Improving Efficiency in Uncertainty Quantification (UQ) Methods

Dr. Bryan Tolson

Associate Professor

Department of Civil and Environmental Engineering University of Waterloo

btolson@uwaterloo.ca

http://www.civil.uwaterloo.ca/btolson

engineering.uwaterloo.ca

STAC Workshop: Assessing Uncertainty in the Chesapeake Bay Program Modeling System

Feb 1, 2016

Broad Research Expertise

My research focuses largely in two general areas:

1. How do we build better env. simulation models?
   - Focus not on physics/chemistry but on procedures for fitting model predictions to measured data
   - Including calibration (optimization), quantifying prediction uncertainty, sensitivity analysis
2. How do we efficiently optimize the design or management of complex water resources systems?
   - This involves one or more numerical simulation models

I specifically concentrate on solving computationally intensive problems environmental modelling problems.

After my talk you will hopefully…

- Understand some ‘entry level’ UQ methods in environmental modelling
- Always think twice before using GLUE method for UQ or at least remember model pre-emption concept
- Remember that UQ methods should be conditioned to system response data if it is available
- Know there are multiple, relatively easy to implement techniques to make modern UQ methods more efficient

WATERLOO
ENGINEERING

Background

- Define $f(x_1, x_2, \ldots, x_n)$ as an environmental simulation model (or model chain) that predicts some key output of interest, y:

$$y = f(x_1, x_2, \ldots, x_n)$$

  - y might be avg. annual P load to Chesapeake Bay
  - where the x_i's are uncertain model inputs (model parameters, spatial data, forcing data like rainfall)
- Interestingly, we always artificially reduce uncertainty in y because we assume that many inputs are known exactly
- Fundamental: begin UQ by precisely defining y
- Note that uncertainty in a random variable y can be described by:
  - Variance of y, Range of y, Probability distribution of y

WATERLOO ENGINEERING

Background

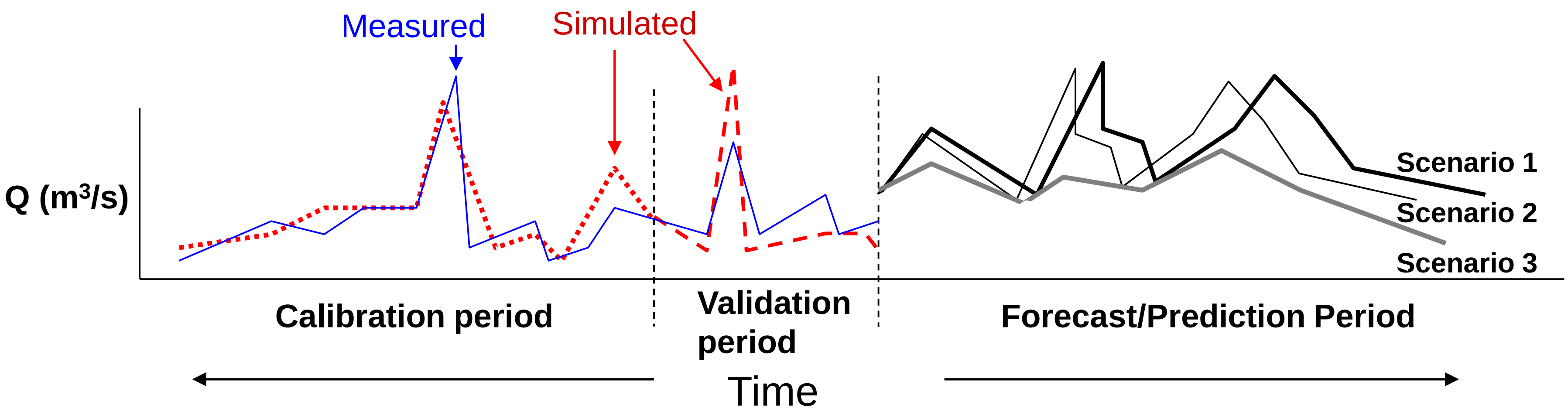
- UQ is 'easy' when environmental models are *not calibrated/fit* to measured system response data
  - Specify input uncertainty probability distributions
  - Perform Monte Carlo sampling experiment to propagate input uncertainties into model output(s) of interest

- UQ is much harder when we build models and want to *calibrate* them to site specific response data
  - Why? Simply put, the above MC sampling experiment will generate uncertain model parameter sets yielding simulated outputs that are completely inconsistent with system response data … and these should be ignored

WATERLOO ENGINEERING

Model Simulation Stages

- In general, when building/applying a model, if historical system response data is available, model simulations are conducted in three distinct steps:
  1. CALIBRATION 2. VALIDATION
  3. FORECAST/FUTURE/SCENARIO PREDICTION



- UQ would describe the uncertainty of the above deterministic responses using some form of prediction limits / quantiles / credible intervals

WATERLOO ENGINEERING

Some Observations of Current Literature and Practice

- Modern UQ methods are:
  - often applied only to model calibration period
  - less often to validation period
  - even less often to the scenario analysis period!!!

- Sensitivity analysis guides which input variable uncertainties must be considered as explicit uncertain variables
  - Unfortunately, the sensitivity analysis is typically conducted for the model calibration period
  - Sensitive variables in scenario analysis period not always sensitive in calibration period (and vice versa)

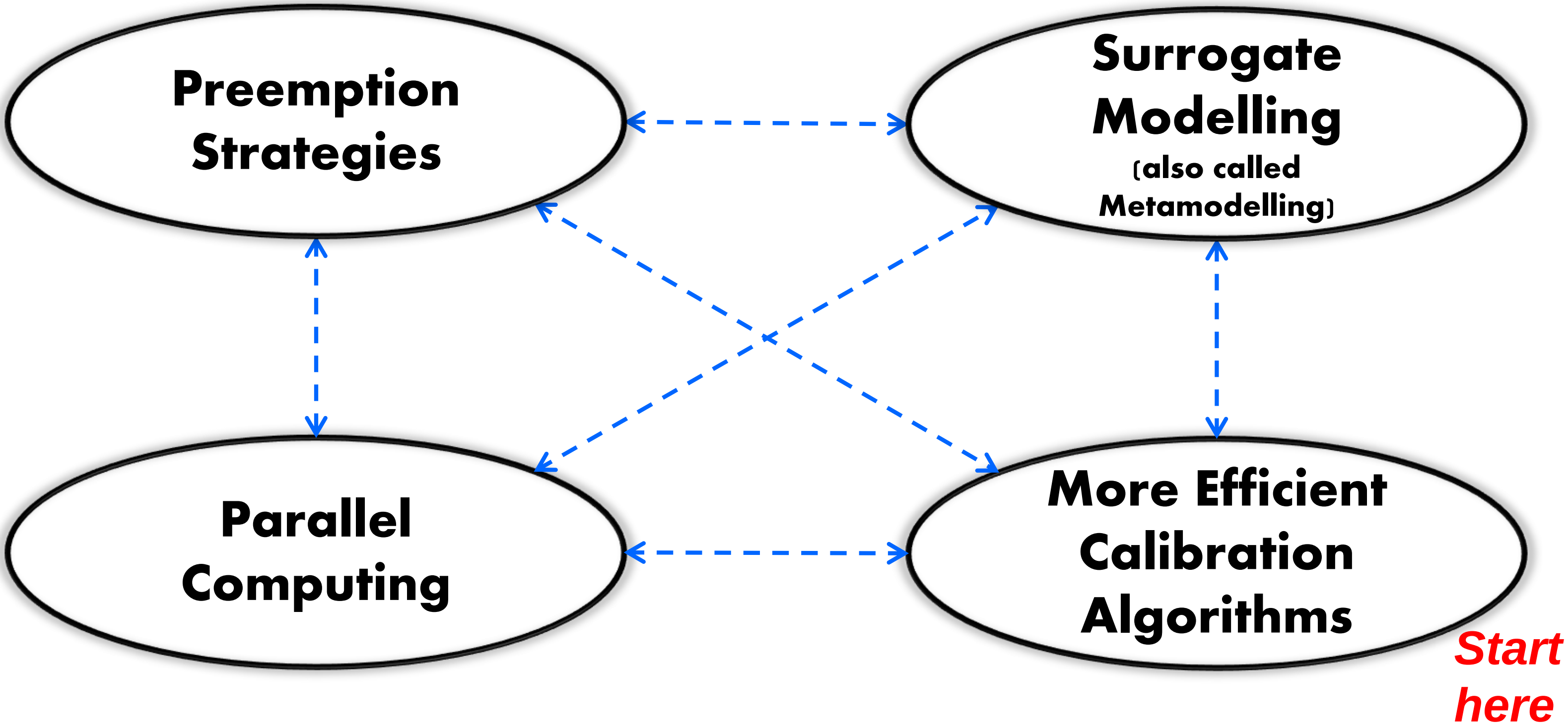
WATERLOO ENGINEERING

Scope of remaining material

- Scope of UQ methods I discuss is limited to describing the uncertainty in output variable *y* when our model is to be conditioned on measured system response data
- Focus on how we improve efficiency of modern UQ methods to describe uncertainty in *y*
- Outside of scope: How uncertain *y* values are used in decision-making

WATERLOO
ENGINEERING

Solution Approaches
to Circumvent Computational Burdens in Calibration/Optimization/UQ



WATERLOO ENGINEERING

GLUE Method (Beven & Binley, 1992)

- Extremely simple UQ method to understand and thus implement
- Original GLUE paper has more than 3000 citations
- Method based upon Spear & Hornberger (1980) global sensitivity analysis method
- Embraces "equifinality" concept (Beven & Freer, 2001):
  - there are multiple acceptable model parameter sets (model structures) that predict the calibration data
- Acceptable parameter sets termed "behavioural" and identified based their likelihood function (pseudo-likelihood) being higher than a subjectively defined threshold
  - e.g., Nash-Sutcliffe coefficient > 0.5 **in model calibration period**
- Behavioural parameter sets describe the joint posterior probability distribution of model parameters → can be sampled from for UQ in validation or scenario analysis stage

WATERLOO ENGINEERING

GLUE Need to Know

- "The GLUE methodology does not formally respect the rules of Bayesian inference." (Thiemann et al., 2001)
- "… may require massive computing resources for highly dimensioned parameter spaces." (Kuczera & Parent, 1998)
- GLUE results are conditioned on many subjective decisions (Montanari 2005; Shafii et al. 2015)
- GLUE prediction limits should not be interpreted as statistically valid confidence limits (Montanari, 2005)
- Vast additional debate in literature on validity of GLUE

GLUE Need to Know

- Original GLUE developers almost exclusively rely on simple Monte Carlo sampling (uniform PDFs) to search for behavioural parameter sets
- Monte Carlo sampling is an **absolutely terrible way** to search for high likelihood (good quality) model parameter sets ... especially in high-dimensional calibration problems
  - e.g. Brazier et al. (2000) required ~3 million model runs for 16 parameter soil erosion model
- This sampling approach often forces GLUE users to lower their behavioural thresholds (often needlessly!)

