Model Abstraction Techniques: an Overview of Applications in Contaminant Hydrology

Yakov Pachepsky
Martinus van Genuchten

ARS-USDA
USDA

Diederik Jacques

Belgian Nuclear
Research Centre

Andrey Guber

Lomonosov Moscow
State University

Marcel Schaap
Jiri Simunek

University of California Riverside

Ralph Cady
Thomas Nicholson

Office of Research
U.S. Nuclear Regulatory
Commission

Layout

1. Why model abstraction?
2. Review of model abstraction techniques
3. Case study for soil water flow in a humid region
4. What did we learn so far?

What is model abstraction?

Model abstraction is a methodology

for <u>reducing the complexity</u> of a simulation model

while <u>maintaining the validity</u> of the simulation results

<u>with respect to the question</u> that the simulation is being used

to address

Why model abstraction?

Simulations of complex engineering and military systems show (Frantz, 2002) that:

- **Increased level of detail *does not necessarily imply* increased accuracy.**

- **Increased computational complexity *does not necessarily* imply increased accuracy.**

- **Increased level of detail usually *does imply* increased computational complexity.**

- **Computational complexity *does imply* increase in computer runtime**

Why model abstraction?

In flow and transport simulations, the issue of data collection and parameter estimation is as essential as computational complexity.

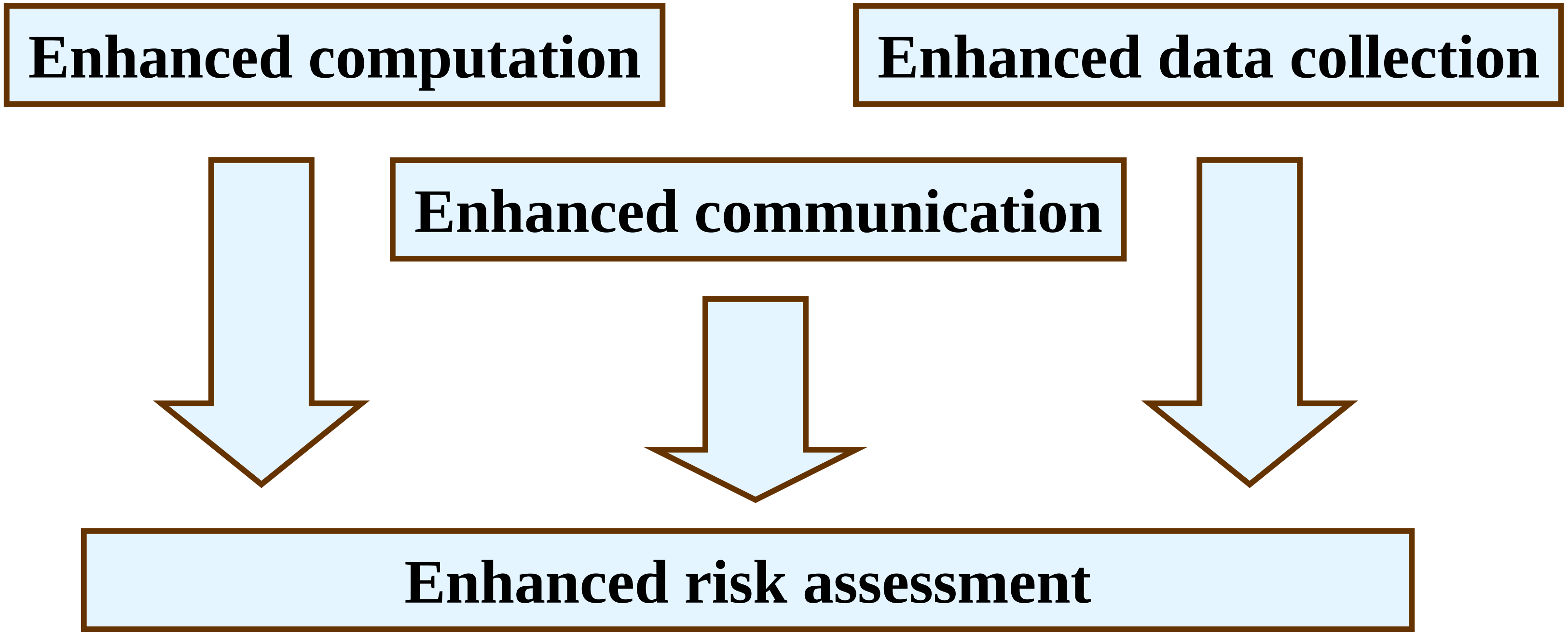
- **Increased level of detail *does not necessarily imply* increased accuracy.**
- **Increased data collection density *does not necessarily imply* increased accuracy.**
- **Increased level of detail usually *does imply* increased data collection density.**
- **High data collection density is usually *prohibitive* in large-scale projects.**

Progress in computational speed is more substantial than in data collection.

