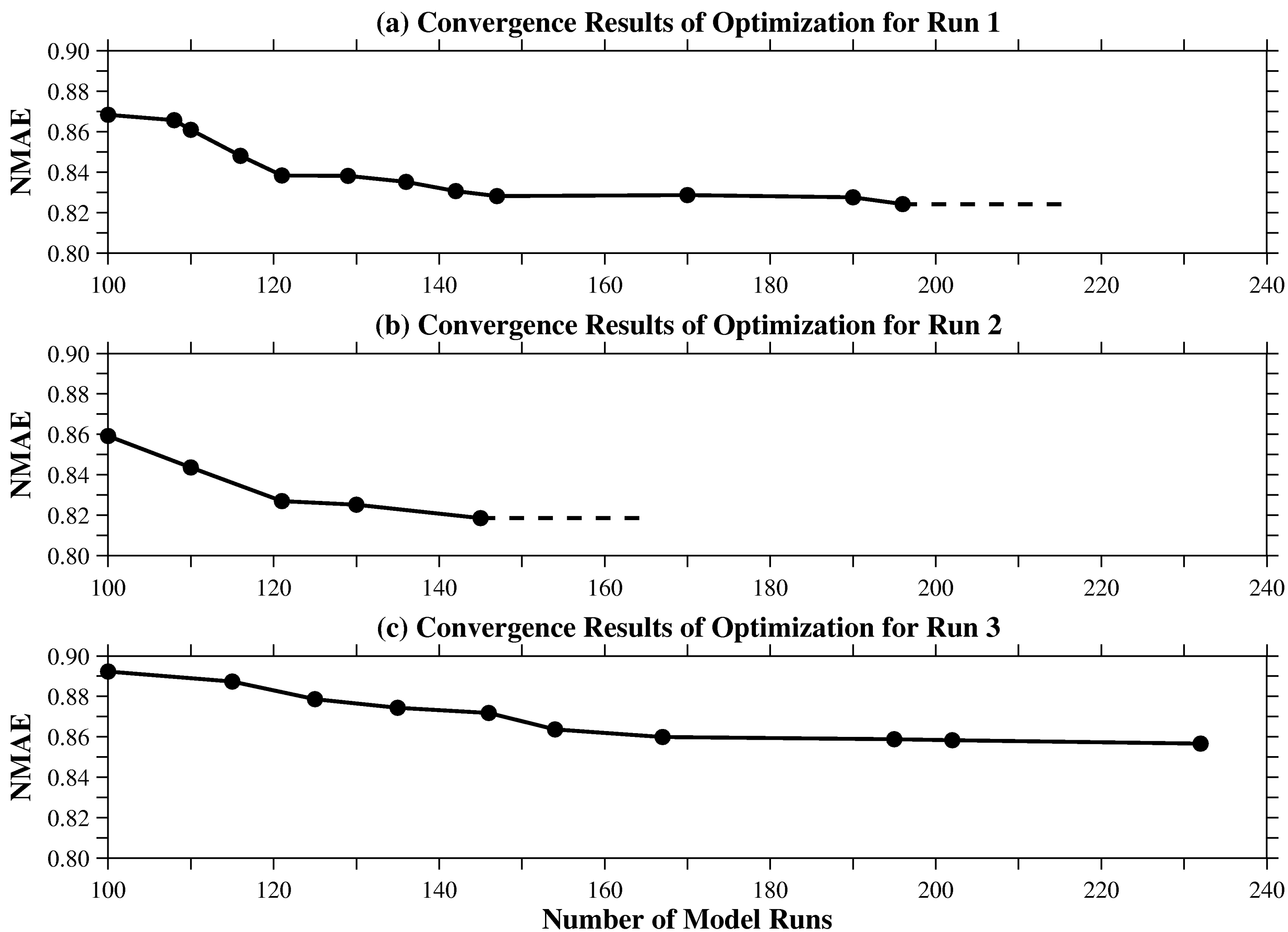
Optimization Experiment Setup

- Adaptive Surrogate Modeling based Optimization (ASMO) method is used to optimize the eight most sensitive parameters found by global sensitivity analysis:
  - Parameter optimized: P3、P4、P5、P8、P10、P12、P16、P21
  - GP surrogate model is created with 100 initial samples generated using LPtau design
  - Adaptive search is then conducted to update the GP surrogate model (i.e., by adding more samples points based on existing response surface)
  - Objective Function Used – Mean Absolute Error (MAE):