WATERLOO ENGINEERING

DDS-AU: An entry level UQ method

- DDS-AU: Dynamically Dimensioned Search - Approximation of Uncertainty
- Conceptually similar *but even simpler* than GLUE method
- DDS-AU is *orders of magnitude* faster than GLUE method
- If you have defined an objective function (for model calibration purposes), you can immediately apply the DDS-AU approach

WATER RESOURCES RESEARCH, VOL. 44, W04411, doi:10.1029/2007WR005869, 2008

Click Here for Full Article

Efficient prediction uncertainty approximation in the calibration of environmental simulation models

Bryan A. Tolson[1] and Christine A. Shoemaker[2]

DDS-AU: An entry level UQ method

- In a nutshell, DDS-AU solves the calibration problem with multiple, relatively quick, independent optimization trials
- At most, one behavioural parameter set is retained per optimization trial and used in ensemble to generate model prediction bounds
- Tolson and Shoemaker (2008) bound predictions rather than making probabilistic statements
- Fewer subjective decisions in DDS-AU compared to GLUE
- DDS-AU is efficient because it relies on the simple & robust DDS optimization algorithm (Tolson & Shoemaker, 2007)

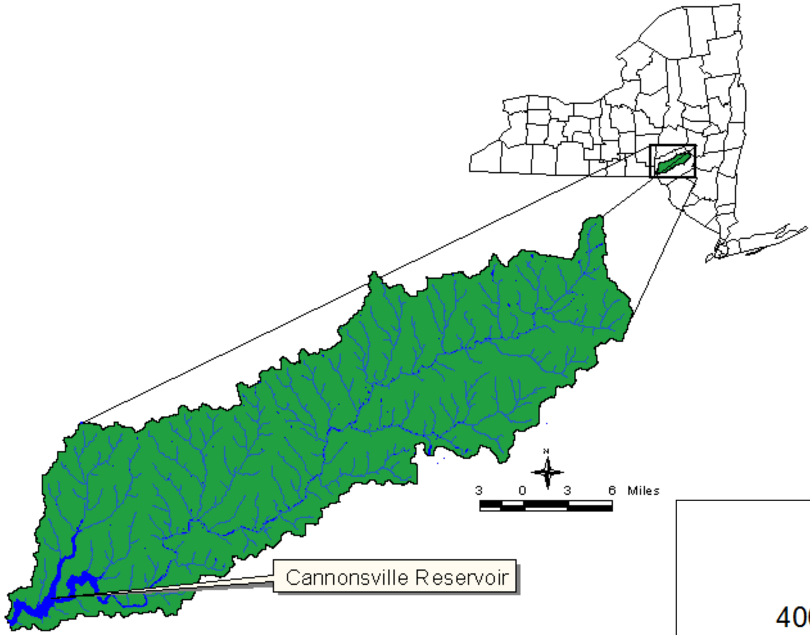
WATER RESOURCES RESEARCH, VOL. 43, W01413, doi:10.1029/2005WR004723, 2007

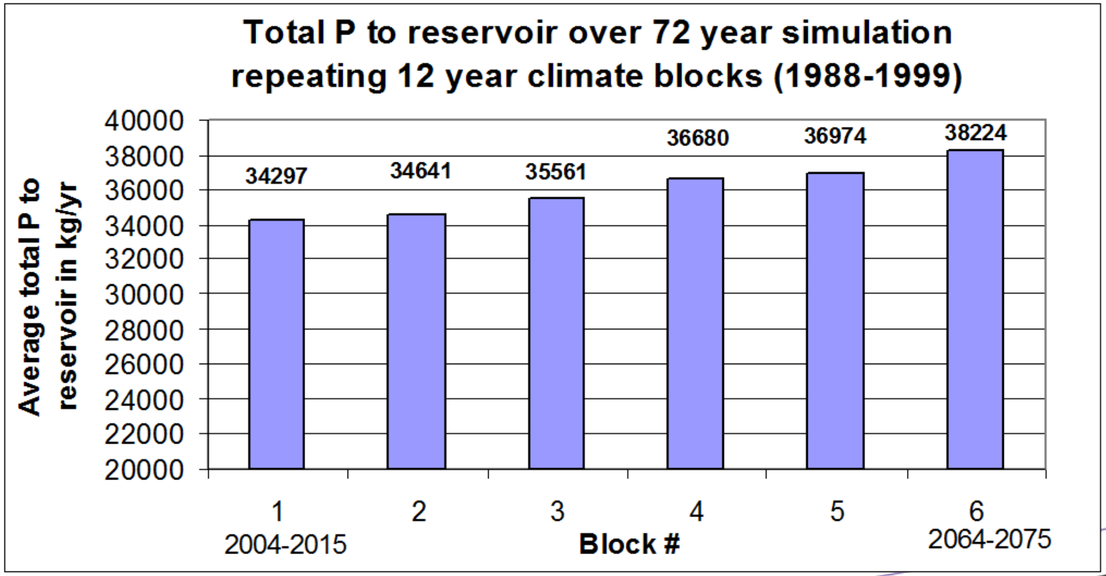
Click Here for Full Article

Dynamically dimensioned search algorithm for computationally efficient watershed model calibration

Bryan A. Tolson[1] and Christine A. Shoemaker[2]

SWAT2000 Model Case Study





WATERLOO ENGINEERING

Comparing DDS-AU to GLUE for SWAT2000 Model Calibration

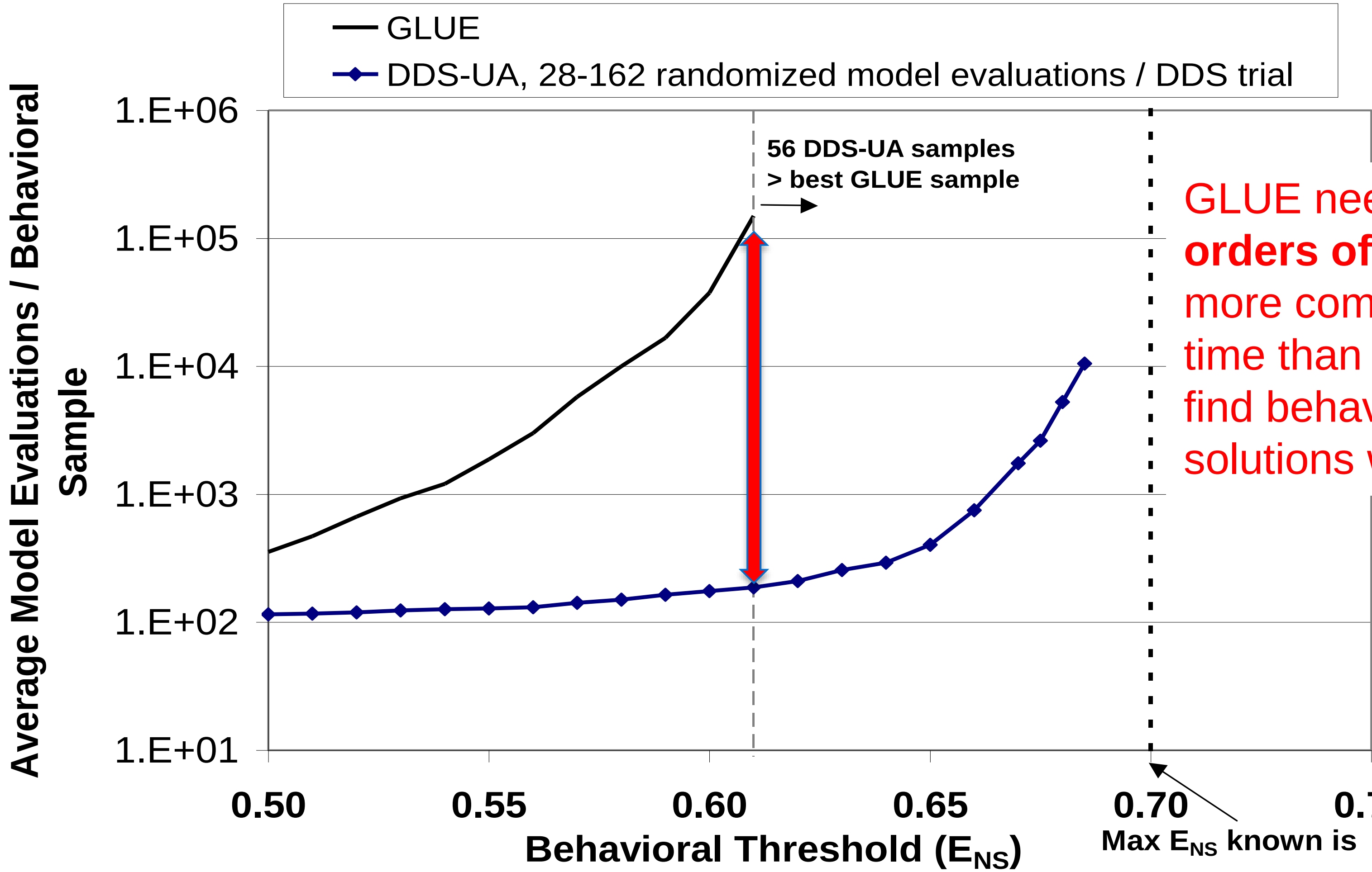
UQ Problem 1:

- Maximize pseudo-likelihood function, $L = E_{NS}$ (Nash coeff. on avg daily Q)
- Define behavioural threshold of $E_{NS} \geq 0.5$
- Default SWAT2000 parameter ranges for 14 flow parameters
- Use 120000 model evaluations for GLUE
- Use 10000 model evaluations for DDS-UA
- DDS-UA:
  - find 100 behavioural solutions using 100 independent DDS optimization trials
  - Each optimization trial uses 10000/100 = 100 model evaluations

WATERLOO ENGINEERING

Comparing DDS-AU to GLUE for SWAT2000 Model Calibration

UQ Problem 1 results:

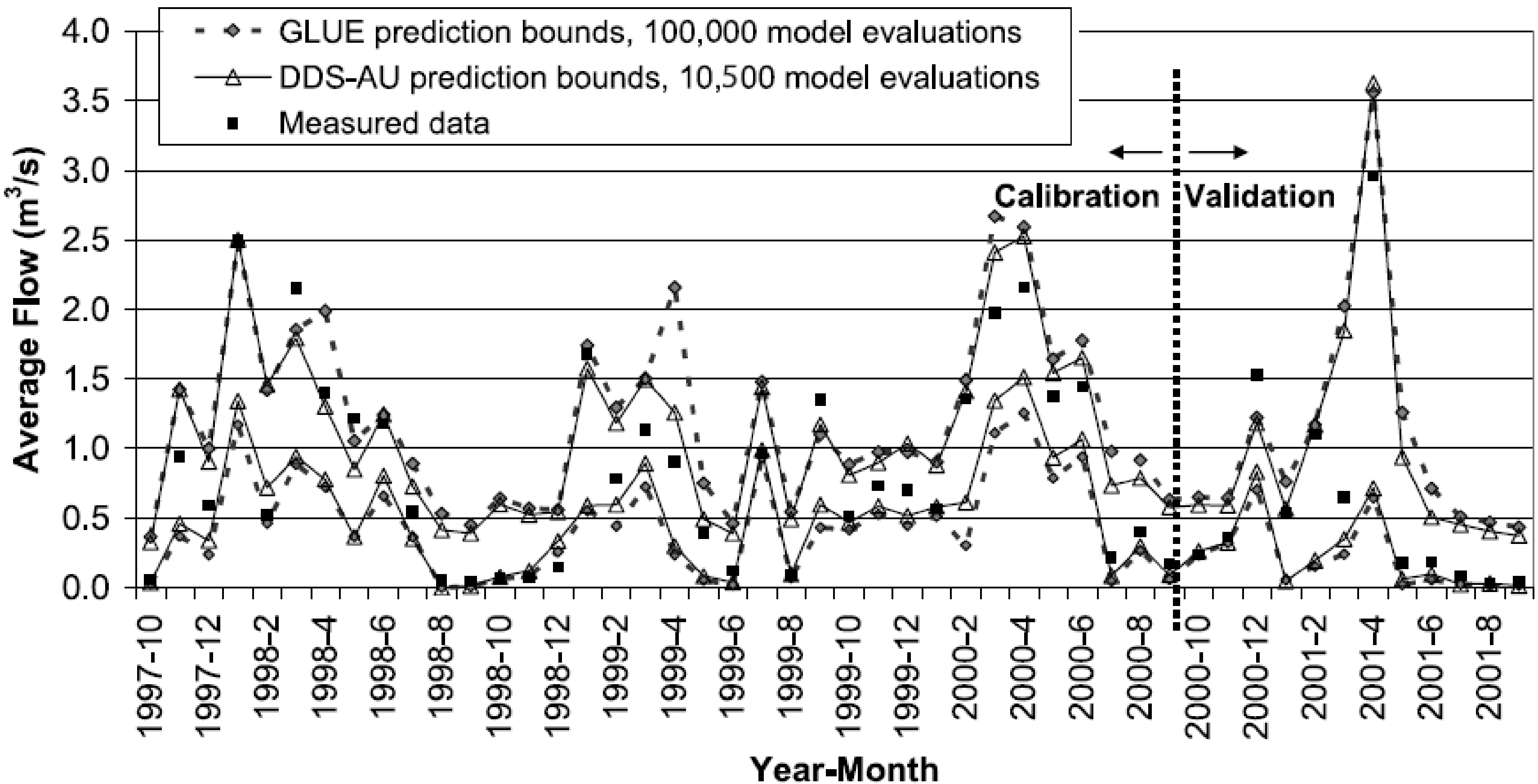


GLUE needs **3 orders of magnitude** more computation time than DDS-AU to find behavioural solutions w E_{NS} > 0.6

