Land Surface Model Parameter Sensitivity Analysis and Estimation in the Amazon Basin

Rafael Rosolem

W. J. Shuttleworth, H. V. Gupta, L. G. G. de Goncalves, X. Zeng, and N. Restrepo-Coupe

LBA-DMIP Workshop
April 19, 2011

THE UNIVERSITY OF ARIZONA®

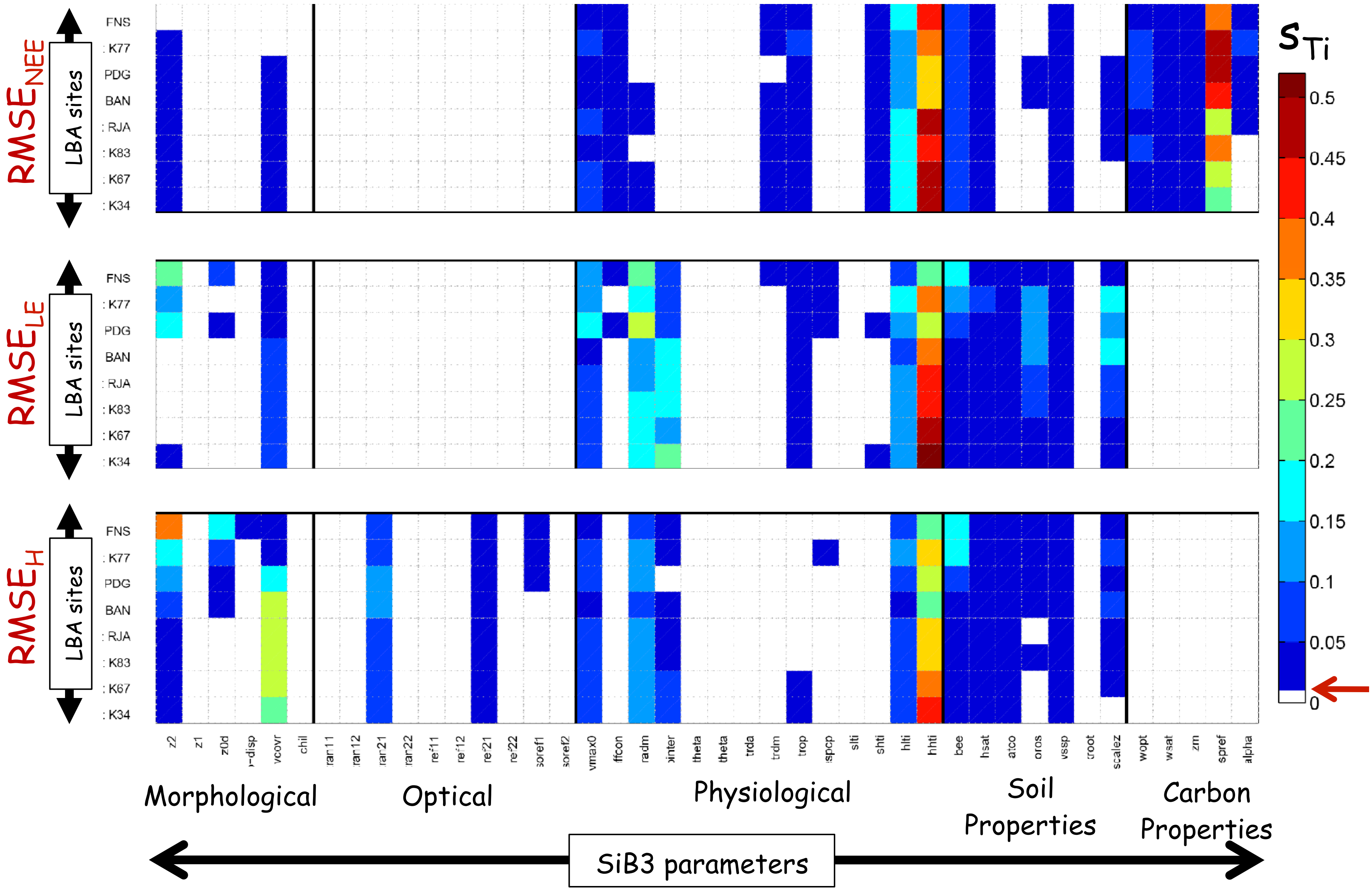
Outline

Paper #1 (Parameter Sensitivity Analysis): Assess parameter sensitivity analysis across all sites using SiB3' model using a newly proposed approach based on variance-based Sobol method accounting for full multi-output nature of land surface models

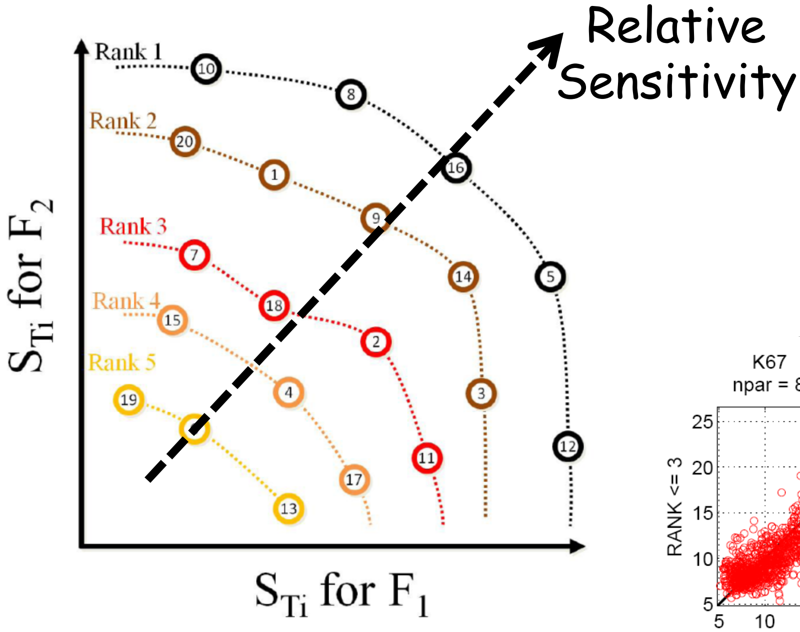
Paper #2 (Parameter Estimation): Substantial improvement (10-30% reduction of RMSE) in energy, water and CO_2 flux simulations after calibration but also identifying the sources of uncertainty

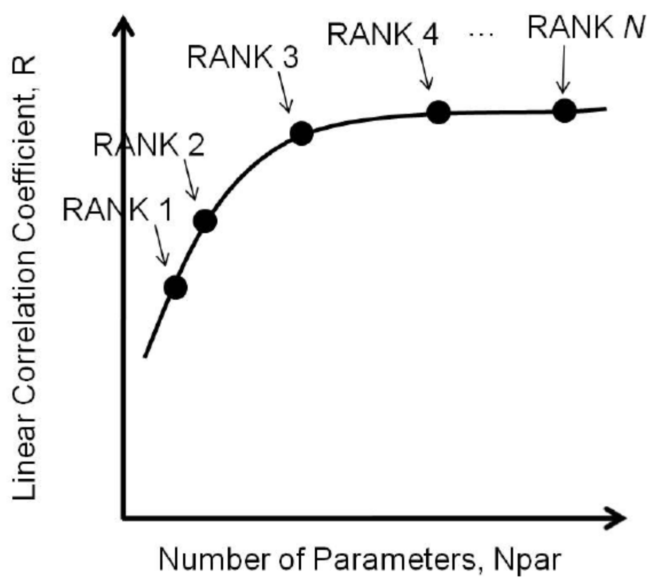
How could these ideas be applied to LBA-DMIP models?

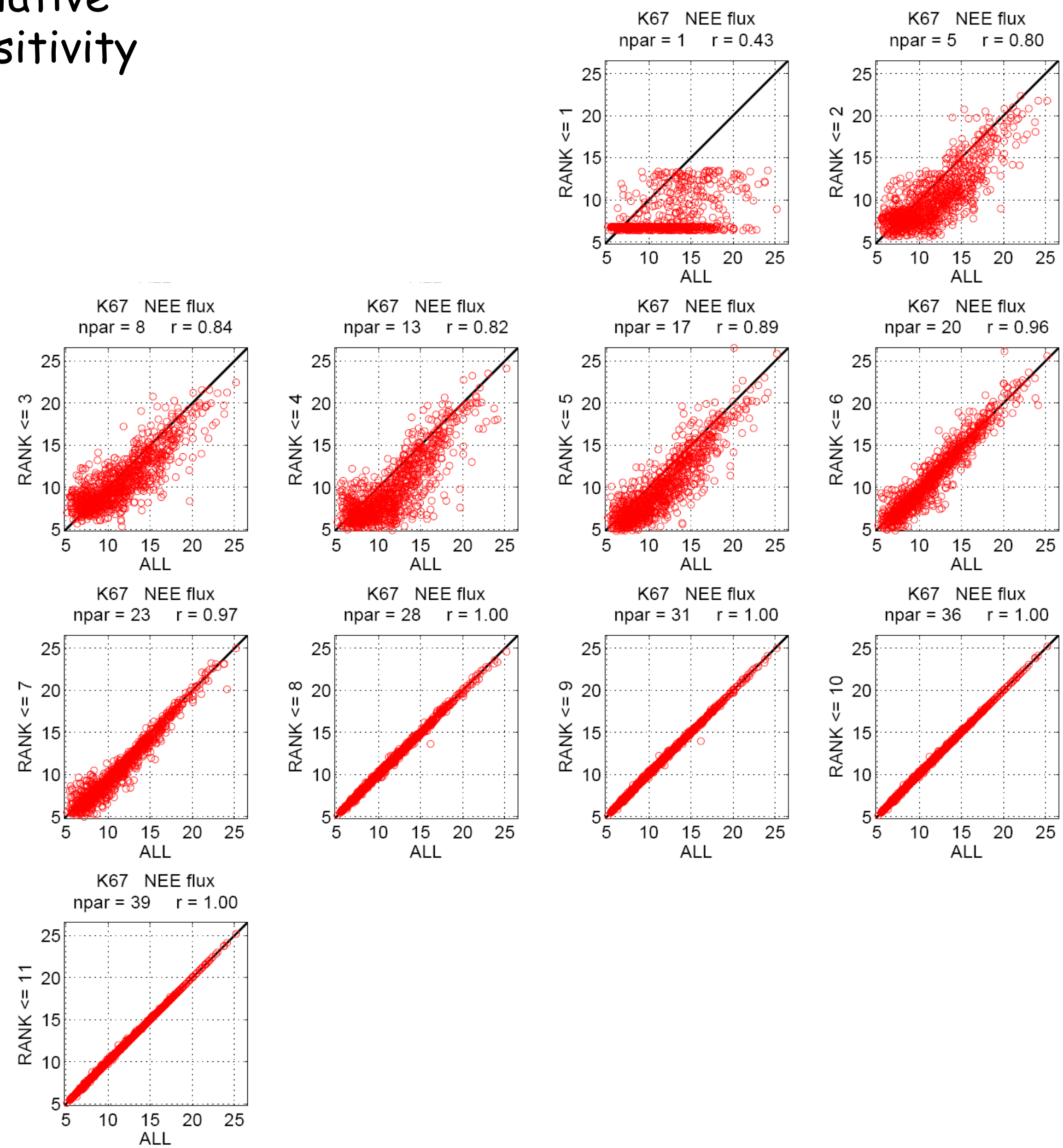
Sobol application: Conventional approach