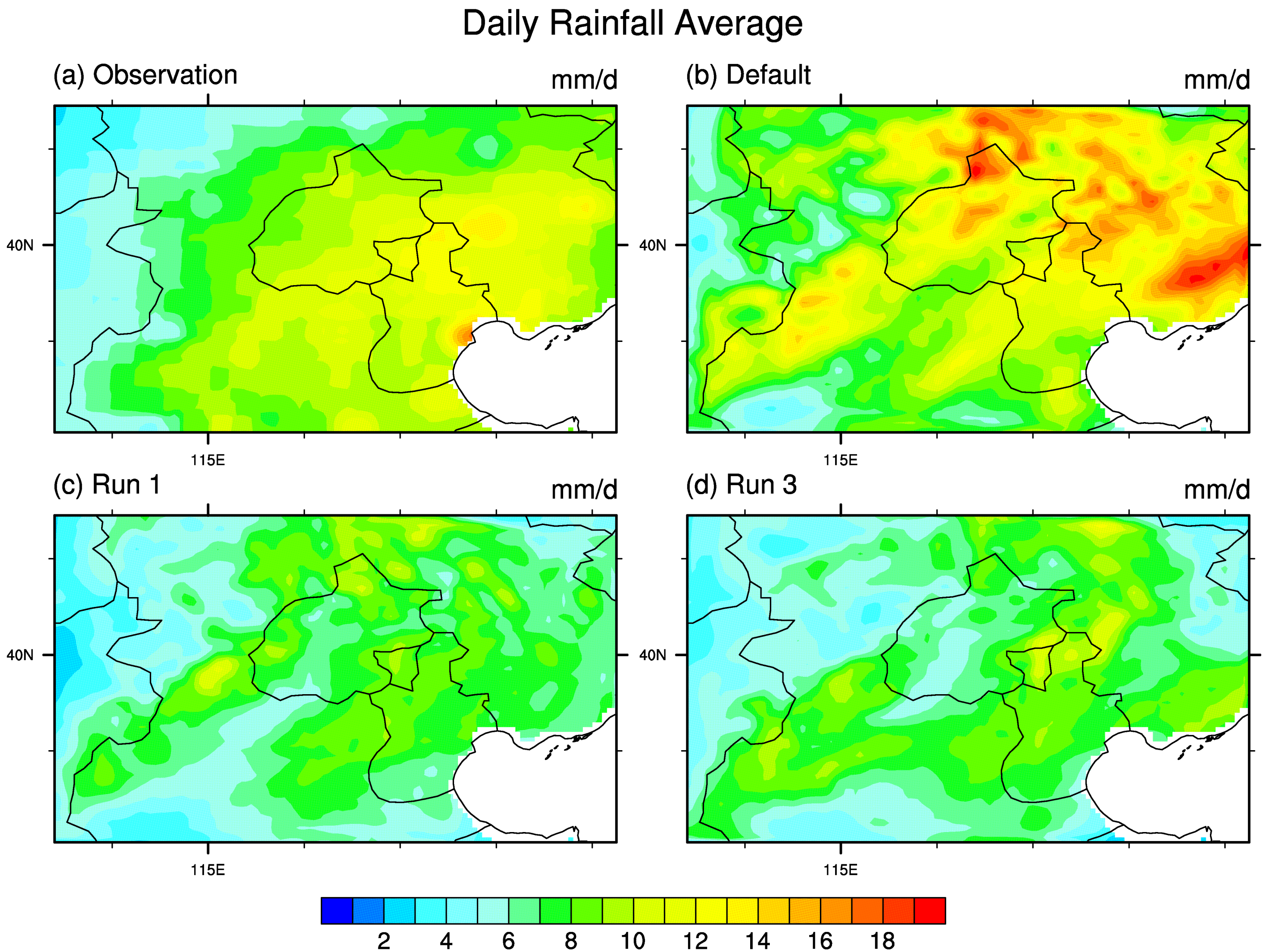
$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Sim_i - Obs_i\right|$$

  - Three Optimization Runs:
    - Optimize P only
    - Optimize SAT only
    - Optimize both P and SAT