Comparing DDS-AU to GLUE for SWAT2000 Model Calibration

UQ Problem 1 results:

100% prediction limits for GLUE and DDS-AU



Comparing DDS-AU to GLUE for SWAT2000 Model Calibration

UQ Problem 2:

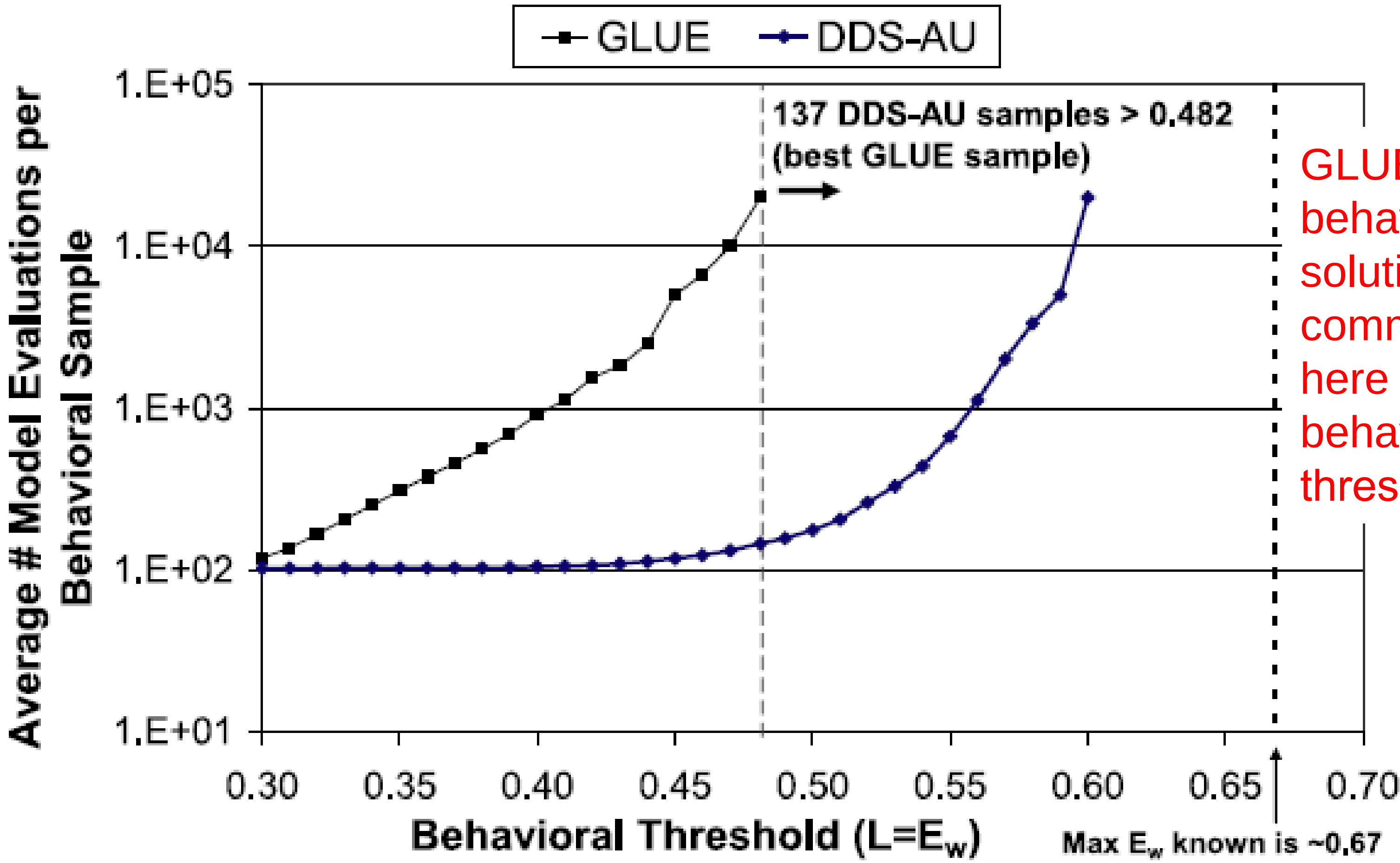
- 30 uncertain parameters with reduced parameter ranges based on experience
- Maximize Likelihood, L, defined as E_w below:

$$\begin{aligned}\underset{\boldsymbol{x}}{\mathrm{Max}}\;\; E_w(\boldsymbol{x}) = {} & 0.5\left(E_{NS}^{Q} - \max[0, |\%B^{Q}| \text{-} 10]/100\right) + \\ & 0.2\left(E_{NS}^{TSS} - \max[0, |\%B^{TSS}| \text{-} 30]/100\right) + \\ & 0.3\left(E_{NS}^{TP} - \max[0, |\%B^{TP}| \text{-} 30]/100\right)\end{aligned}$$

- Above L is similar to a weighted E_{NS}
- Based on prior manual calibration, wanted L > 0.5
- 20,000 model evaluations for **both** GLUE & DDS-UA: **40 days**
- DDS-UA: 200 behavioural samples desired
  - find 200 behavioural solutions using 200 independent DDS optimization trials
  - Each optimization trial uses 20000/200 = 100 model evaluations

Comparing DDS-AU to GLUE for SWAT2000 Model Calibration

UQ Problem 2 results:

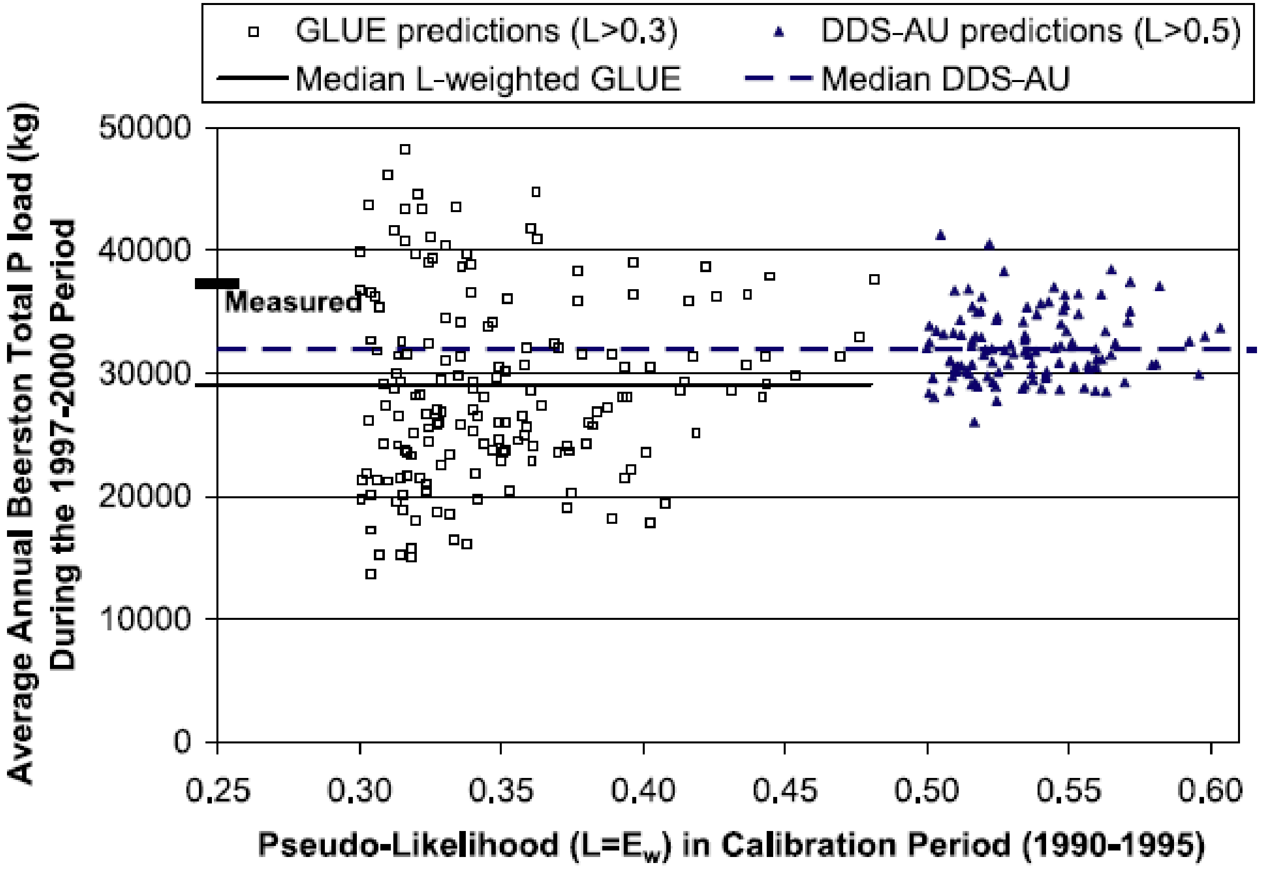


GLUE returns 0 behavioural solutions → common practice here is to reduce behavioural threshold …

Comparing DDS-AU to GLUE for SWAT2000 Model Calibration

UQ Problem 2 results:

impact of artificially reducing behavioral threshold



DDS-AU Remarks

- DDS-AU much less sensitive to prior parameter ranges than GLUE (results not shown)
- Like GLUE, DDS-AU is just as simple to run in parallel (each processor assigned one optimization trial)
- With any calibration objective function you can immediately apply the DDS-AU approach to approximate parameter uncertainty
- Working on extensions to method currently

WATERLOO ENGINEERING

Surrogate Modelling in Water Resources

- Also called Metamodelling
- Essentially this is developing a model of another model
  - Replace hydrologic model with a statistical model (e.g., ANN)
  - Replace complex hydrologic model with a simpler, faster model
  - The metamodel used in place of original model when model must be simulated thousands of times (e.g., auto-calibration)
- Razavi et al. (2012a) review this literature
- Razavi et al. (2012b) perform numerical tests of the effectiveness of metamodels to improve optimization efficiency
-

Razavi, S., B. A. Tolson and D. H. Burn (2012a), Review of surrogate modeling in water resources, *Water Resources Research*, 48, W07401, doi:10.1029/2011WR011527.
Razavi, S. S., B. A. Tolson and D. H. Burn (2012b), Numerical assessment of metamodelling strategies in computationally intensive optimization, *Environmental Modelling and Software*. 34(0), 67-86. doi:10.1016/j.envsoft.2011.09.010

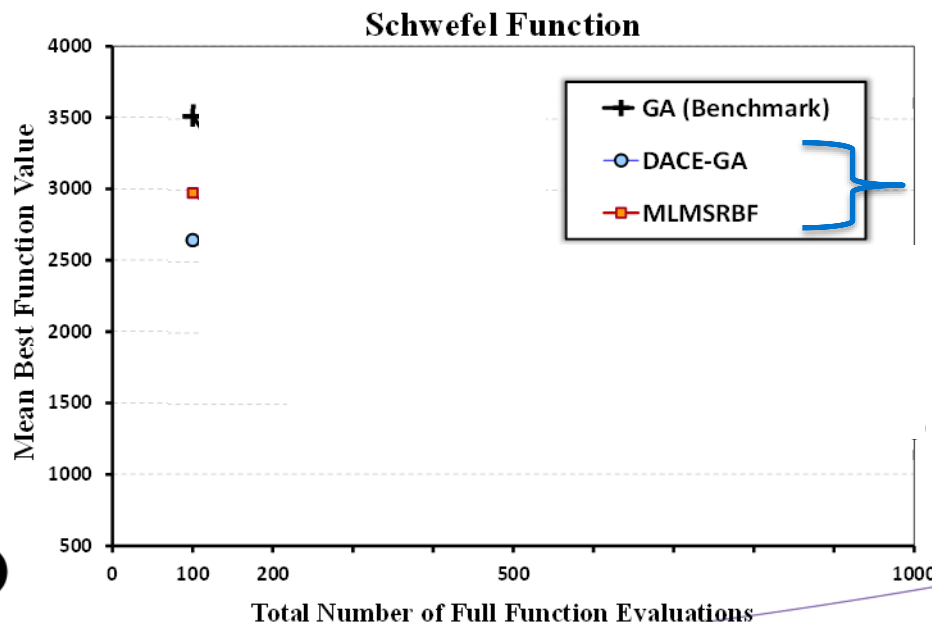
WATERLOO ENGINEERING

From the Literature Introducing Metamodel Enabled Optimizers

Literature suggests:

Response surface surrogates yields enhanced solution efficiency and effectiveness of computationally intensive optimization problems.