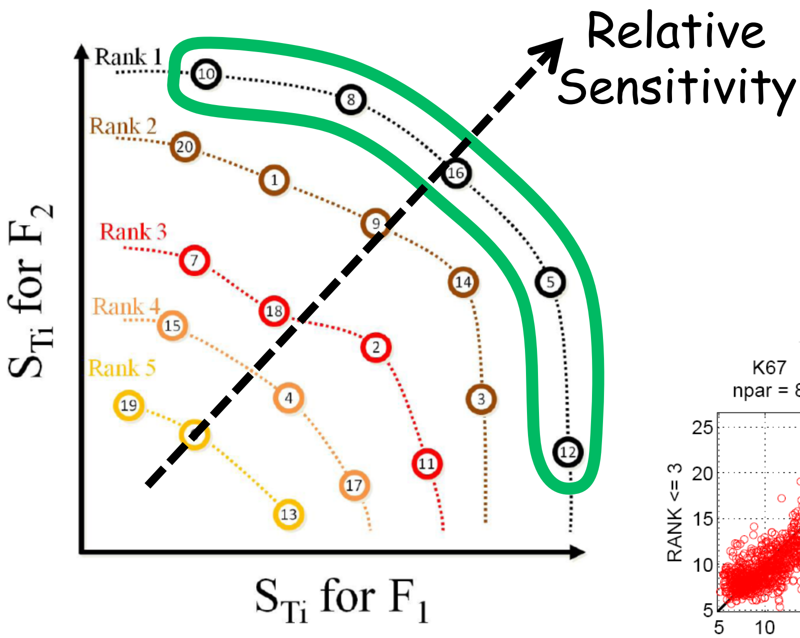
Sobol application: New multi-objective approach

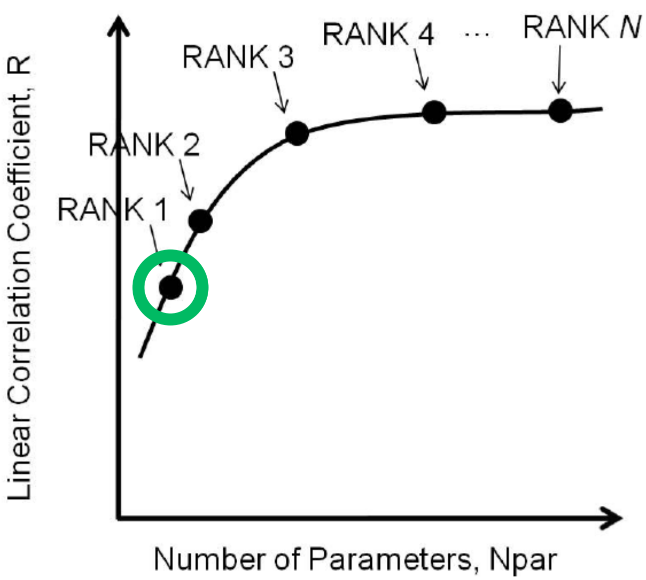


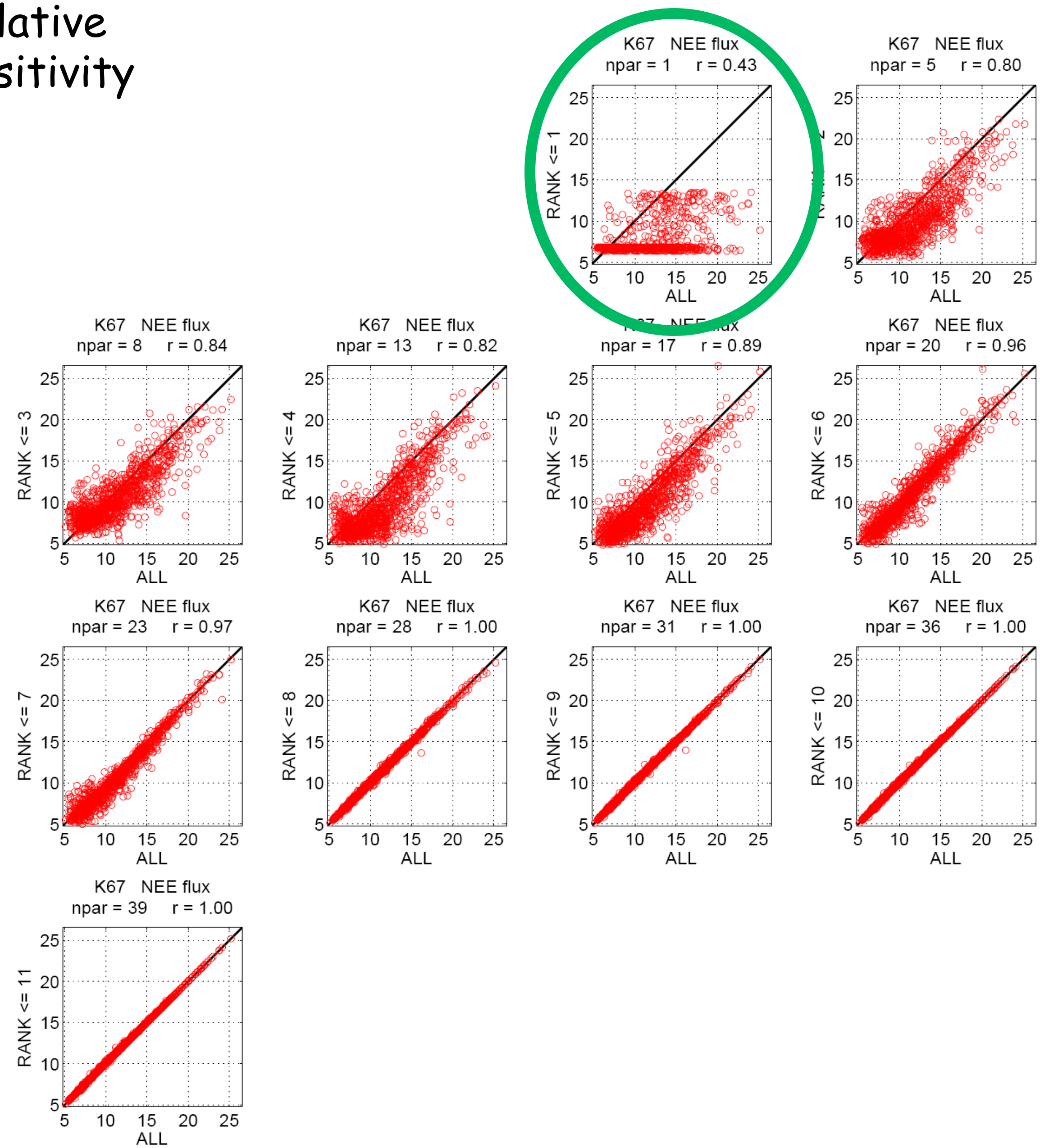




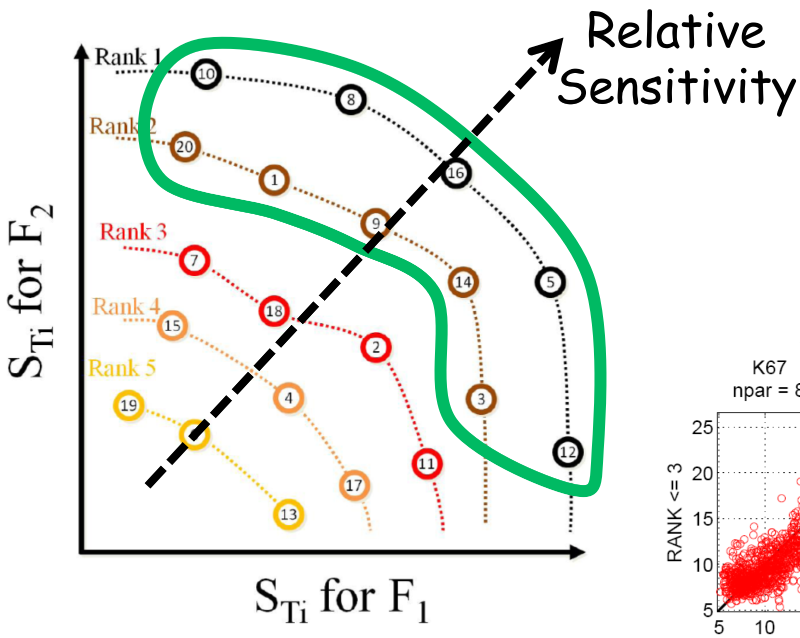
Sobol application: New multi-objective approach