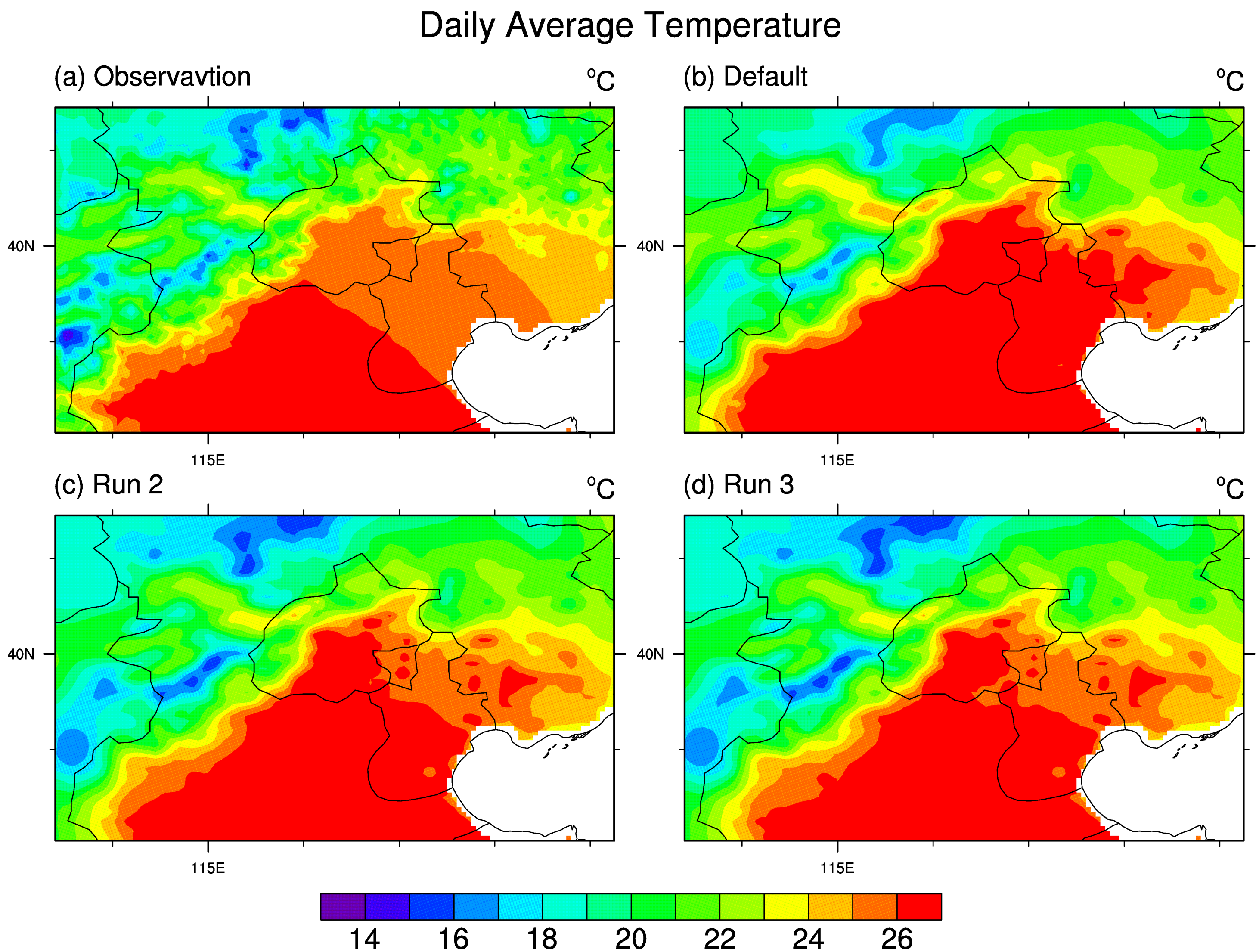
The Optimization Results



Optimization Results: Average Rainfall Values

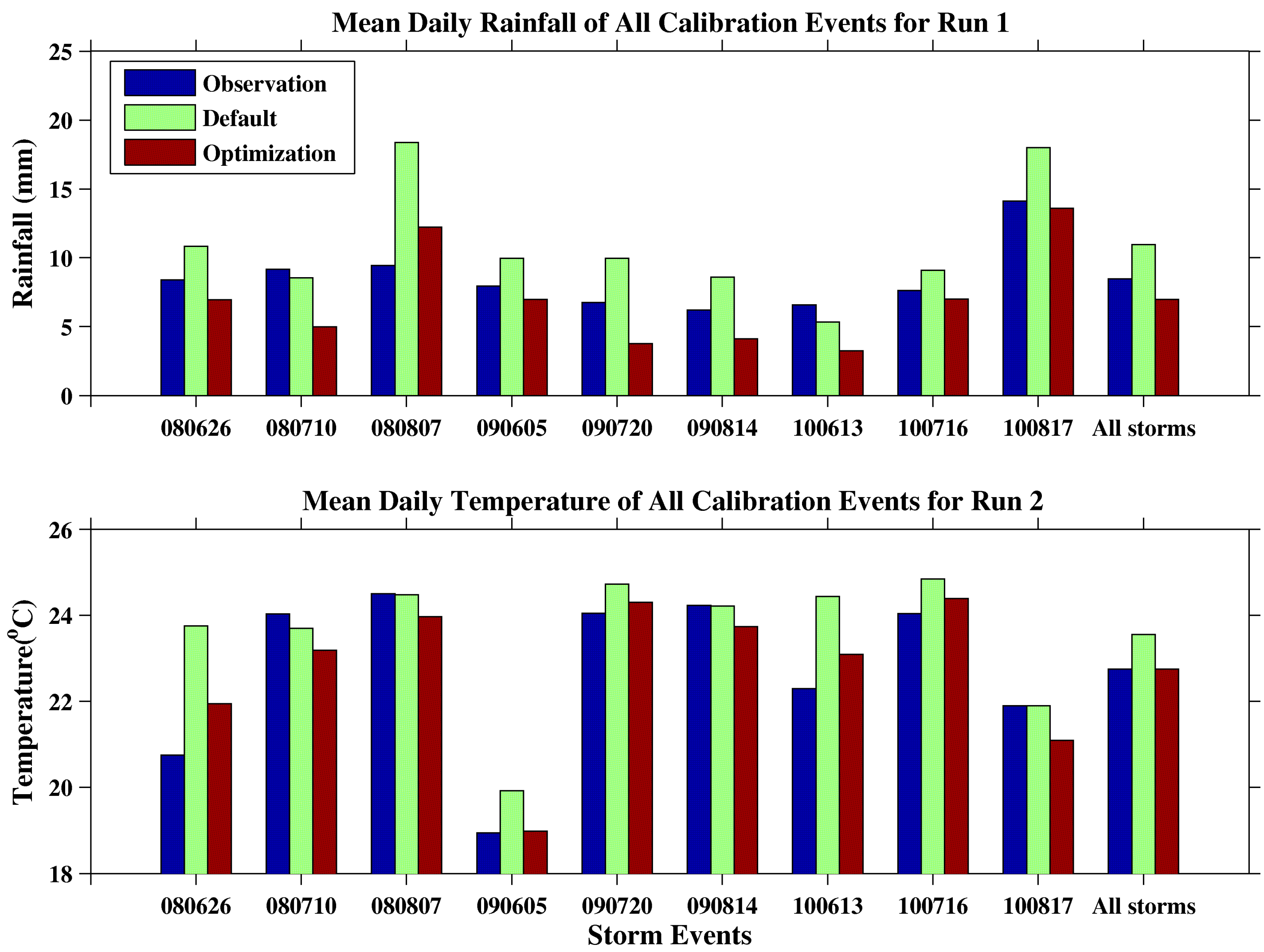


Optimization Results:
Average Surface Air Temperature Values

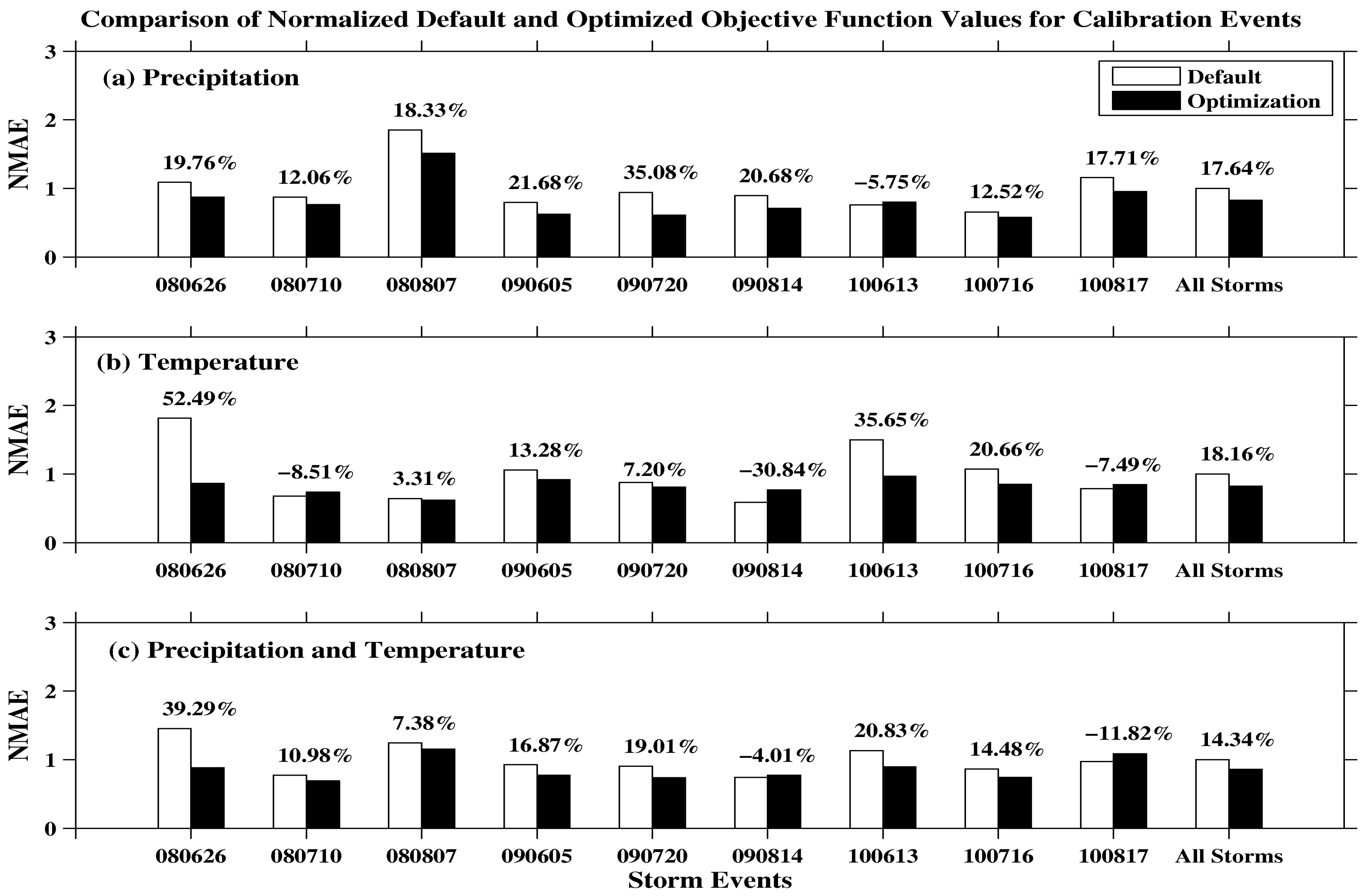


Improvement in Performance Skill

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science

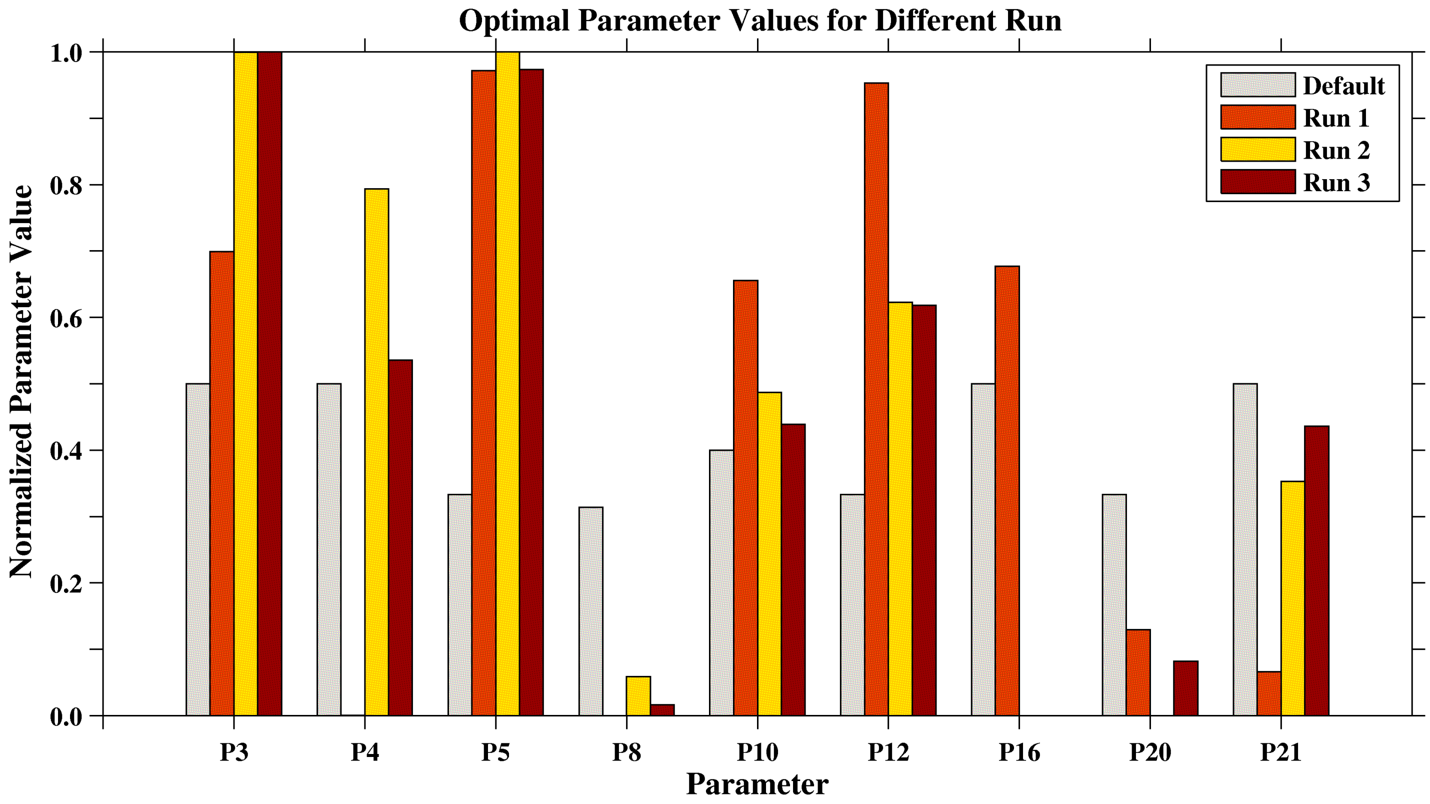


Improvement in Performance Skill



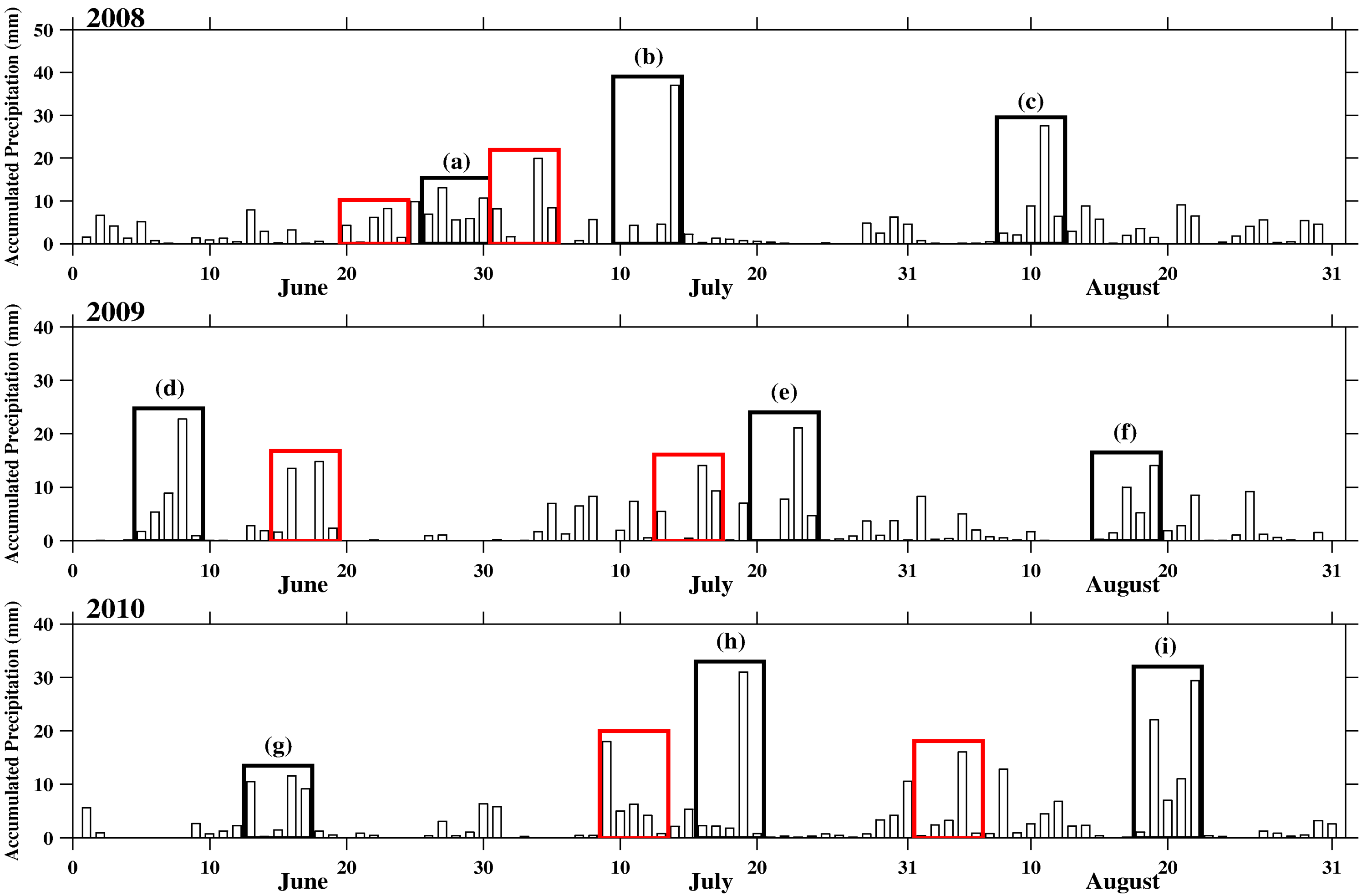
Optimized Parameters

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science