Typically based on an analysis like this:



Metamodel enabled optimizers

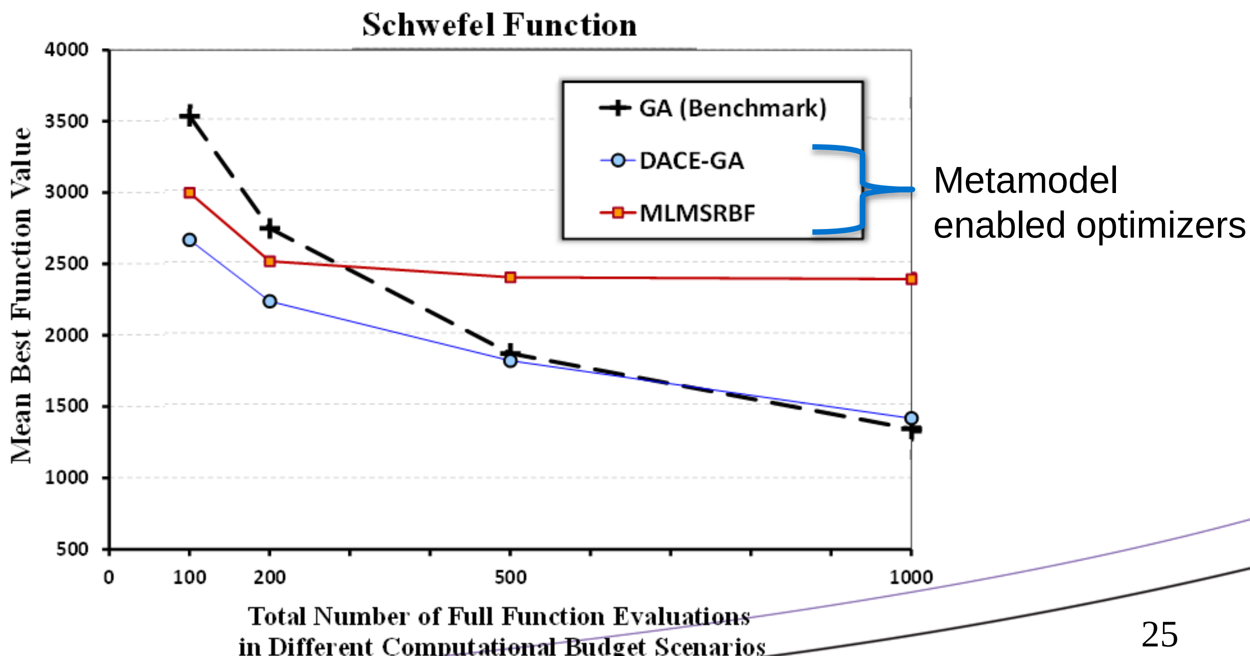
WATERLOO ENGINEERING

From the Literature Introducing Metamodel Enabled Optimizers

Literature suggests:

Response surface surrogates yields enhanced solution efficiency and effectiveness of computationally intensive optimization problems.

My past student Razavi performed comparison like this:



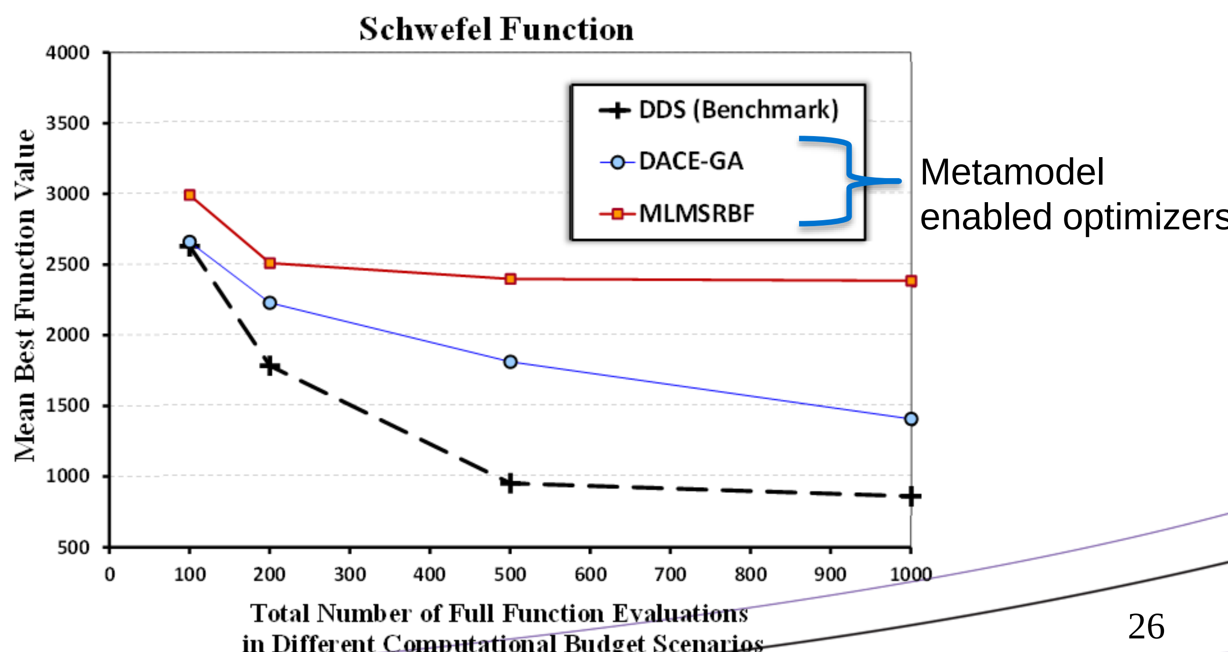
WATERLOO ENGINEERING

From the Literature Introducing Metamodel Enabled Optimizers

Literature suggests:

Response surface surrogates yields enhanced solution efficiency and effectiveness of computationally intensive optimization problems.

With DDS as benchmark, we found this in 67% of comparisons:



WATERLOO ENGINEERING

Surrogate Modelling for UQ Speedup?

Conclusions re Optimization in Razavi et al. (2012b):

metamodelling can be misleading and a hindrance, such that better solutions are achieved with optimizers not involving metamodels.

- I would *hypothesize* that the same could be said about surrogate models/metamodels in UQ context
- If you go down this road for UQ Speedup please:
  - Do a very thorough literature search (see Razavi et al. 2012a)
  - Realize there are an unwieldly number of subjective decisions
  - Many surrogate model approaches tend to become unviable as dimensionality (# parameters uncertain increases)
  - Consider lower-fidelity physically based surrogates → this is effectively what you are doing with Phase 6 models (see Razavi)

WATERLOO ENGINEERING

The Model Preemption Concept

- *The most straightforward contribution I will ever make*

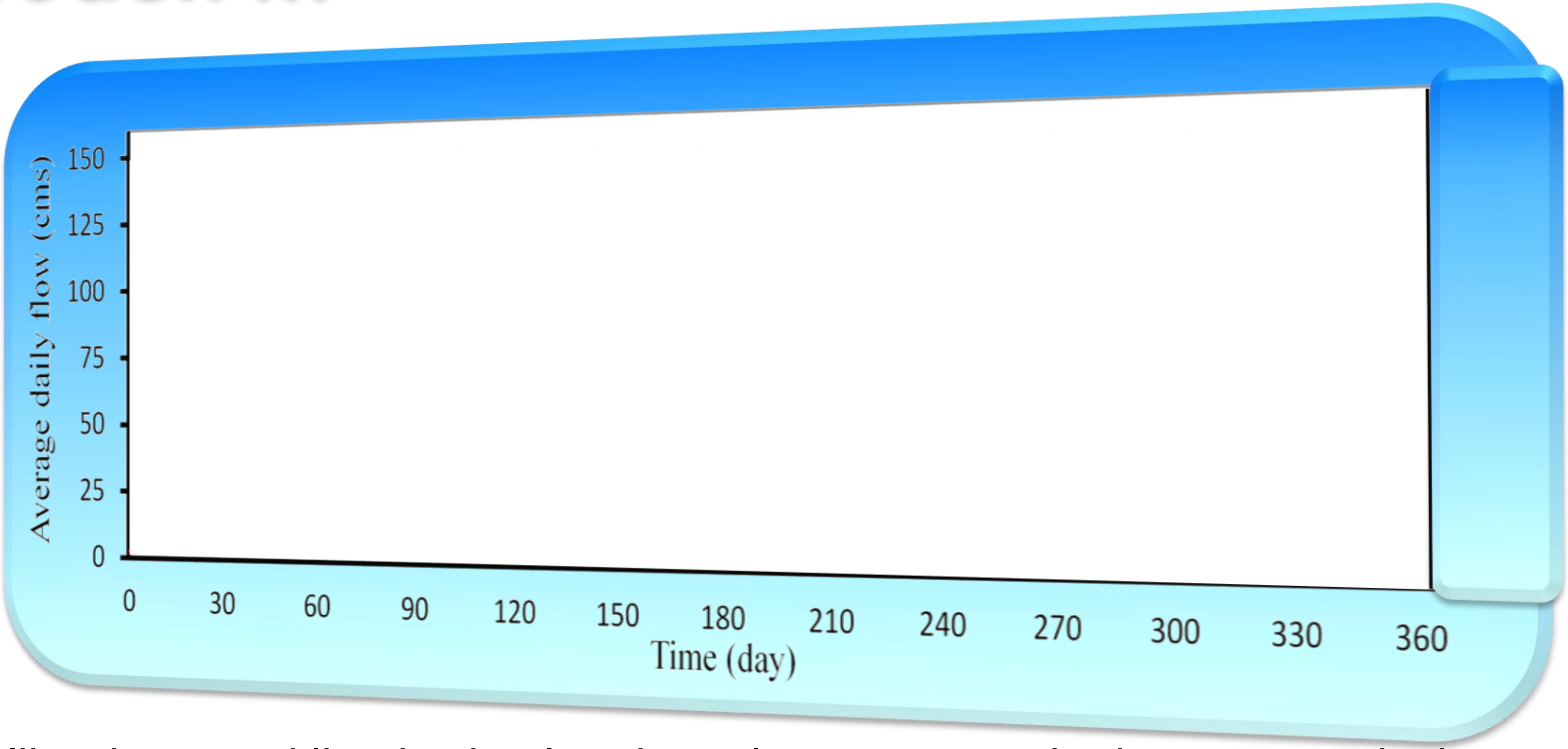
WATER RESOURCES RESEARCH, VOL. 46, W11523, doi:10.1029/2009WR008957, 2010

Reducing the computational cost of automatic calibration through model preemption

Saman Razavi,[1] Bryan A. Tolson,[1] L. Shawn Matott,[1] Neil R. Thomson,[1] Angela MacLean,[1] and Frank R. Seglenieks[1]

WATERLOO ENGINEERING

Simulating manual calibration approach …



Manual calibration: Red line is simulated result you are monitoring on your desktop

… assume calibration period is actually another 2-3 years. Are you going to wait for model to finish?