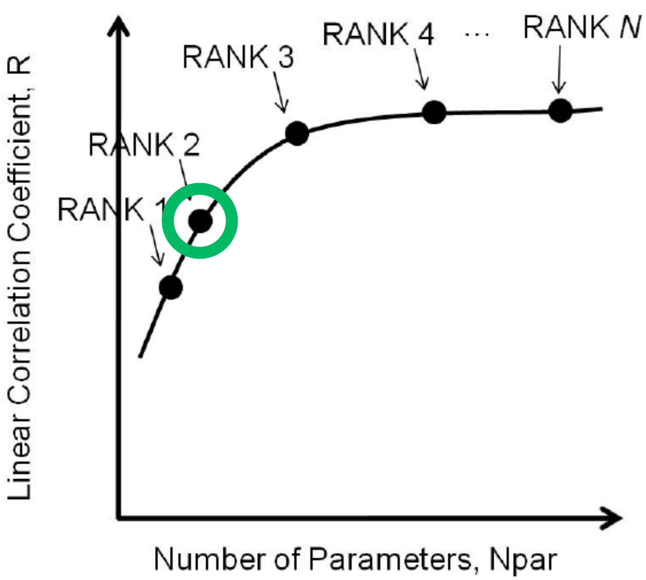


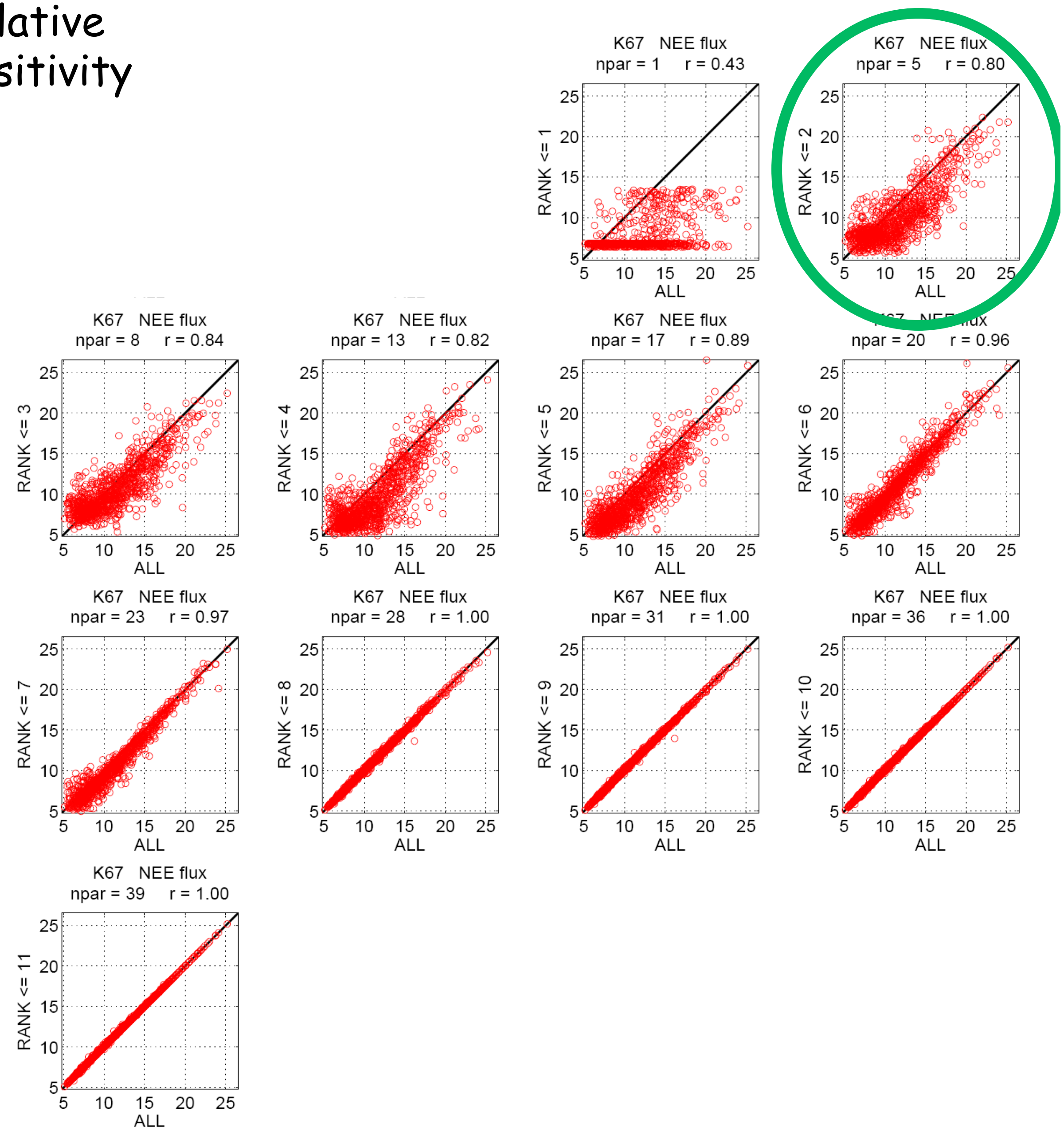




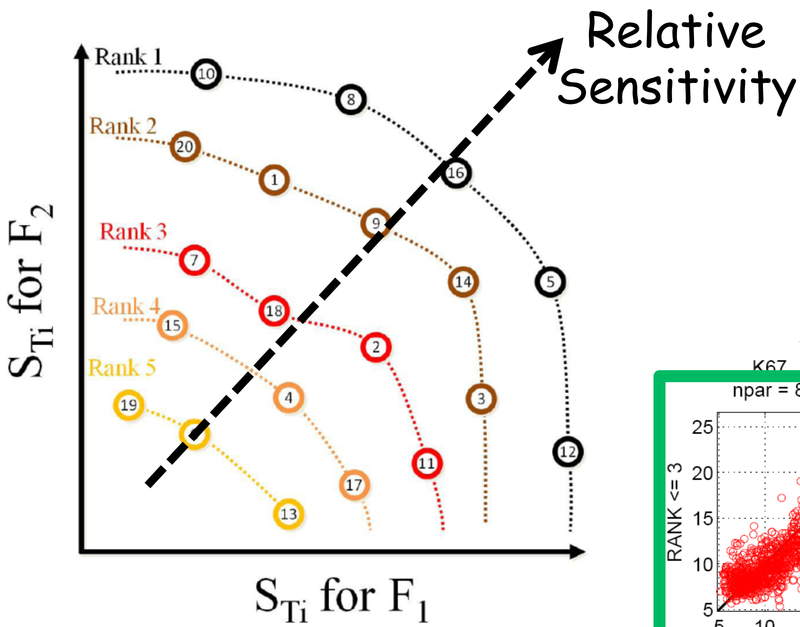
Sobol application: New multi-objective approach