The Validation Events

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science

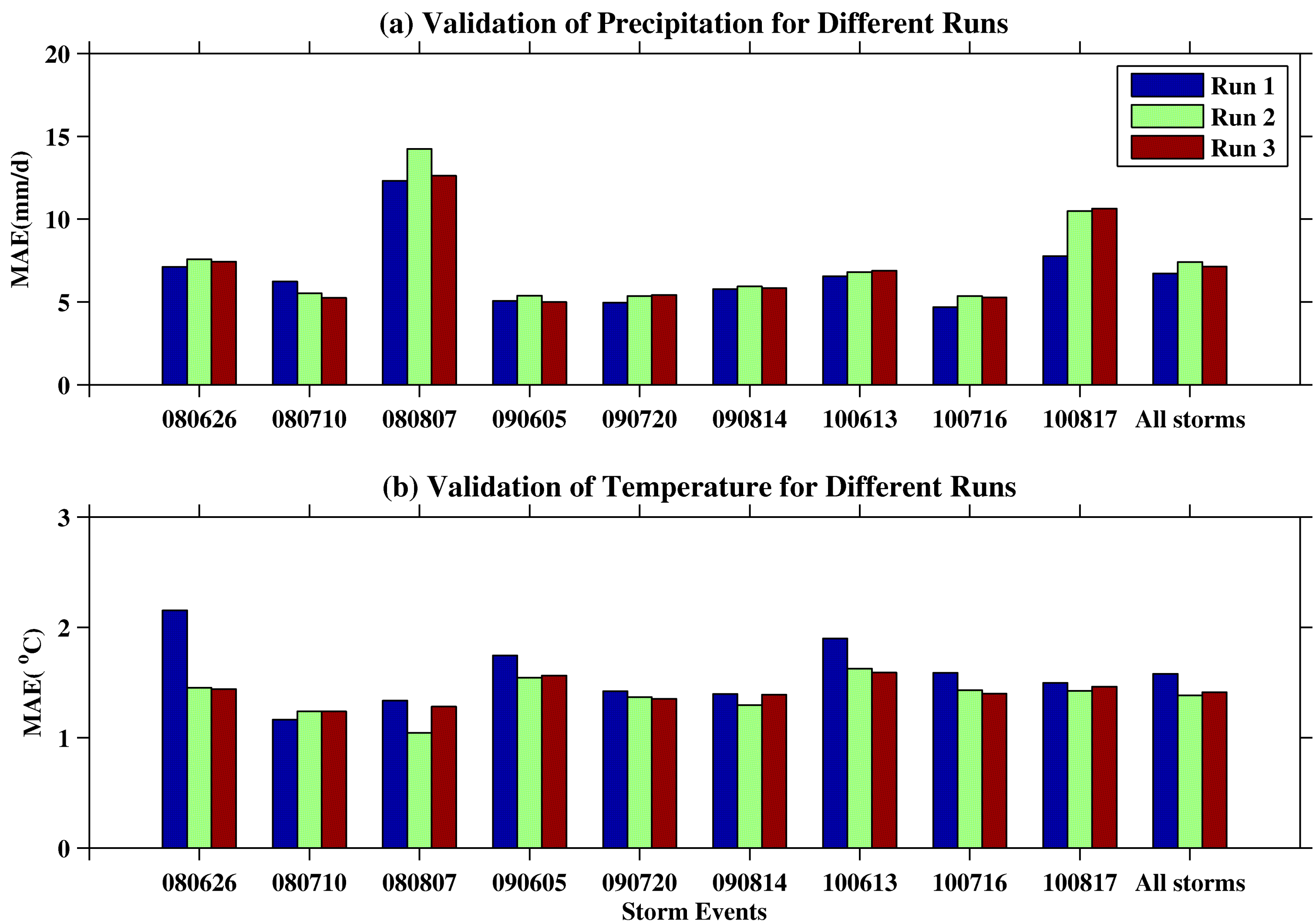


Black box: Calibration

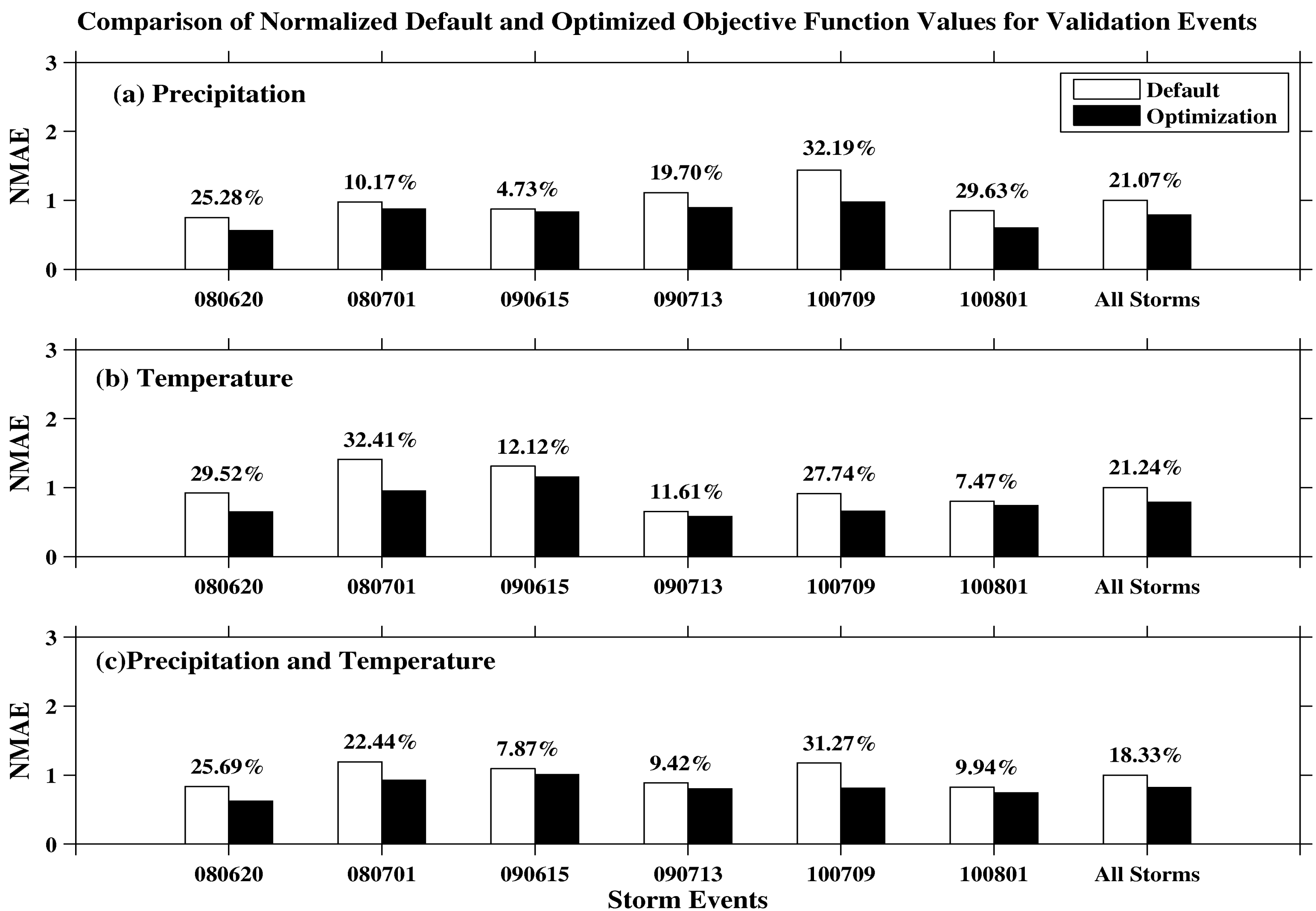
Red box: Validation

Improvement in Validation Events

北京師範大學
Beijing Normal University
全球变化与地球系统科学研究院
College of Global Change and Earth System Science



Improvement in Validation Events



Summary and Discussion of WRF Parametric Uncertainty Research