WATERLOO ENGINEERING

Pre-emption is simply stopping simulation model when you know the solution is terrible

Model Pre-emption for UQ

- Pre-emption improves efficiency of:
  - Various optimization algorithms
  - DDS-AU
  - GLUE
  - MCMC samplers for Bayesian calibration
- Deterministic pre-emption is “free”!
- Two catches to pre-emption:
  - The calibration objective function must monotonically degrade in quality with simulation period length (e.g., SSE)
  - The search algorithm must have an objective pre-emption threshold that separates promising/non-promising solutions
- Model pre-emption also applicable to mgt. optimization (post-calibration) as in Asadzadeh et al. (2014)

Results of Actual Computational Savings due to deterministic model pre-emption

Case Study	DDS	PSO	SCE	GLUE Behavioral threshold $E_{ns}=0$	GLUE Behavioral threshold $E_{ns}=0.5$	DDS-AU
SWAT-1	14%	34%	---	---	---	---
SWAT-2	21%	53%	5%	52%	70%	18%
MESH	49%	---	---	95%	**96%**	---
DFRTT	37%	**59%**	---	---	---	---

WATERLOO ENGINEERING

Model Pre-emption results for MCMC

- Work by my group in Shafii et al. (2015)
- Applied to DREAM (Vrugt et al. 2009) and Sequential Monte Carlo (SMC) (see Jeremiah et al. 2011) sampling algorithms
- Our likelihood functions monotonically degrade with simulation length

- Overall pre-emption computational time savings range from 5% to 21%
- Pre-emption savings greatest in burn-in period (to 39%)
- Pre-emption highest utility when users fine-tuning MCMC

Model Pre-emption in MCMC

- DREAM and SMC decide to jump from a current state ($\boldsymbol{\theta}_{\mathbf{n}}$) to a candidate state ($\boldsymbol{\theta}^{*}$) based on the ratio of the posterior densities of the two states: $p(\boldsymbol{\theta}^{*} \mid \mathbf{Y}) \big/ p(\boldsymbol{\theta}_{\mathbf{n}} \mid \mathbf{Y})$
- The candidate state is accepted if

$p(\boldsymbol{\theta}^{*} \mid \mathbf{Y}) \big/ p(\boldsymbol{\theta}_{\mathbf{n}} \mid \mathbf{Y}) > Z$, where Z is a random number uniformly distributed between 0 and 1;

- Pre-emption threshold: $p(\boldsymbol{\theta}^{*} \mid \mathbf{Y})_{\min} = Z \times p(\boldsymbol{\theta}_{\mathbf{n}} \mid \mathbf{Y})$
- Compute above before candidate solution evaluated
- If at any time step, the intermediate posterior density is smaller than the threshold, the simulation can be stopped

WATERLOO ENGINEERING

Additional Ideas to Improve Calibration and UQ

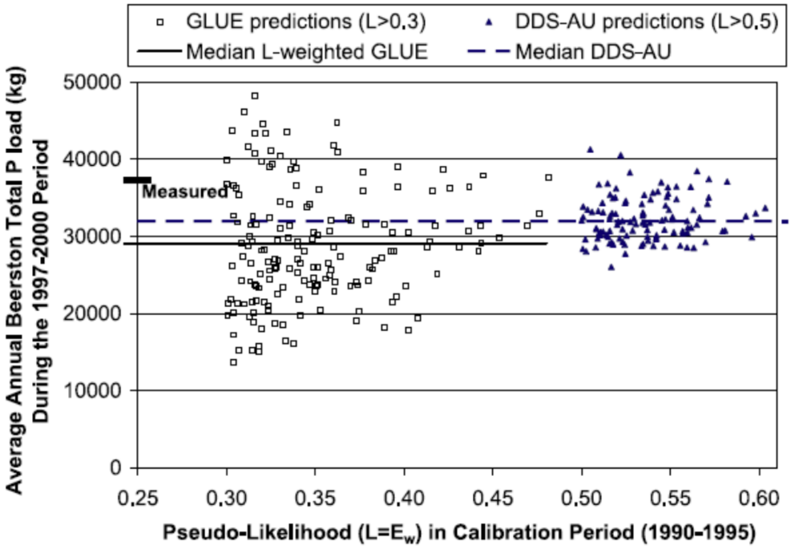
1. Calibrating on short subsets of historical system response data (Razavi and Tolson, 2013)
2. Addressing subjectivity in GLUE (Shafii et al. 2015)
3. Non-converged MCMC performance reasonably consistent with a formal, fully-converged Bayesian approach (despite large differences in # of model evaluations). See Shafii et al. (2014)
4. Calibrating to numerous hydrological signatures simultaneously with MO tools (Shafii & Tolson, 2015)

Concluding Remarks

1. It is entirely possible that the final UQ results indicate a **massive** amount of uncertainty such that the "confidence" in mgt. impacts could be very low
   a. Step 1 is quantify uncertainty in output of interest (UQ)
   b. Step 2 is investigate which sources of input uncertainty can/should be reduced (Uncertainty *Analysis*)
2. True uncertainty in model predictions are greatest if the model parameters/structure are not conditioned on the measured system response data

conditioning → calibration

must calibrate!!!



WATERLOO ENGINEERING

Concluding Remarks

3. Model independent parameter estimation software with my research group's algorithms embedded (DDS, DDS-AU, PA-DDS, parallel computation versions):
   - OSTRICH software by Dr. L.S. Matott at U. Buffalo
   - Linux, Windows PCs & clusters
   - http://www.eng.buffalo.edu/~lsmatott/Ostrich/OstrichMain.html
   - U. Buffalo super-computing center has programs to serve industry and research consortiums like this one
4. THANK YOU! And thanks to my graduate students contributing to above work: Saman, Masoud, Mahyar

QUESTIONS?

See: www.civil.uwaterloo.ca/btolson/
btolson@uwaterloo.ca

WATERLOO ENGINEERING

References

Beven, K., and A. Binley (1992), The future of distributed models—Model calibration and uncertainty prediction, Hydrol. Processes, 6(3), 279–298.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249(1 –4), 11–29.

Brazier, R. E., K. J. Beven, J. Freer, and J. S. Rowan (2000), Equifinality and uncertainty in physically based soil erosion models: Application of the glue methodology to WEPP—The water erosion prediction project for sites in the UK and USA, Earth Surf. Processes Landforms, 25(8), 825– 845.

Jeremiah, E., Sisson, S., Marshall, L., Mehrotra, R., Sharma, A., 2011. Bayesian calibration and uncertainty analysis of hydrological models: A comparison of adaptive Metropolis and sequential Monte Carlo samplers. Water Resources Management, 47(W07547).

Montanari, A. (2005), Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall runoff simulations, Water Resour. Res., 41, W08406, doi:10.1029/2004WR003826.

Razavi, S., and B. A. Tolson (2013), An efficient framework for hydrologic model calibration on long data periods, Water Resour. Res., 49, 8418–8431, doi:10.1002/2012WR013442.

Shafii, M., B. Tolson, and L. Matott (2014), Uncertainty-based multi-criteria calibration of rainfall-runoff models: a comparative study, Stoch. Env. Res. Risk. A., 28(6), 1493-1510.

Shafii, M., B. A. Tolson, and L. S. Matott (2015), Addressing subjective decision-making inherent in GLUE-based multi-criteria rainfall–runoff model calibration, J. Hydrol., 523, 693-705, doi:10.1016/j.jhydrol.2015.01.051.

Shafii, M. H., B. A. Tolson and L. S. Matott (2015), Improving the efficiency of Monte Carlo Bayesian calibration of hydrologic models via model pre-emption, J. Hydroinform., 17 (5), 763-770; doi:10.2166/hydro.2015.043.

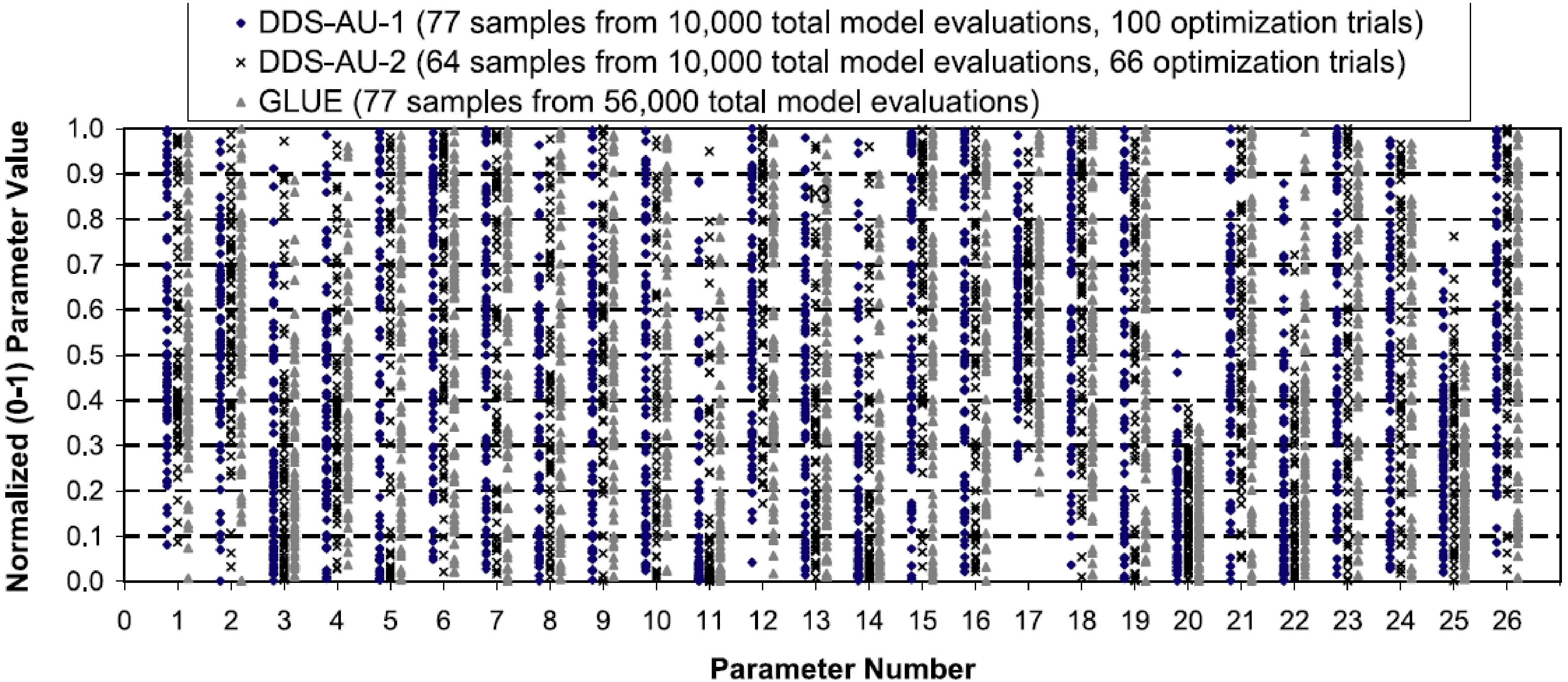
Spear, R. C., and G. M. Hornberger (1980), Eutrophication in Peel Inlet. 2. Identification of critical uncertainties via generalized sensitivity analysis, Water Res., 14(1), 43–49.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, Water Resour. Res., 37(10), 2521– 2535.

Vrugt, J.A. et al., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. International Journal of Nonlinear Sciences and Numerical Simulation, 10(3): 273-290.