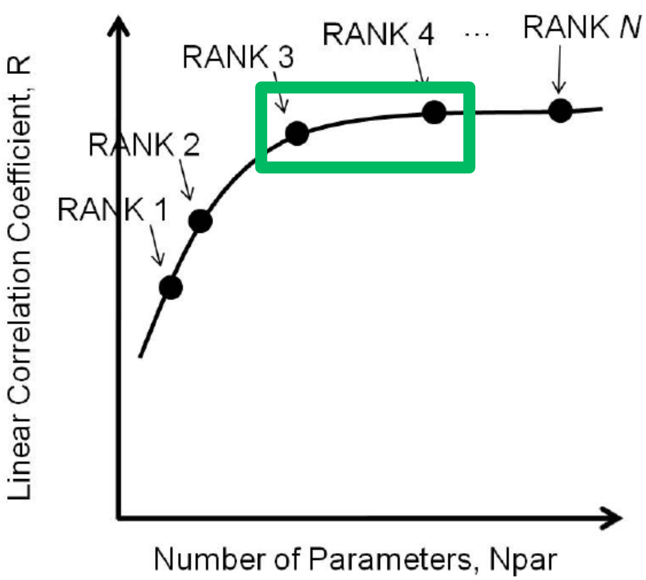


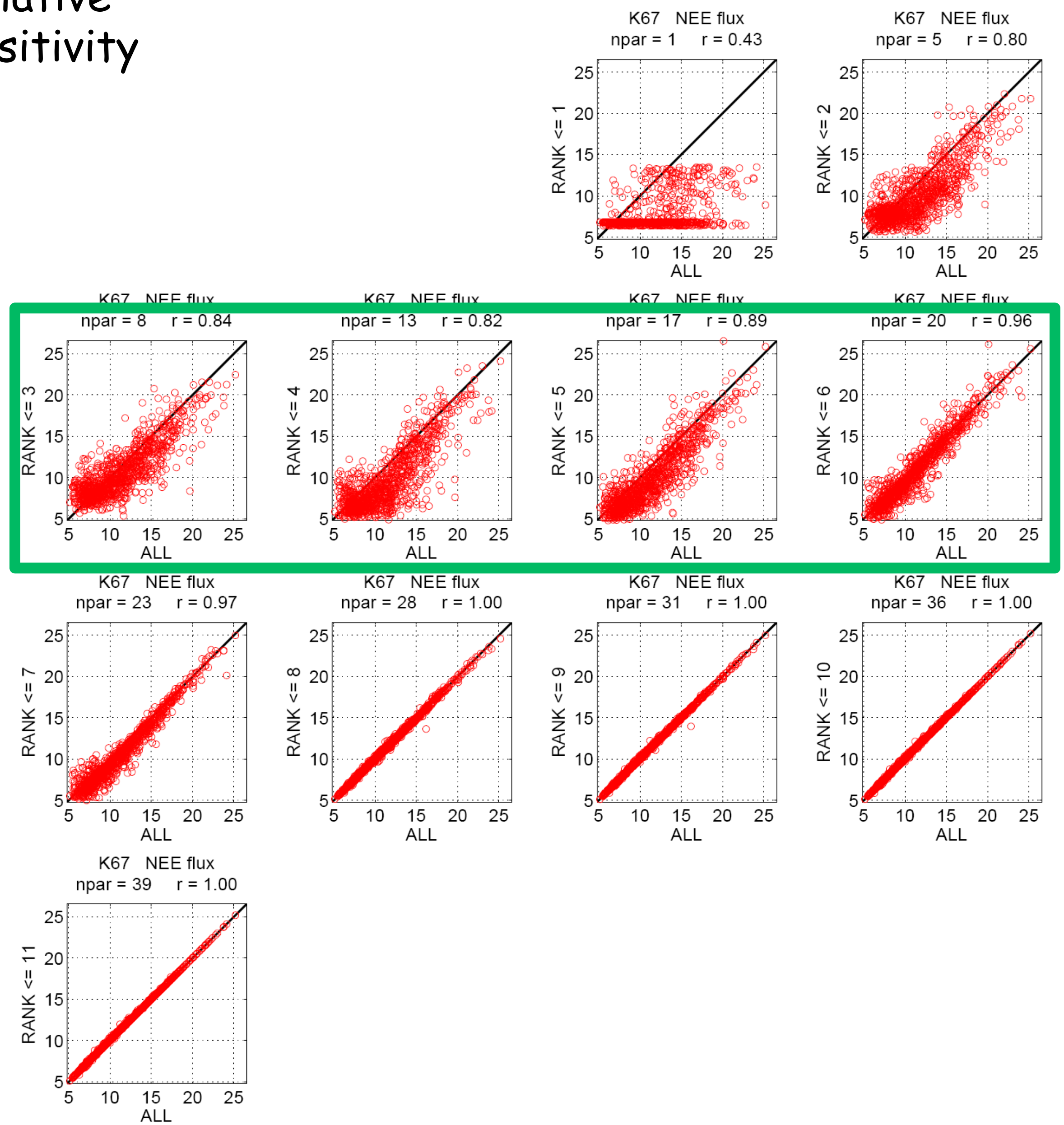




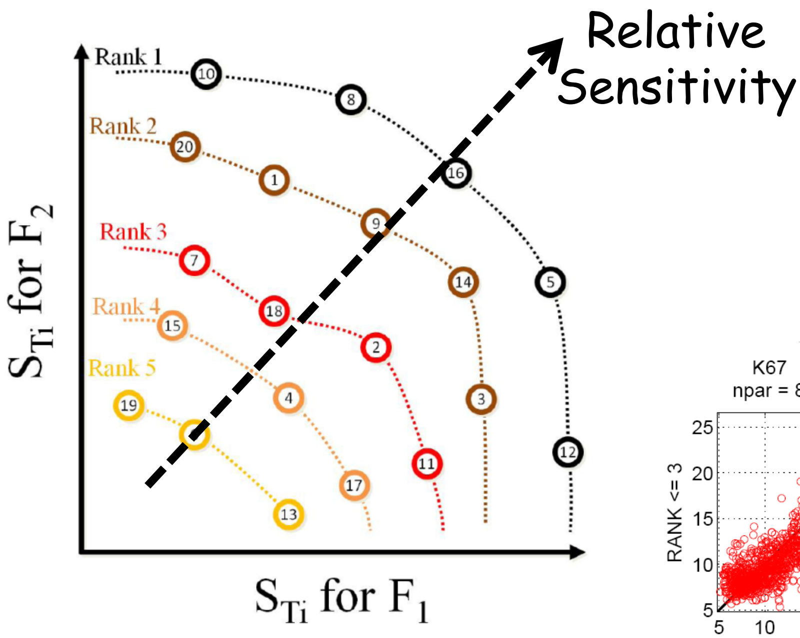
Sobol application: New multi-objective approach

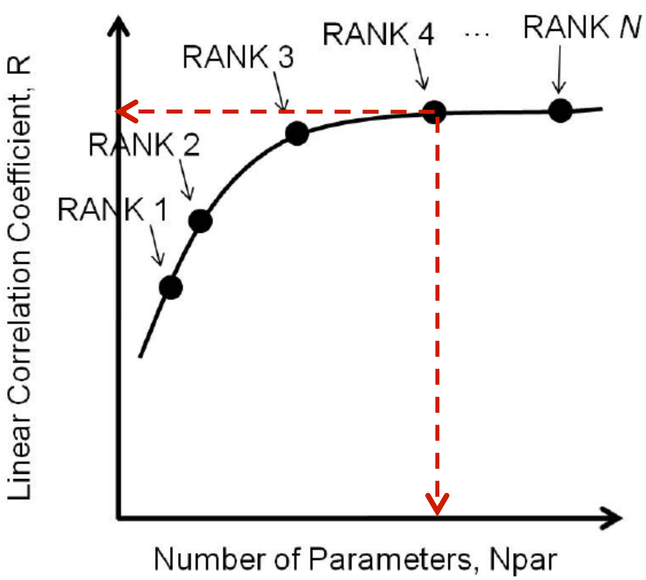


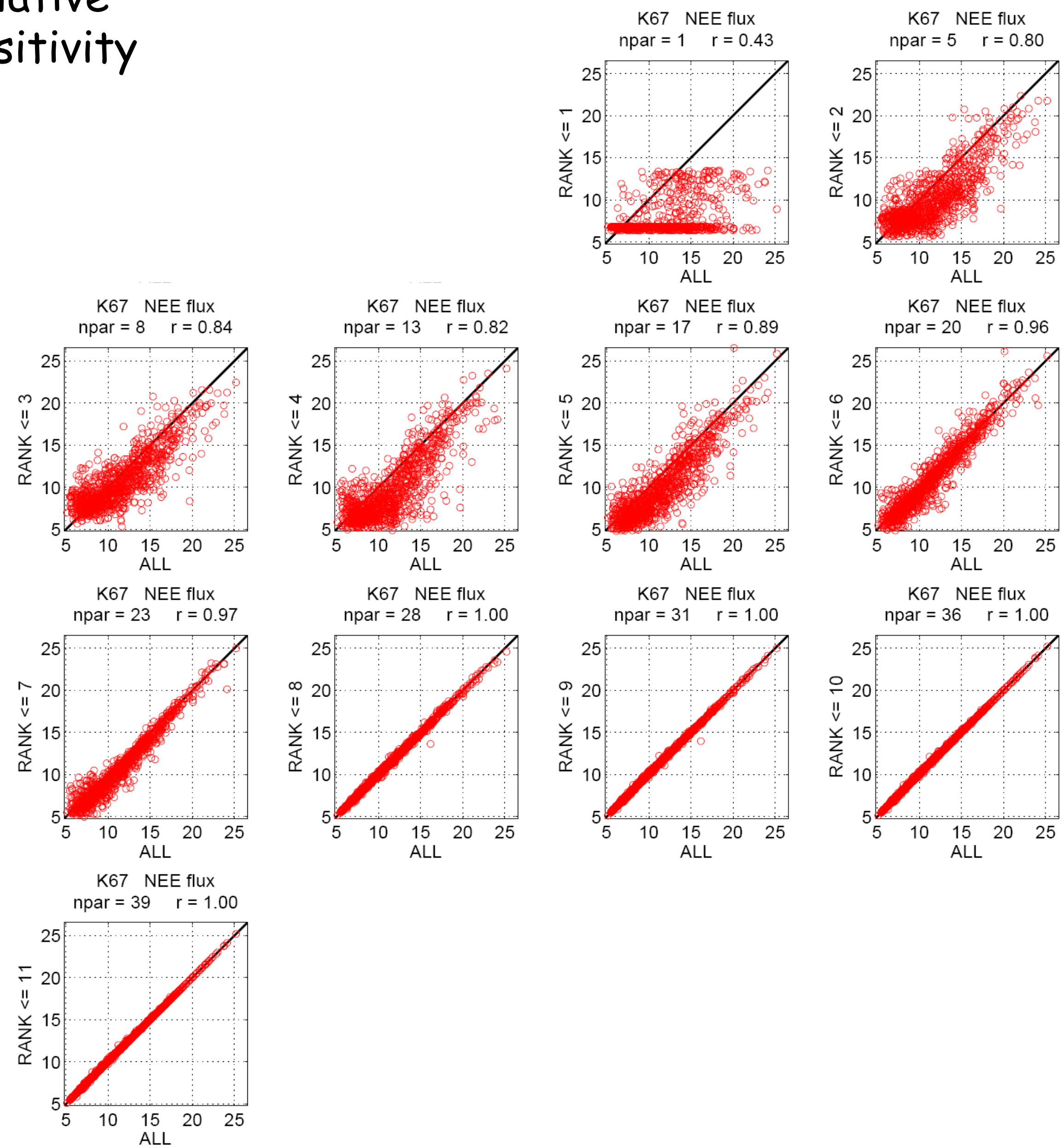




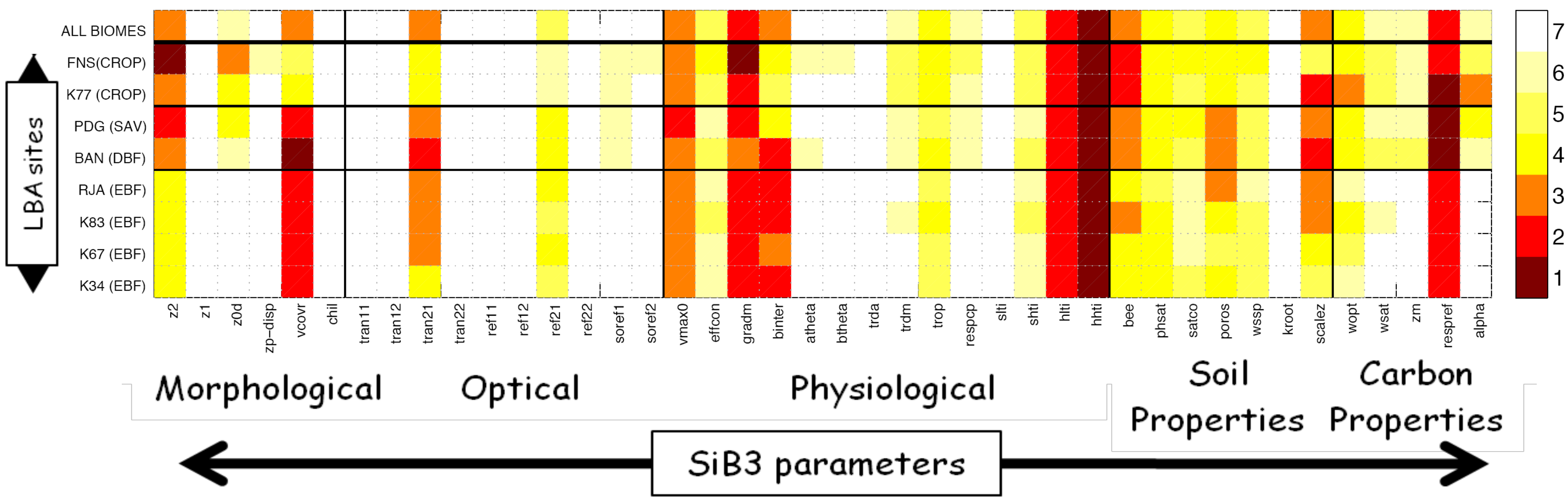
Sobol application: New multi-objective approach