- 240 model runs are used to identify the most important parameters in WRF that exert great influence on precipitation forecasting skill in Beijing area. 140-230 model runs are needed to optimize the sensitive parameters

- The most sensitive parameters identified are:
  - P3, P4, P5, P8, P12, P16, P18, and P21

- Optimization experiments with the eight most sensitive parameters has improved the model performance by 14-17%

- Validation using independent storm data shows an improved model performance by 18-21%

Overall Summary

- UQ concept explained
- Introduced main UQ techniques:
  - Parameter screening methods
  - Global sensitivity analysis methods
  - Surrogate modeling methods
  - Optimization methods
- Cases studies with CoLM and WRF models

- TAKE HOME MESSAGE: **UQ is an essential tool for improving model performance and automatic optimization of complex geophysical models such as CoLM and WRF is possible**

References

– Li, J.,, et al., 2013. Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis. Hydrol. Earth Syst. Sci. 17, 3279-93.
– Gan Y, et al., 2013. A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. Environmental Modelling & Software 51, 269-85.
– Wang, C., et al., 2014. An evaluation of adaptive surrogate modeling based optimization with two benchmark problems, Environmental Modelling & Software, 60, P167-179
– Gong, W., et al., 2014. Multi-objective parameter optimization of common land model using adaptive surrogate modelling, Hydrol. Earth Syst. Sci., 19, 2409-2425, doi:10.5194/hess-19-2409-2015
– Di, Z., et al., Assessing WRF Model Parameter Sensitivity, Geophysical Research Letter, DOI: 10.1002/2014GL061623
– Gan, Y., X.-Z. Liang, Q. Duan, H. I. Choi, Y. Dai, and H. Wu (2015), Stepwise sensitivity analysis from qualitative to quantitative: Application to the terrestrial hydrological modeling of a Conjunctive Surface-Subsurface Process (CSSP) land surface model, J. Adv. Model. Earth Syst., 07, doi:10.1002/2014MS000406.
– Quan, J., Z. Di, W. Gong, C. Wang, Q. Duan, Y. Gan, A. Ye, C. Miao, (2015), A Multi-objective Evaluation of Parametric Sensitivities of Different Meteorological Forecasts by WRF Model, submitted to Mon. Wea. Rev.

Questions ?