Extra slides

Demonstration of Equifinality



WATERLOO
ENGINEERING

Dynamically Dimensioned Search

- Stochastic global optimization algorithm (direct search method)
- Originally designed for automatic calibration of environmental simulation models:
  - Simple to implement & no algorithm parameter-tuning needed
  - Scales search to yield calibration result **in modeller’s time frame**
  - Calibration philosophy: to **find good or acceptable solutions**, not the precise globally optimal solution

WATER RESOURCES RESEARCH, VOL. 43, W01413, doi:10.1029/2005WR004723, 2007

Click Here for Full Article

Dynamically dimensioned search algorithm for computationally efficient watershed model calibration

Bryan A. Tolson[1] and Christine A. Shoemaker[2]

WATERLOO ENGINEERING

DDS Algorithm Description

- Algorithm scales to user-specified computational limits
- DDS mimics the manual calibration process

STEP 1. Define DDS inputs for *D* dimensional problem:
-neighborhood perturbation size parameter, *r* (0.2 is default)
-maximum # of function evaluations, *m*

STEP 2. Evaluate objective function at initial solution

STEP 3. Randomly select a subset of the *D* decision variables for perturbation from the current best solution.

→**Size of subset decreases as maximum function evaluation limit approached**

STEP 4. Perturb the decision variables selected in Step 3 from their current best solution

STEP 5. Evaluate new solution and update current best solution if necessary

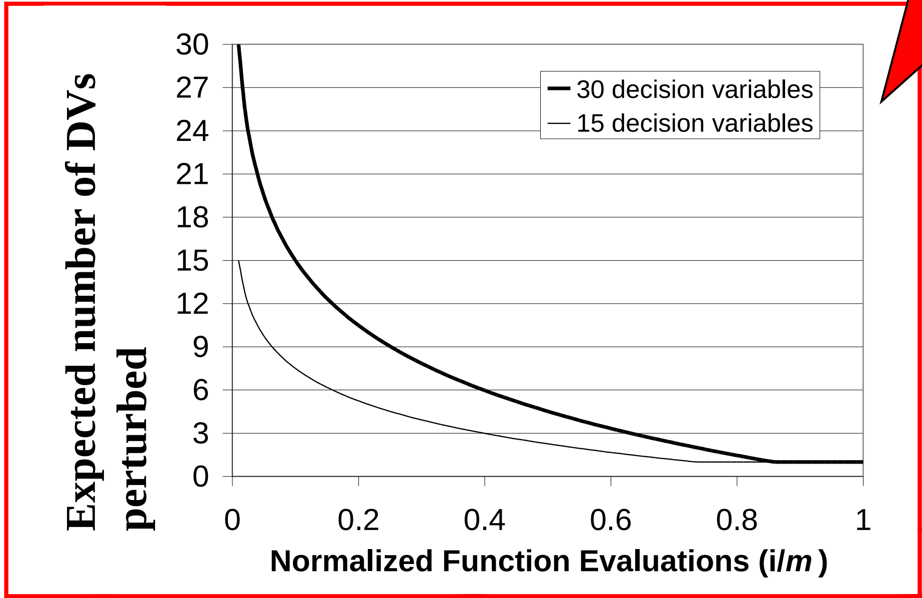
STEP 6. Update function evaluation counter, i=i+1, and check stopping criterion:
-IF i = *m* → STOP
-ELSE repeat STEP 3

WATERLOO ENGINEERING

DDS Algorithm Description

STEP 3. Randomly select a subset of the D decision variables for perturbation from the current best solution

- Probability each decision variable perturbed is: $P(i) = 1 – \ln(i)/\ln(m)$
  where i is obj. function evaluation counter
  m is maximum number of obj. functions



WATERLOO ENGINEERING

DDS Algorithm Description

- Algorithm scales to user-specified computational limits
- DDS mimics the manual calibration process

STEP 1. Define DDS inputs for *D* dimensional problem:
-neighborhood perturbation size parameter, *r* (0.2 is default)
-maximum # of function evaluations, *m*

STEP 2. Evaluate objective function at initial solution

STEP 3. Randomly select a subset of the *D* decision variables for perturbation <u>from the current best</u> solution.

→**Size of subset decreases as maximum function evaluation limit approached**

STEP 4. Perturb the decision variables selected in Step 3 from their current best solution

→ **normally distributed perturbations with adequate variance ensures global search**
→ **perturbations beyond decision variable boundary are reflected**

STEP 5. Evaluate new solution and update current best solution if necessary

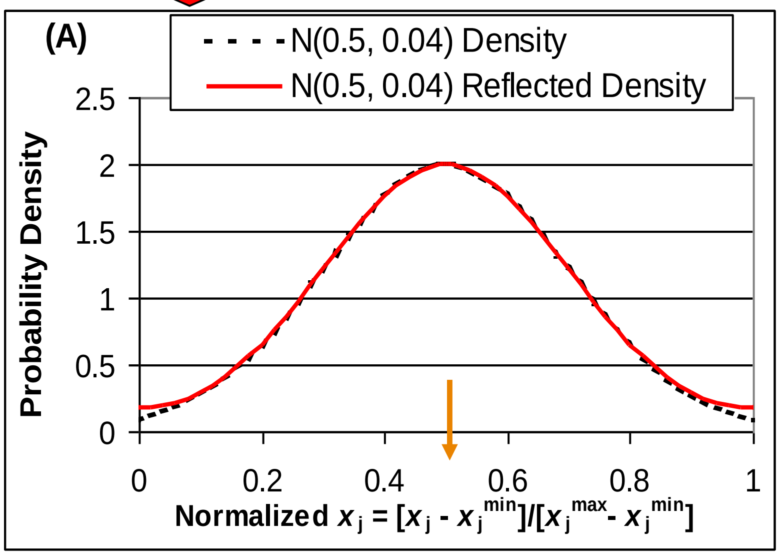
STEP 6. Update function evaluation counter, i=i+1, and check stopping criterion:
-IF i = *m* → STOP
-ELSE repeat STEP 3

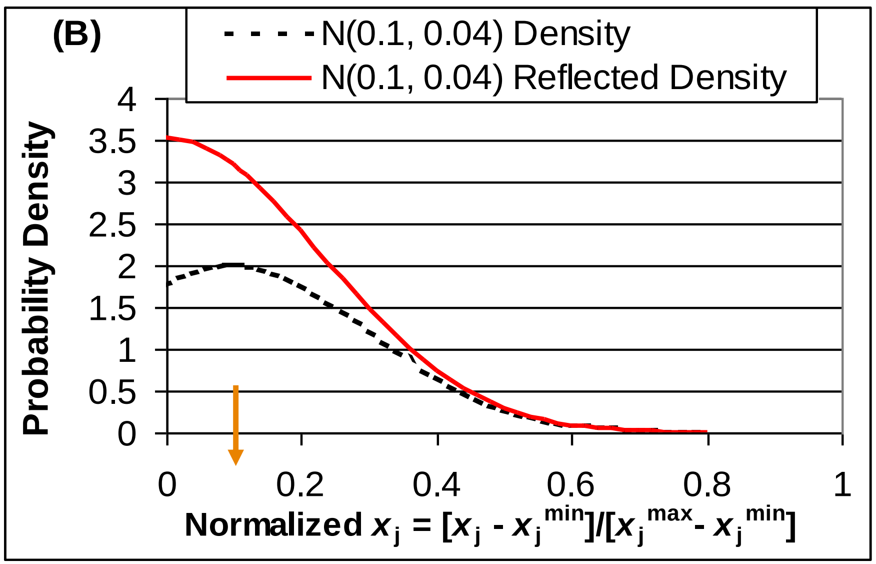
WATERLOO ENGINEERING

DDS Algorithm Description

STEP 4. Perturb the decision variables selected in Step 3 from their current best solution

Example of perturbation probability distributions (one dimension):





$r = 0.2$, normalized $x_j^{best} = 0.5$

$r = 0.2$, normalized $x_j^{best} = 0.1$

r is the only DDS parameter & **$r = \dfrac{\text{std. deviation of perturbation size}}{\text{decision variable range}}$**

DDS Algorithm Description

- Algorithm scales to user-specified computational limits
- DDS mimics the manual calibration process

STEP 1. Define DDS inputs for *D* dimensional problem:
-neighborhood perturbation size parameter, *r* (0.2 is default)
-maximum # of function evaluations, *m*

STEP 2. Evaluate objective function at initial solution

STEP 3. Randomly select a subset of the *D* decision variables for perturbation from the current best solution.
→**Size of subset decreases as maximum function evaluation limit approached**

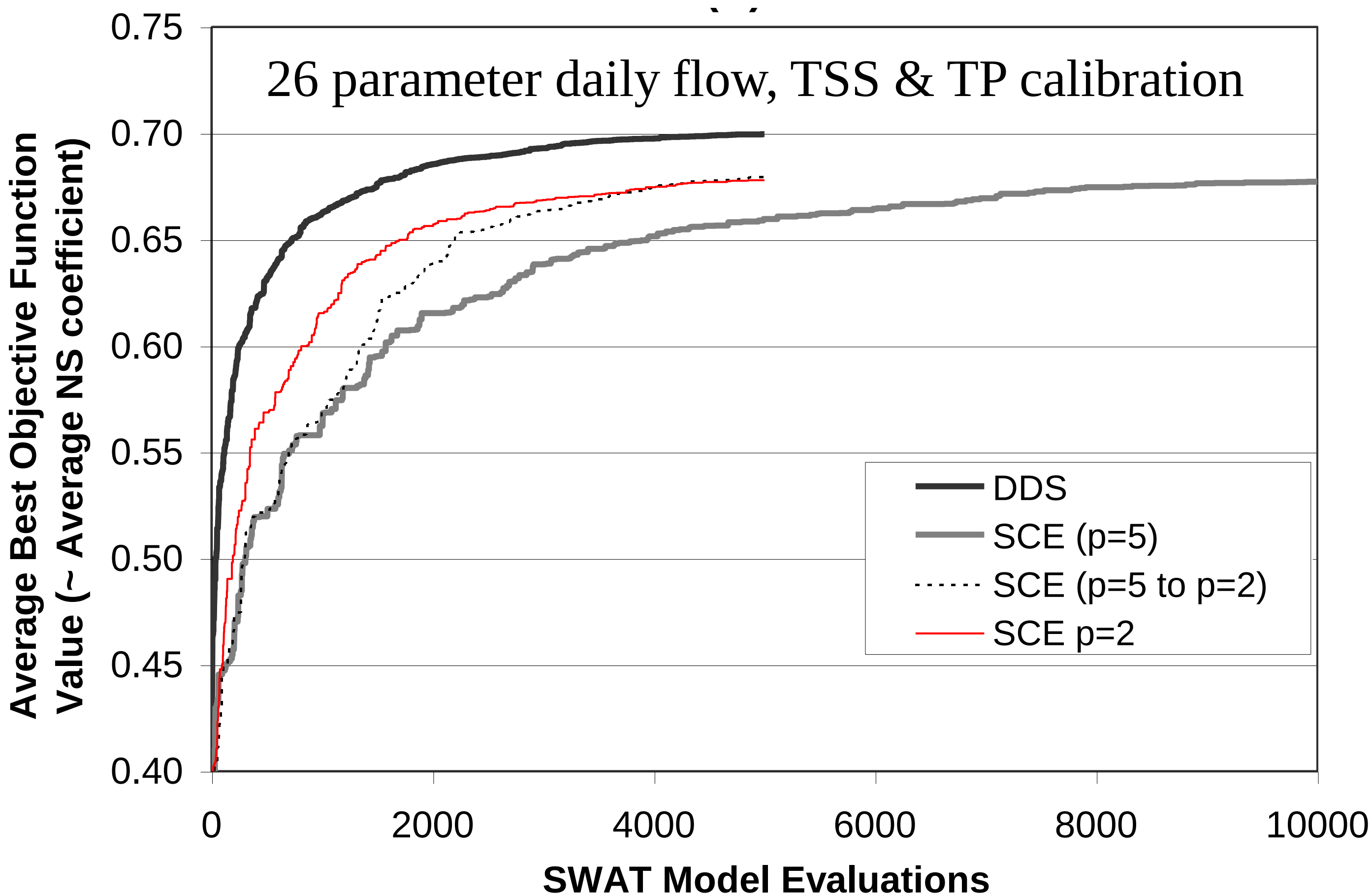
STEP 4. Perturb the decision variables selected in Step 3 from their current best solution
→ **normally distributed perturbations with adequate variance ensures global search**
→ **perturbations beyond decision variable boundary are reflected**

STEP 5. Evaluate new solution and update current best solution if necessary

STEP 6. Update function evaluation counter, i=i+1, and check stopping criterion:
-IF i = *m* → STOP
-ELSE repeat STEP 3

WATERLOO ENGINEERING

DDS Performance in comparison with "Fine-Tuned" SCE



WATERLOO ENGINEERING

Robust Performance of DDS

- My group has applied DDS to numerous case studies:
  - 6 – 60 parameters calibrated, 100 to 100,000 model evaluations
  - Excellent performance relative to other algorithms
- Yen et al. (2014):
  - Compared 6 calibration algorithms and concluded DDS "surpassed all other methods in convergence speed and behavioral statistics"
- Arsenault et al. (2014):
  - Compared 10 global optimization algorithms in 40 case studies
  - Concluded DDS was among top 3 algorithms which "were either as good as or better than the other methods"
- DDS applied above with **same algorithm parameter value**

Arsenault, R., Poulin, A., Côté, P., and Brissette, F. (2013). "A comparison of stochastic optimization algorithms in hydrological model calibration." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000938.