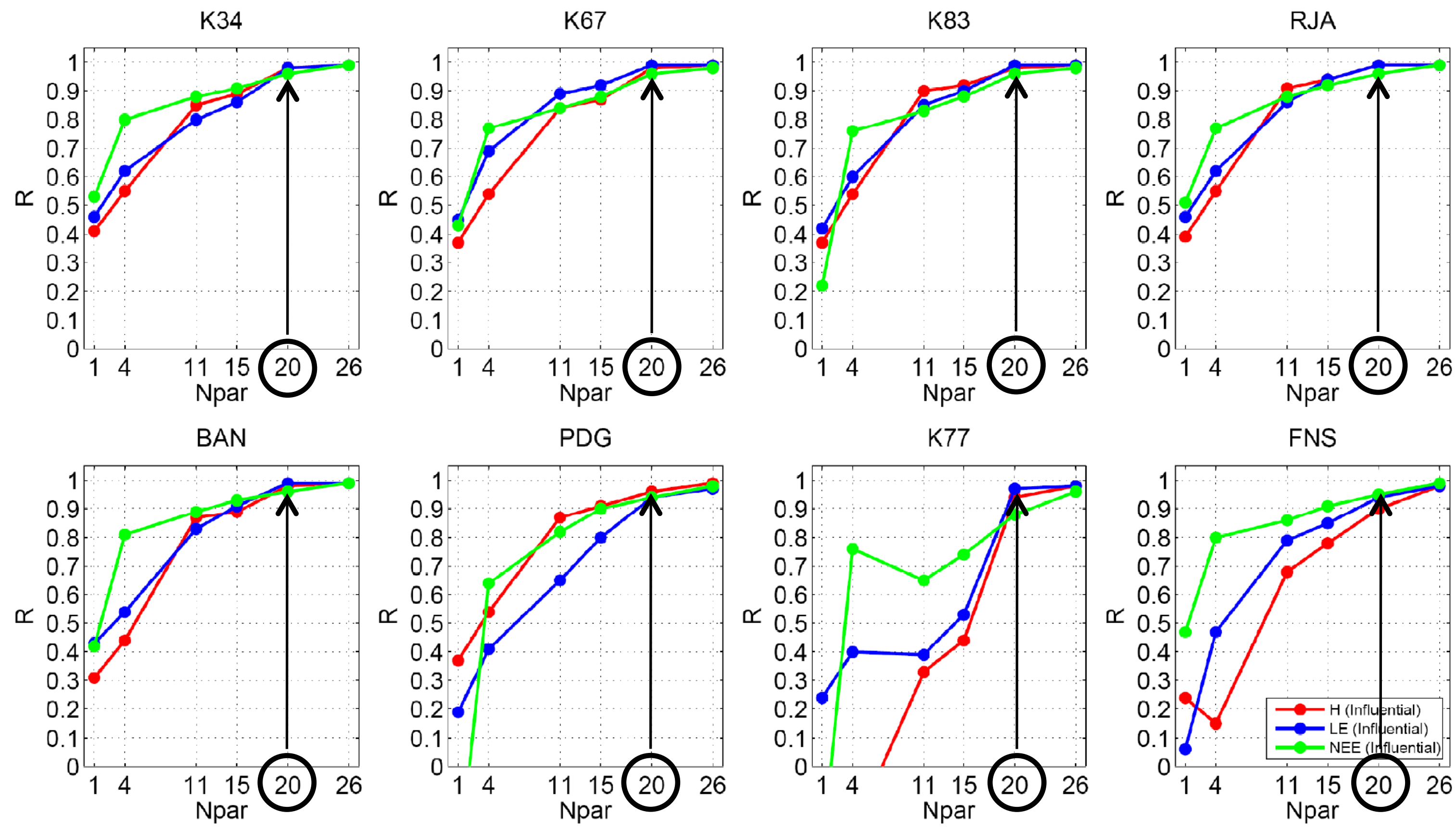
Sobol application: New multi-objective approach



The lower the rank (darker), the more sensitive a parameter is for a given site (taking into account the contribution from all objective functions; i.e. fluxes).

The top row is the weighted average results based on distinct biome types.

Common set of sensitive parameters to all biome types "Influential Ranks" (H, LE, and NEE fluxes)



Influential Rank up to 5 → 20 parameters (~50%)

Physiological Properties = 8 parameters (*vmax0*, *effcon*, *gradm*, *binter*, *trop*, *shti*, *hlti*, *hhti*)
Soil Properties = 6 parameters (*bee*, *phsat*, *satco*, *poros*, *wssp* and *scalez*)
Morphological = 2 parameters (*z2* and *vcovr*)
Optical = 2 parameters (*tran21* and *ref21*)
Soil Respiration = 2 parameters (*wopt* and *respref*)

A Multi-Algorithm Genetically Adaptive Multiobjective method - AMALGAM (Vrugt and Robinson, 2007)

PNAS PNAS PNAS PNAS

Improved evolutionary optimization from genetically adaptive multimethod search

Jasper A. Vrugt* and Bruce A. Robinson

Los Alamos National Laboratory, EES-6, Mail Stop T003, Los Alamos, NM 87545

Communicated by Stirling A. Colgate, Los Alamos National Laboratory, Los Alamos, NM, November 27, 2006 (received for review March 29, 2006)

In the last few decades, evolutionary algorithms have emerged as a revolutionary approach for solving search and optimization problems involving multiple conflicting objectives. Beyond their ability to search intractably large spaces for multiple solutions, these algorithms are able to maintain a diverse population of solutions and exploit similarities of solutions by recombination. However, existing theory and numerical experiments have demonstrated that it is impossible to develop a single algorithm for population evolution that is always efficient for a diverse set of optimization problems. Here we show that significant improvements in the efficiency of evolutionary search can be achieved by running multiple optimization algorithms simultaneously using new concepts of global information sharing and genetically adaptive offspring creation. We call this approach a multialgorithm, genetically adaptive multiobjective, or AMALGAM, method, to evoke the image of a procedure that merges the strengths of different optimization algorithms. Benchmark results using a set of well known multiobjective test problems show that AMALGAM approaches a factor of 10 improvement over current optimization algorithms for the more complex, higher dimensional problems. The AMALGAM method provides new opportunities for solving previously intractable optimization problems.

evolutionary search | multiple objectives | optimization problems | Pareto front

solutions by recombination. These attributes lead to efficient convergence to the Pareto-optimal front in a single optimization run (13). Of these, the nondominated sorted genetic algorithm II (NSGA-II) (14) has received the most attention because of its simplicity and demonstrated superiority over other methods.

Although the multiobjective optimization problem has been studied quite extensively, current available evolutionary algorithms typically implement a single algorithm for population evolution. Reliance on a single biological model of natural

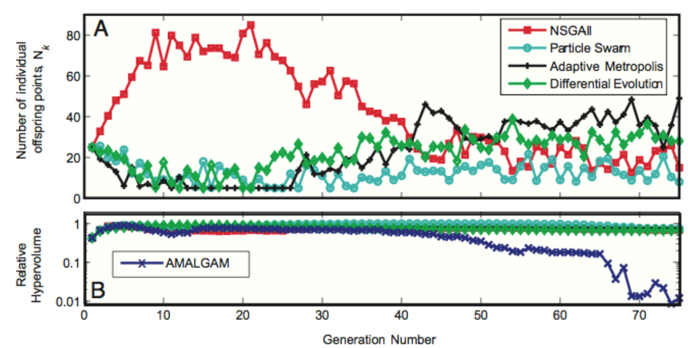
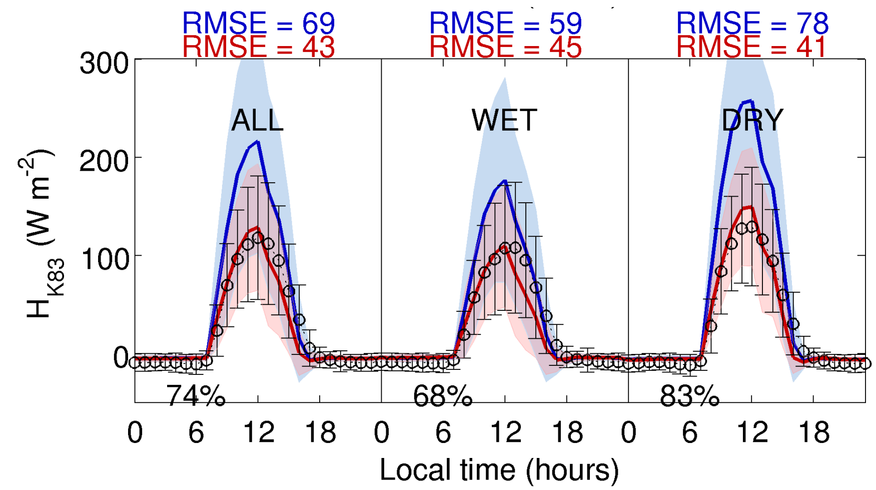


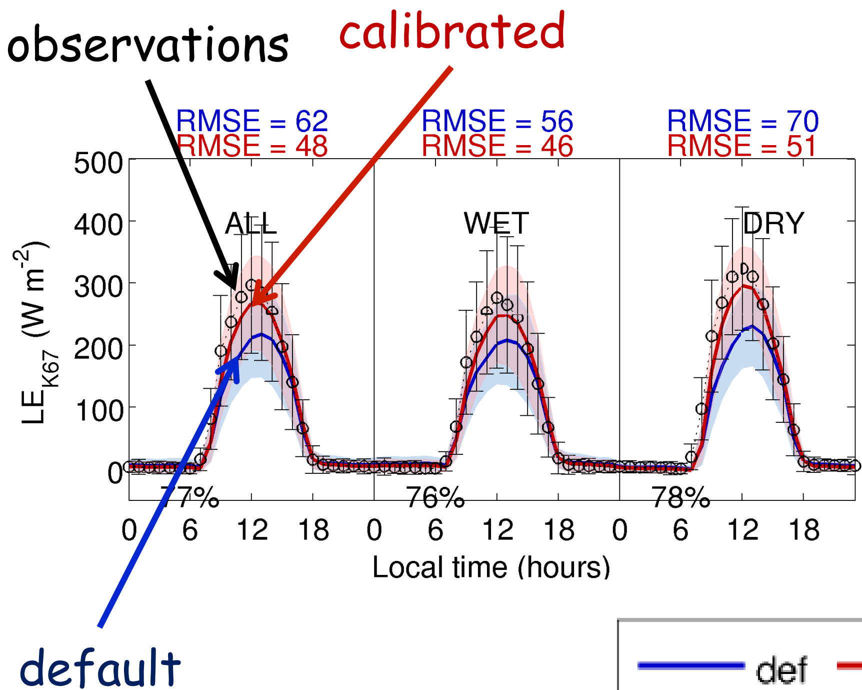
Fig. 2. Illustration of the concept of self-adaptive offspring creation. (*A*) Evolution of the number of offspring points generated with the NSGA-II (squares), PSO (circles), AMS (+), and DE (diamonds) algorithms within AMALGAM's multimethod search as function of generation number for test problem ZDT4. (*B*) The hypervolume convergence metric for AMALGAM and each algorithm used individually. These results illustrate the utility of individual search algorithms during different stages of the optimization, and provide numerical evidence of Wolpert and Macready's "No Free Lunch" theorem, showing that it is impossible to develop a single search algorithm that will always be superior to any other algorithm (16).

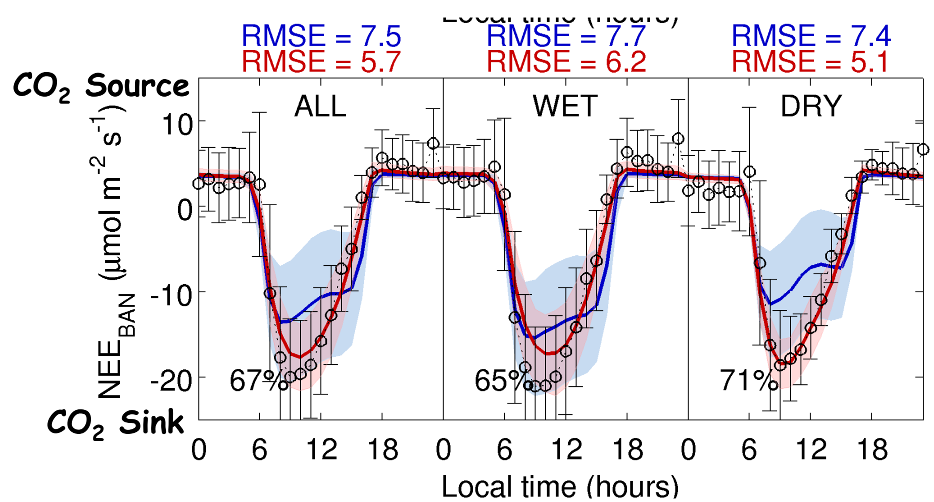
Improvement RMSE for simulated fluxes at diurnal-scale

We select a **'compromise' solution**, for which average normalized RMSE of errors in matching all three fluxes is minimum

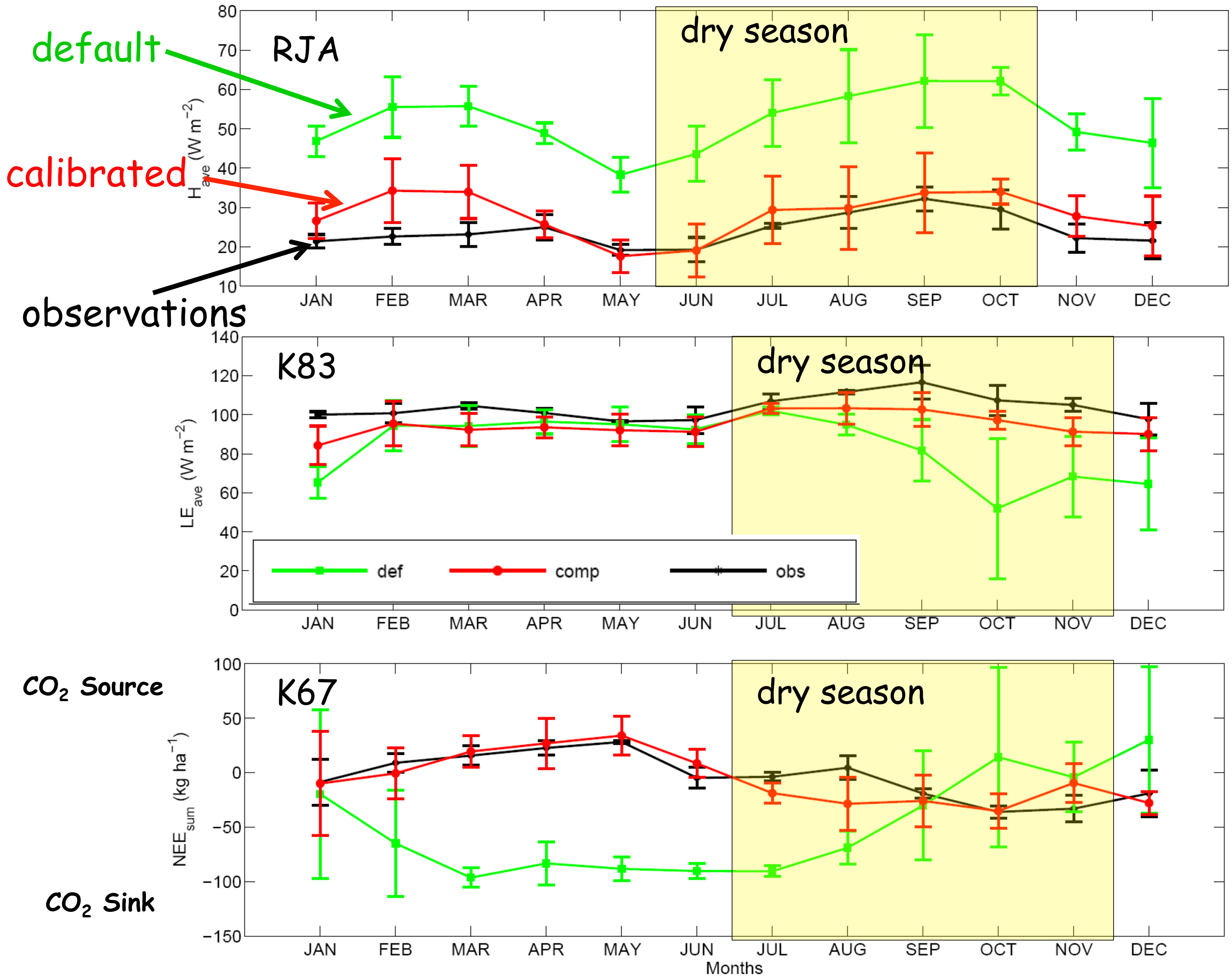
This solution provides a balanced (**equally weighted**) reproduction of all three **fluxes**.