Yen, H., Jeong, J., Tseng, W., Kim, M., Records, R., and Arabi, M. (2014). Computational Procedure for Evaluating Sampling Techniques on Watershed Model Calibration., *J. Hydrol. Eng.* 10.1061/(ASCE)HE.1943-5584.0001095

WATERLOO ENGINEERING

Common Questions and Recommendations when Using DDS

- Should I do a sensitivity analysis to choose a reduced set of calibration parameters?

  … not critical for DDS. Performance will only degrade if there are many (>25%) insensitive parameters.

- Should I try to adjust the value of DDS parameter to something other than *r* = 0.2?

  *… **NO**. Just perform multiple calibration trials and use best solution. When solving real problem, use 2 or 3 calibration trials in case unlucky.*

WATERLOO ENGINEERING

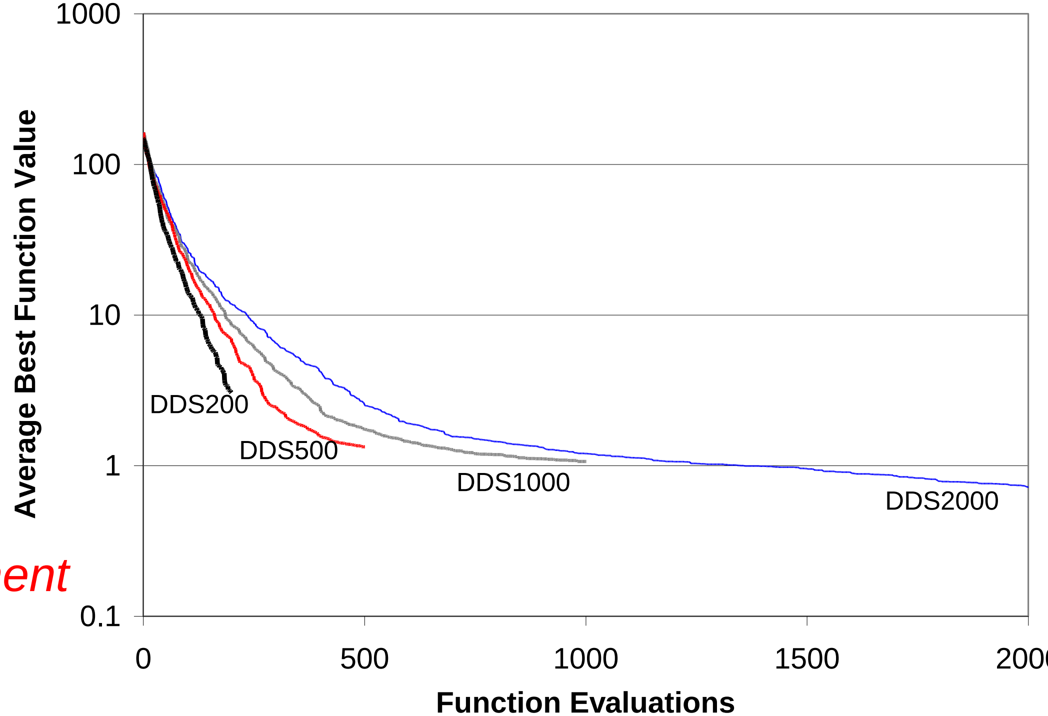
Common Questions and Recommendations when Using DDS

- How many model evaluations (objective function evaluations) should I use?

  … Depends on when you want an answer and avg. model run time

  … 1000 has worked for me calibrating dozens of parameters

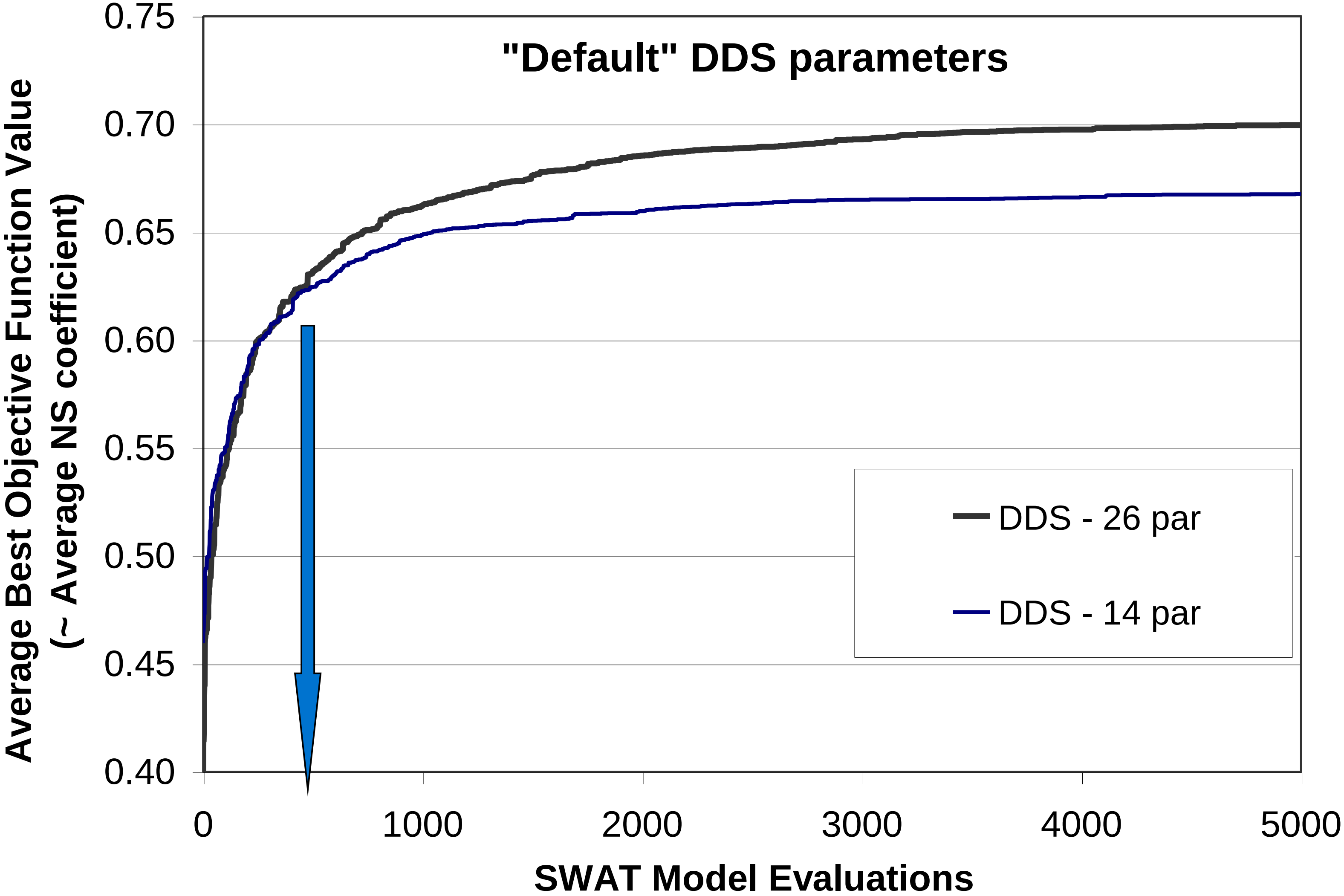
Observed DDS algorithm behaviour as user computational limit changes (not always like this though)



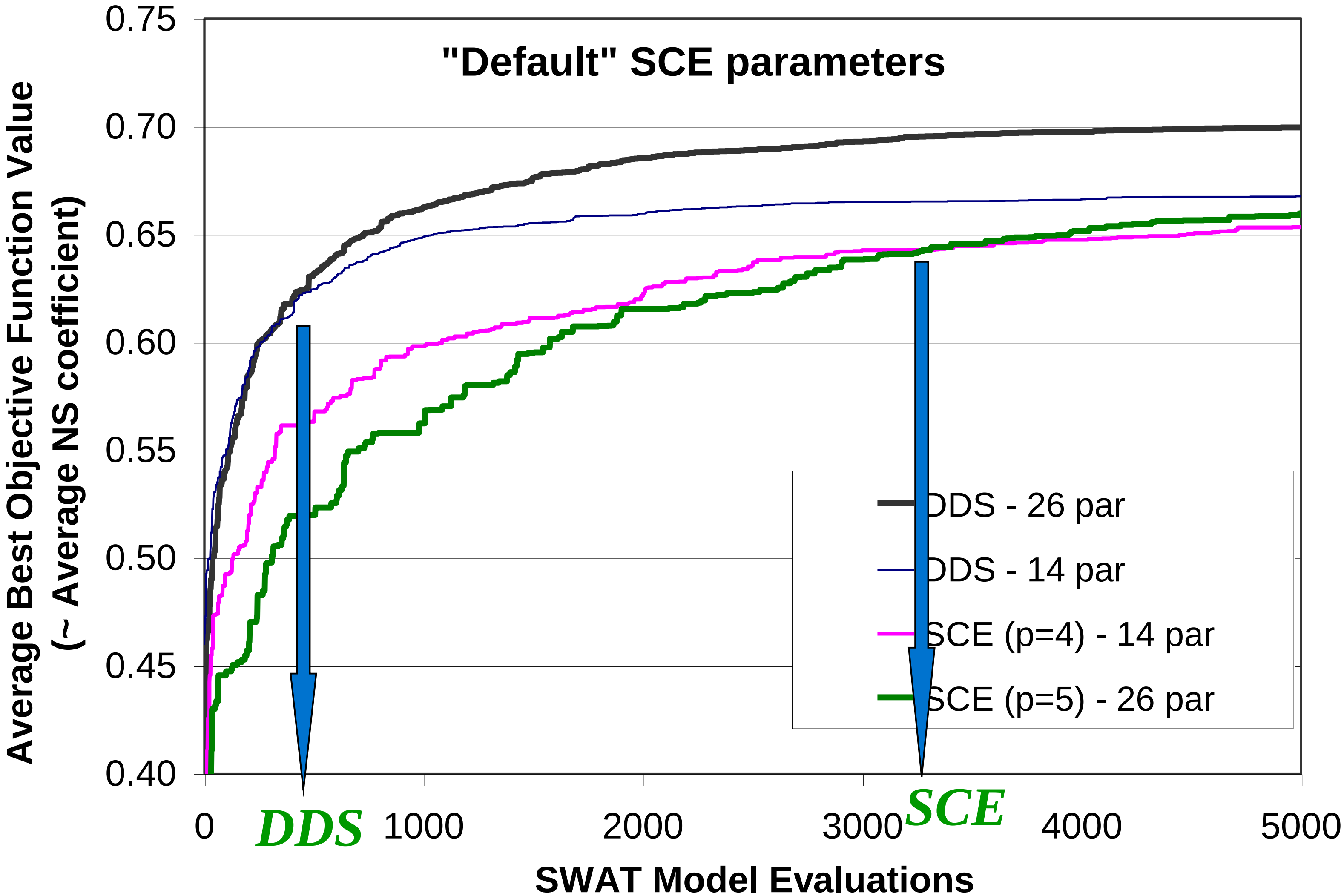
… try a synthetic calibration experiment