Improvement RMSE for simulated fluxes at seasonal-scale

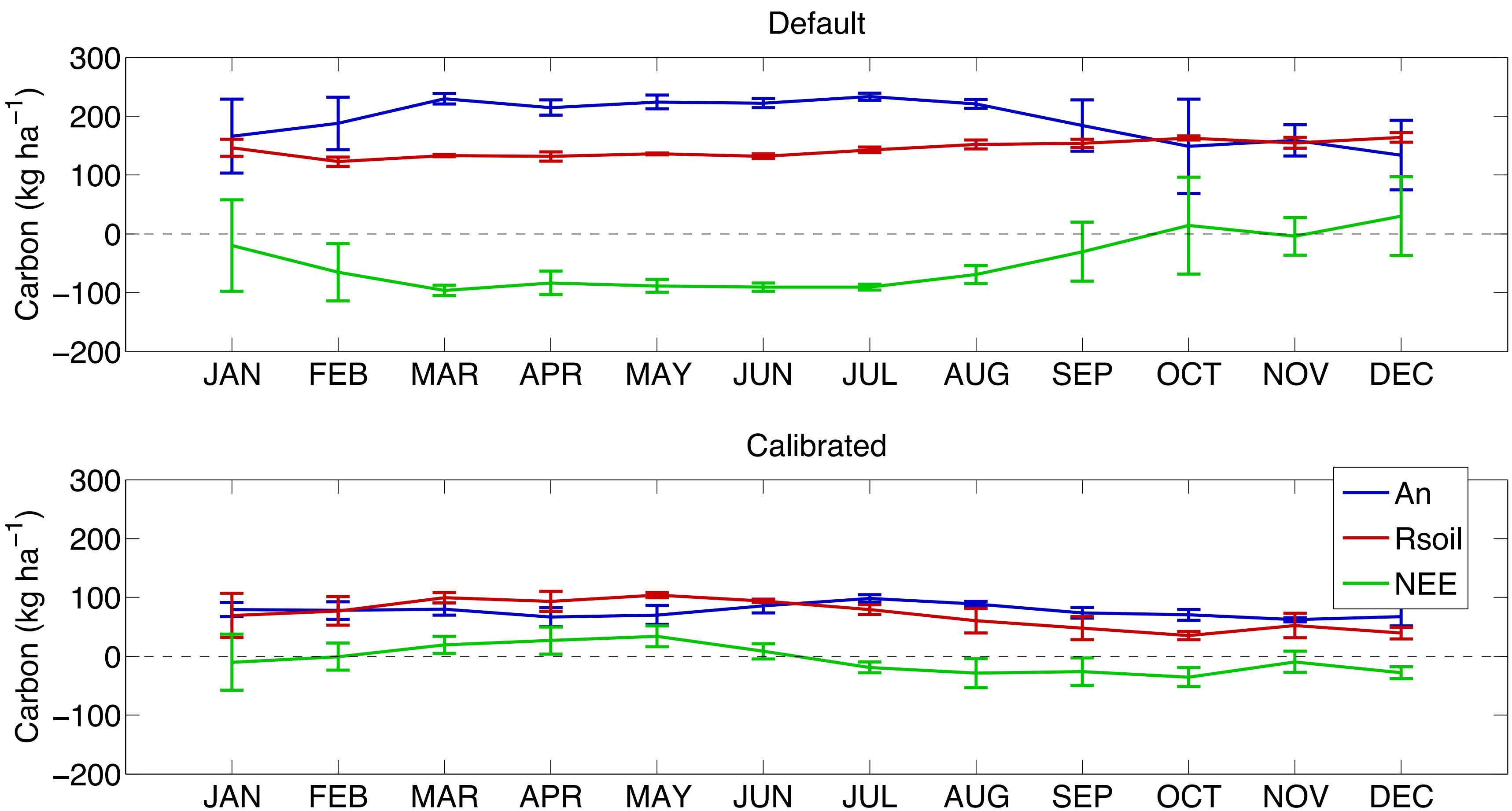


Improvement:
ALL = 41%
WET = 40%
DRY = 43%

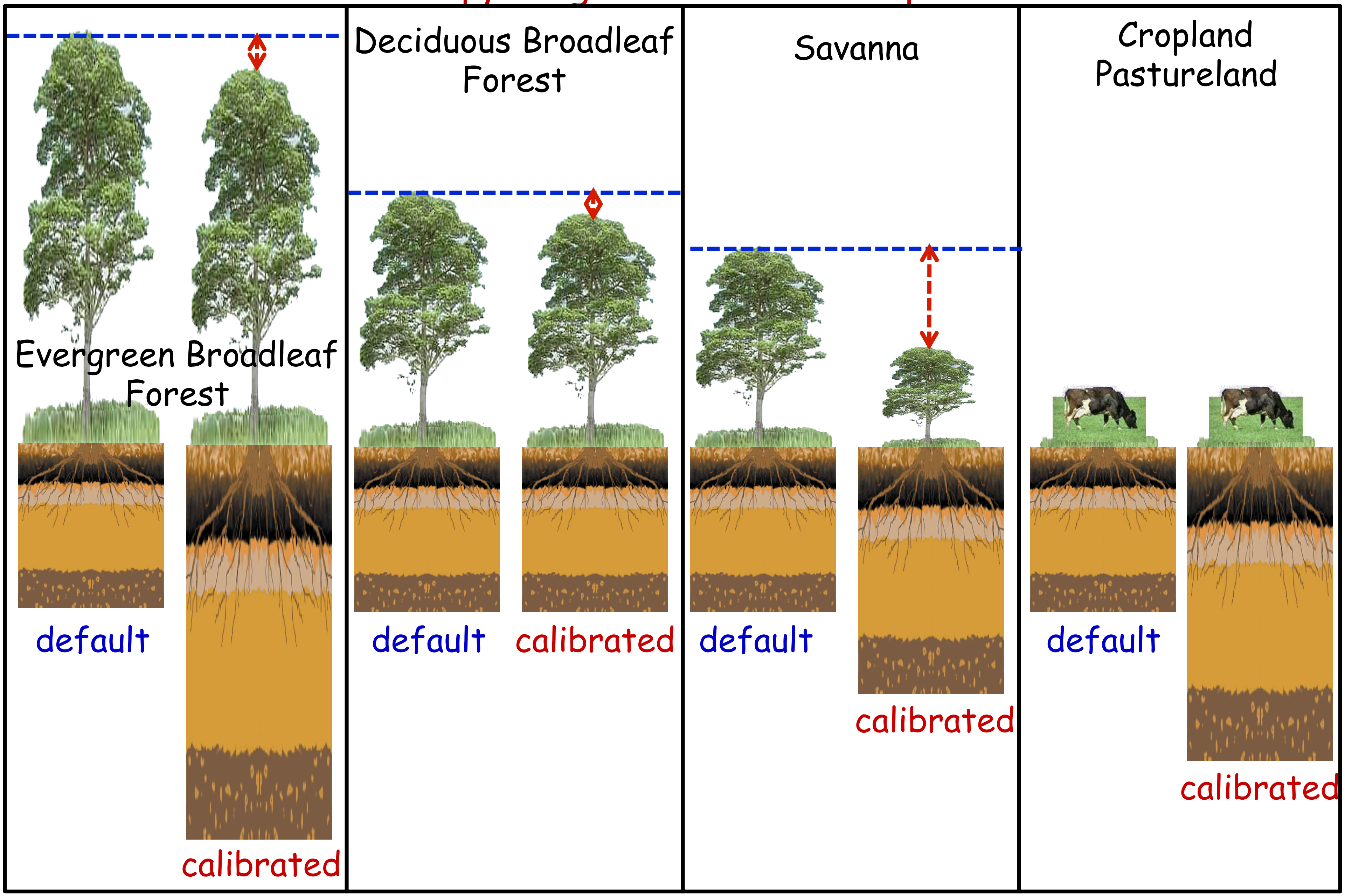
Improvement:
ALL = 14%
WET = 7%
DRY = 21%

Improvement:
ALL = 30%
WET = 31%
DRY = 28%

Improvement RMSE for simulated fluxes at seasonal-scale

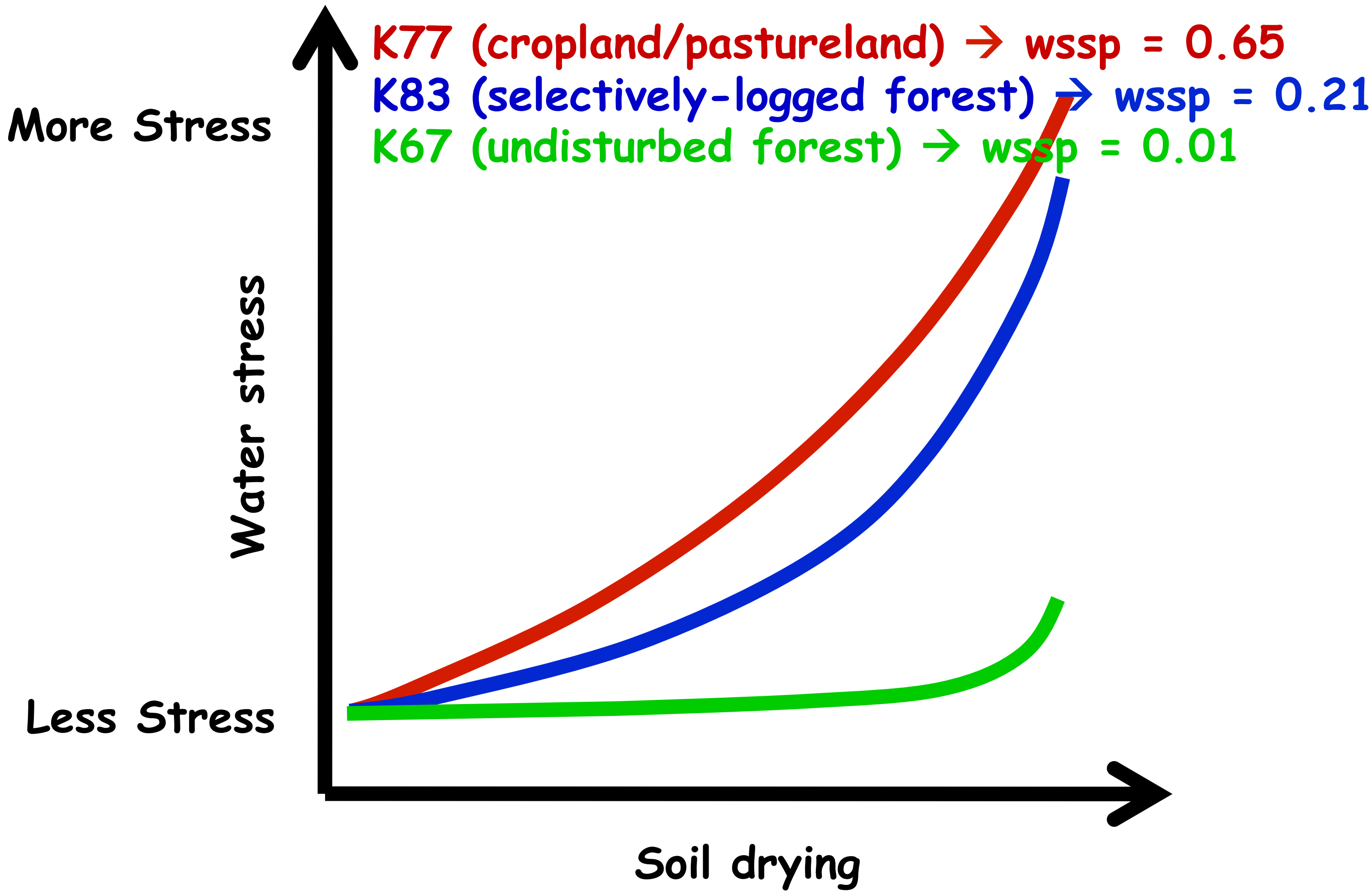


Consistency of calibrated parameters: Canopy Height & Total Soil Depth



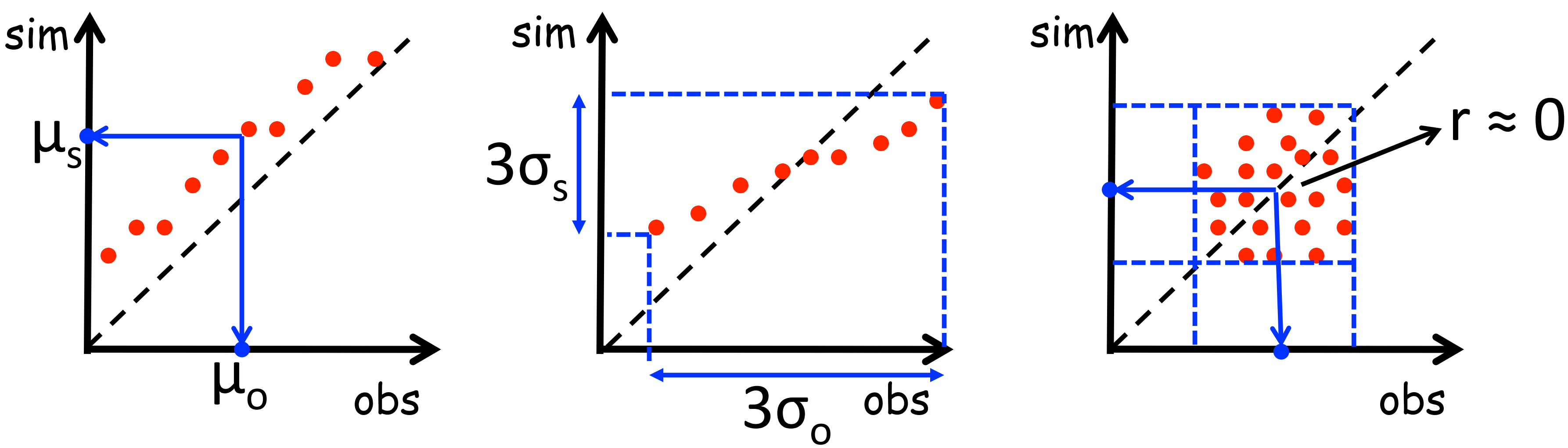
Consistency of calibrated parameters: Soil Water Stress Curvature Parameter

SiB3 originally assumes this to be a constant (wssp = 0.20)