WATERLOO ENGINEERING

Results – Giving Algorithm More Parameters to Fit Data



Results – Giving Algorithm More Parameters to Fit Data



Surrogate Modelling Case Studies

Case Study	Run-time (min)	Number of Parameters	Details
SWAT-1	2	14	SWAT2000 watershed model of the Cannonsville Reservoir watershed, New York (streamflow calibration)
SWAT-2	1	30	SWAT2000 watershed model of the Cannonsville Reservoir watershed (calibration to streamflow, total suspended sediment, & total phosphorous)
MESH	5	62	MESH hydrologic model of Reynolds Creek Watershed, Idaho (streamflow calibration)
FEFLOW	0.5 to 15	56	Pilot Point based Calibration to Transient Drawdowns - Three-layer confined groundwater model, northeast Alberta
DFRTT	40	7	Dipole flow and reactive tracer test interpretation model (Groundwater), Borden, Ontario

Pre-emption logic:

- *DDS pre-emption threshold is dynamically set to the current best objective function value after each iteration:*

$$SSE^* = SSE_{best}$$

- *Some code must be written to monitor pre-emption:*

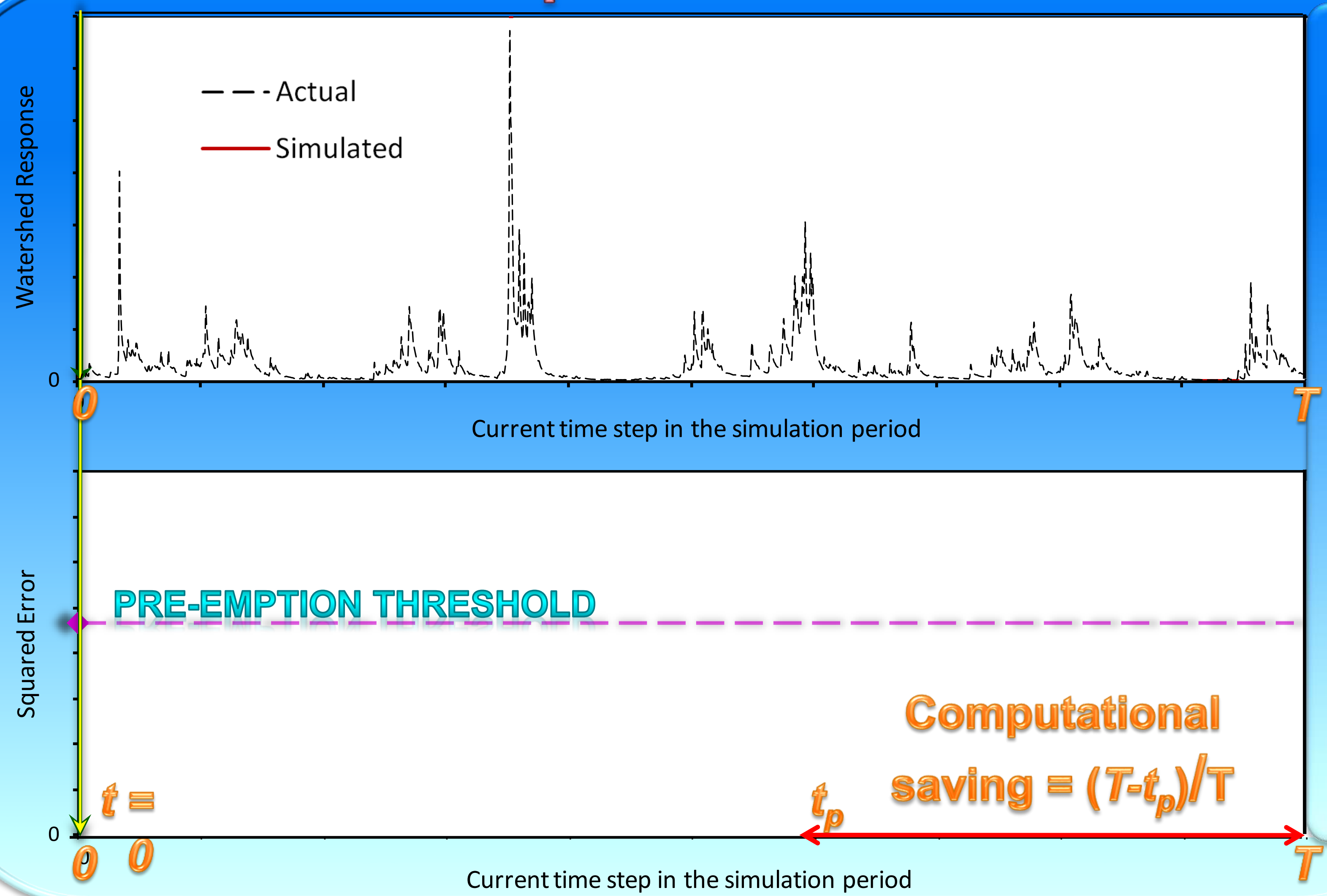
```
FOR t=1 to T                [time period loop in simulation model]

    IF SSE(t) ≥ SSE*
        TERMINATE MODEL SIMULATION
    END

END
```

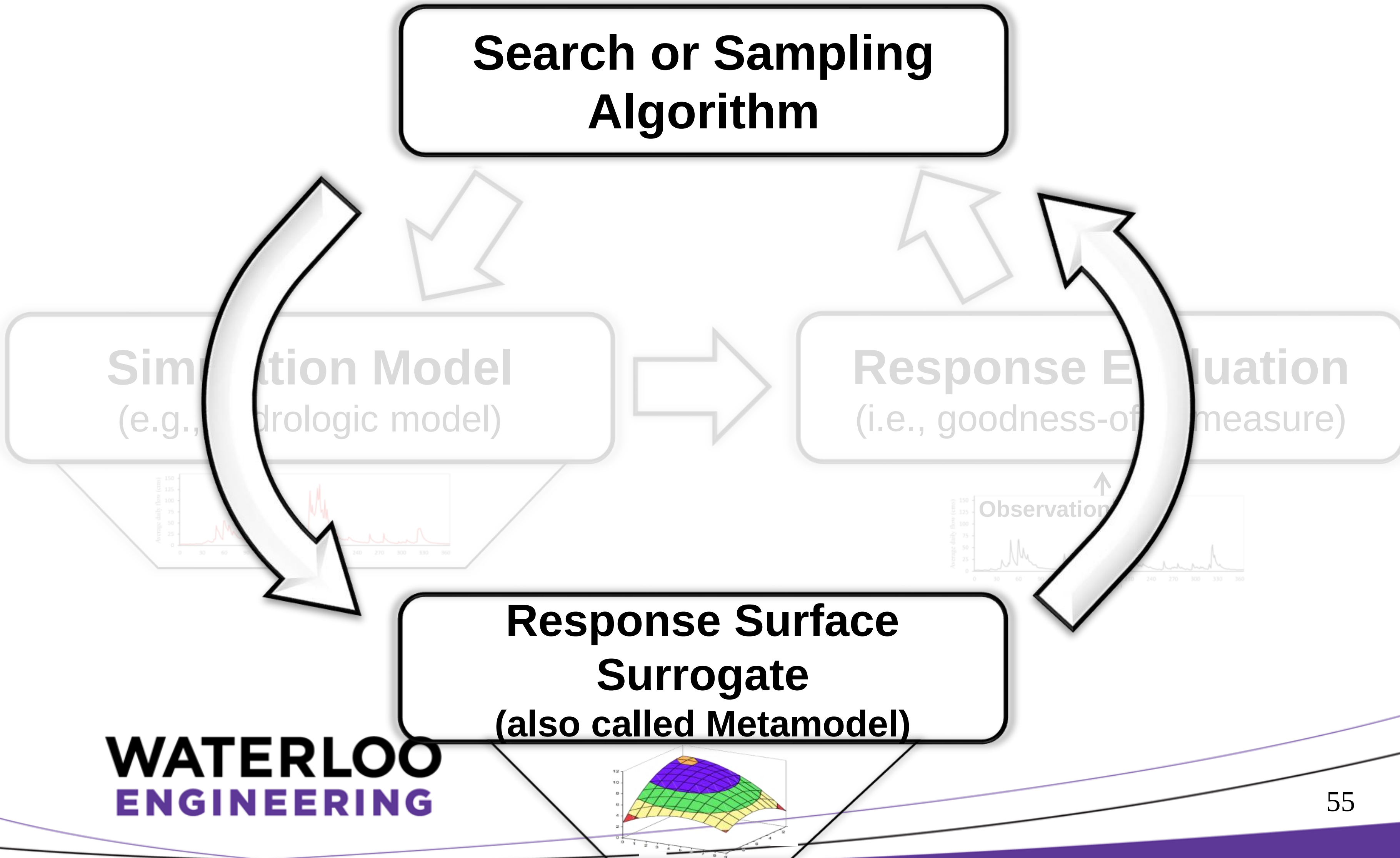
WATERLOO ENGINEERING

An illustrative example



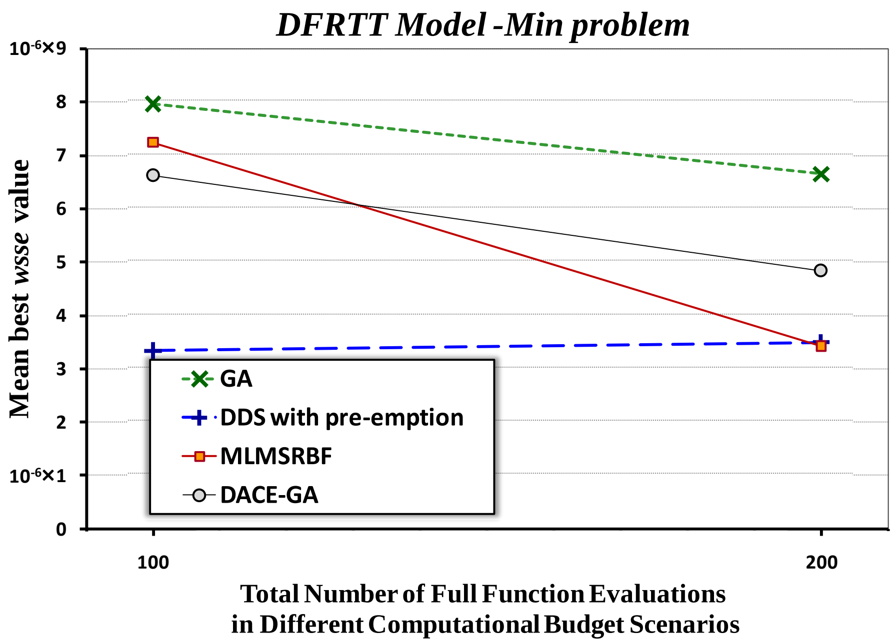
WATERLOO
ENGINEERING

Response Surface Surrogates



WATERLOO ENGINEERING

Sample Results of Numerical Experiments



Dipole flow and reactive tracer test (DFRTT) Model calibration problem:

- calibrate 7 aquifer parameters
- 37 minute runtime

WATERLOO ENGINEERING