Mean-Squared-Error (MSE) Decomposition (Gupta et al. 2009)

$$MSE = \underbrace{\left(\mu_s - \mu_o\right)^2}_{\substack{\text{Error in} \\ \text{Signal Mean}}} + \underbrace{\left(\sigma_s - \sigma_o\right)^2}_{\substack{\text{Error in} \\ \text{Signal Variability}}} + \underbrace{2 \cdot \sigma_s \cdot \sigma_o \cdot \left(1 - r\right)}_{\substack{\text{Error in} \\ \text{Timing and shape}}}$$



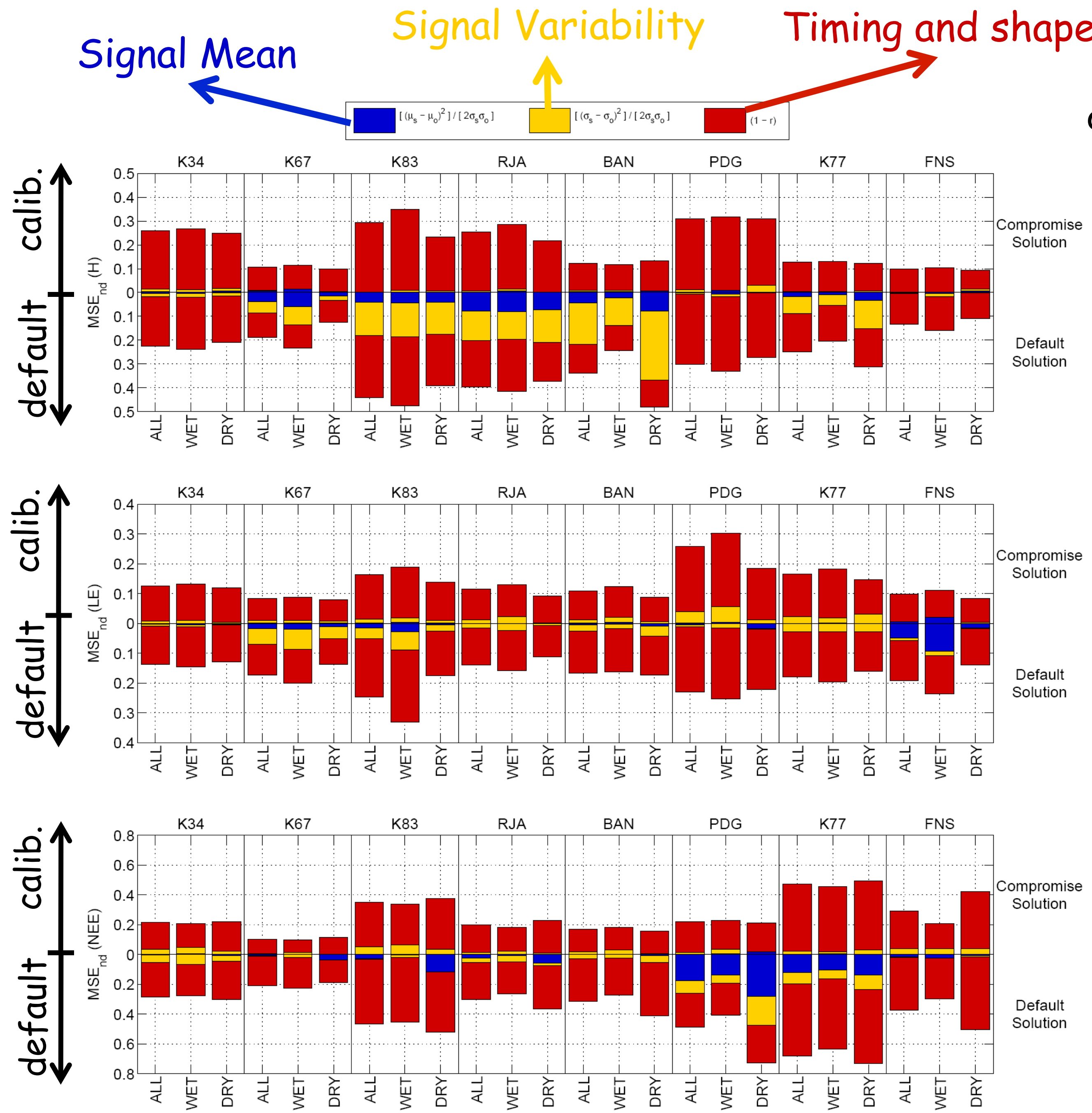
Mean-Squared-Error (MSE) Decomposition (Gupta et al. 2009)

$$MSE = \underbrace{(\mu_s - \mu_o)^2}_{\substack{\text{Error in} \\ \text{Signal Mean}}} + \underbrace{(\sigma_s - \sigma_o)^2}_{\substack{\text{Error in} \\ \text{Signal Variability}}} + \underbrace{2 \cdot \sigma_s \cdot \sigma_o \cdot (1 - r)}_{\substack{\text{Error in} \\ \text{Timing and shape}}}$$

Rearranging terms so individual signals are directly related to cross-correlation coefficient (r):

$$MSE_{non-\mathrm{dim}ensional} = \frac{MSE}{2 \cdot \sigma_s \cdot \sigma_o} = \underbrace{\frac{(\mu_s - \mu_o)^2}{2 \cdot \sigma_s \cdot \sigma_o}}_{\substack{\text{Signal} \\ \text{Mean}}} + \underbrace{\frac{(\sigma_s - \sigma_o)^2}{2 \cdot \sigma_s \cdot \sigma_o}}_{\substack{\text{Signal} \\ \text{Variability}}} + \underbrace{(1 - r)}_{\substack{\text{Timing} \\ \text{and shape}}}$$

MSE Decomposition: Reduced error in signal mean and signal variability



Percent contribution to overall uncertainty from individual components

2% 3% 95%

9% 23% 68%

2% 9% 89%

6% 12% 82%

1% 10% 90%

11% 9% 80%

Summary

1. **Newly proposed parameter sensitivity analysis approach** based on variance-based **Sobol method** accounting for full **multi-output nature** of land surface models

2. **Substantial improvement** (10-30% reduction of RMSE) in **energy**, **water** and **CO_2** flux simulations **after calibration**

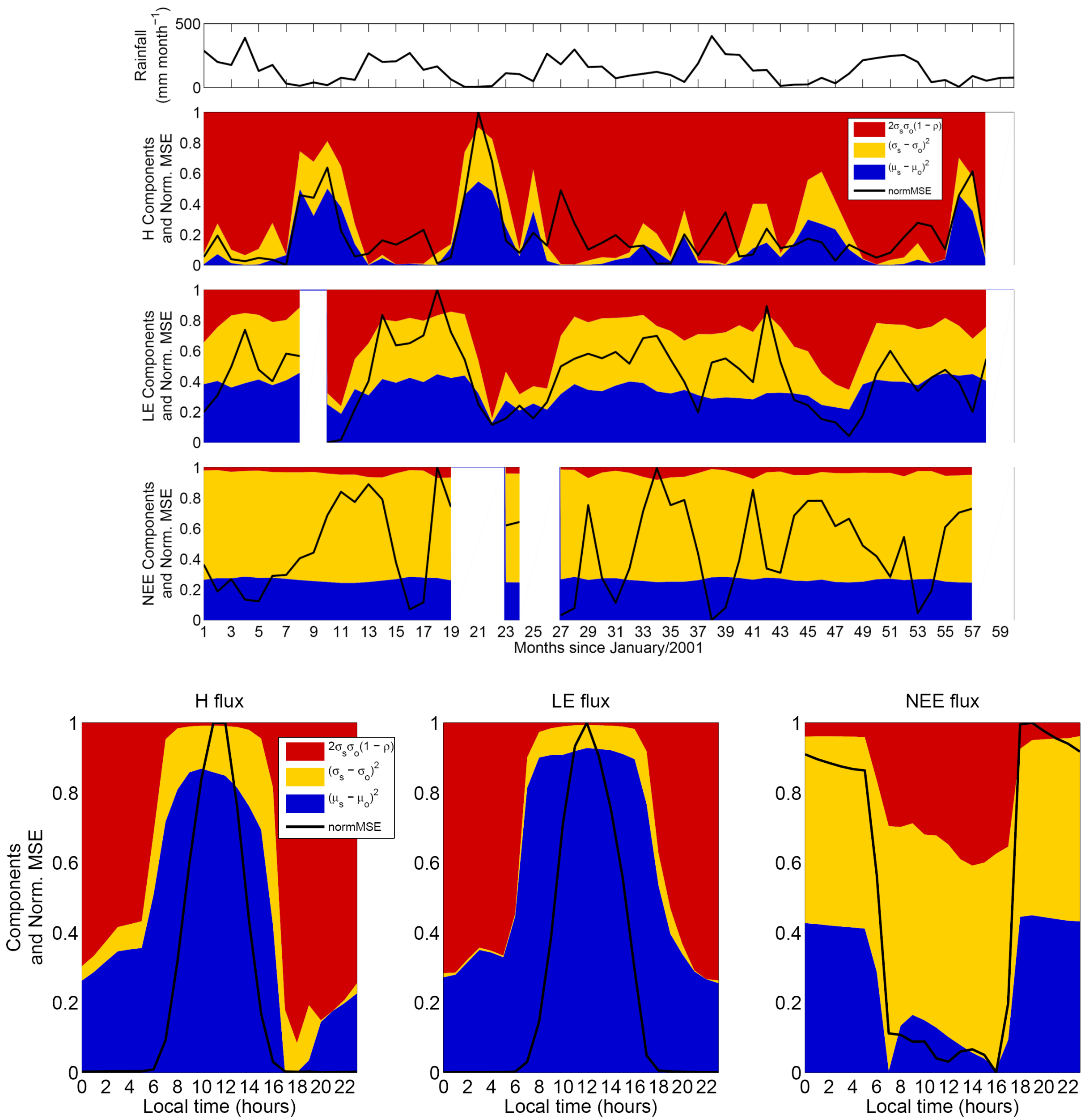
Additional:

- **Deeper effective soil depth** needed;

- **Soil water stress curvature parameter** different among biome types;

- Improvement in model performance is achieved by **reducing** the **signal mean and signal variability** components of parameter uncertainty but the timing/shape (i.e., **system dynamics**) is **little affected**.

Uncertainty sources at seasonal and diurnal scales



This is another optimization exercise (randomly generated parameters) and parameter set is selected based on individual minimization of each flux