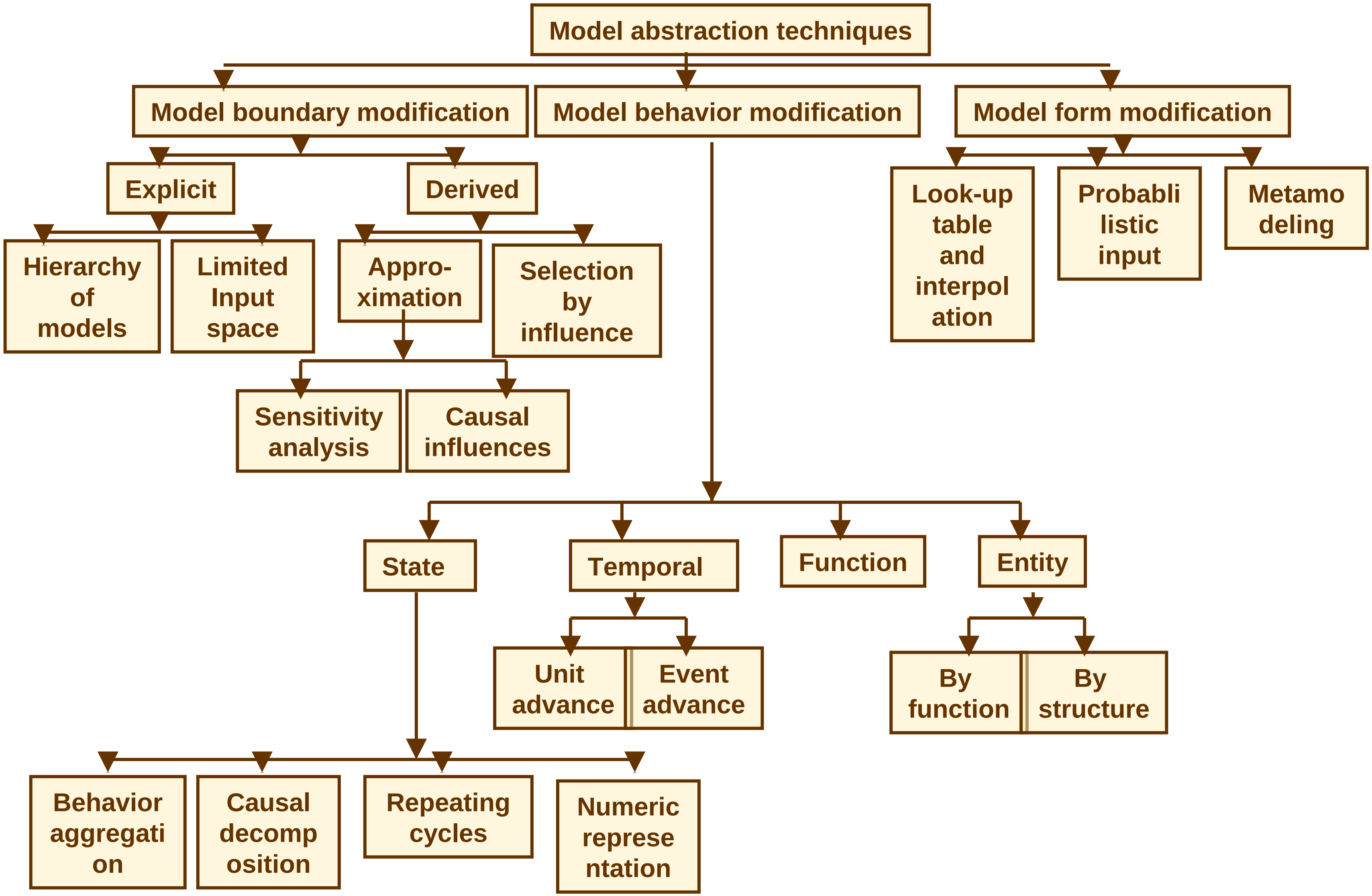
Why model abstraction?

- more easily understandable description of the problem
- discussions to be focused on the most important aspects.
- it is often imperative to explicitly acknowledge the abstraction strategy.
- model abstraction explicitly deals with uncertainties.

Why model abstraction?



General taxonomy of model abstraction techniques (Caughlin and Sisti, 1997; Frantz, 2002)



MODEL ABSTRACTION IN FLOW AND TRANSPORT SIMULATIONS

Boundary	Hierarchy
Behavior	Input space
Form	Approximation

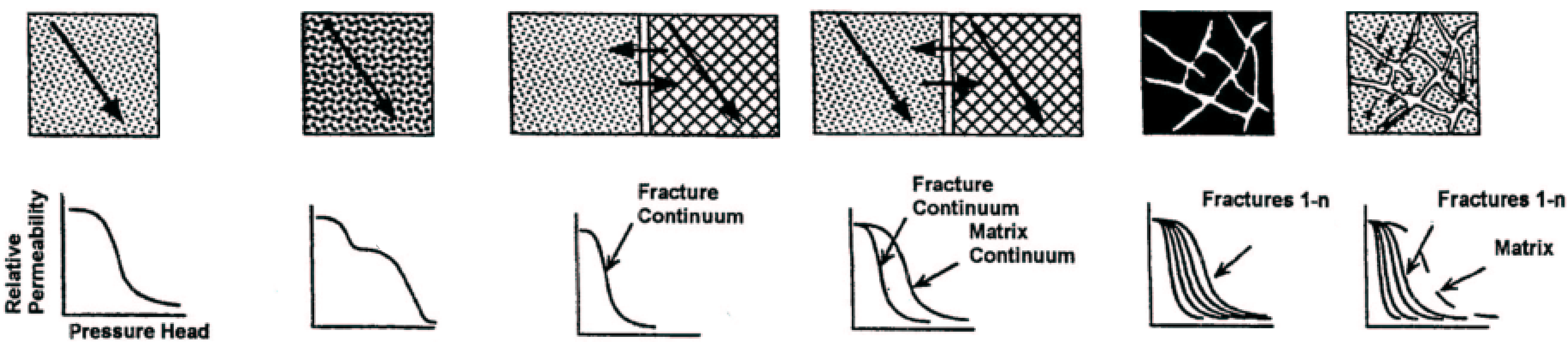
A pre-defined hierarchy of models includes simplified and more complex models.

Models to simulate flow and transport in structured soils or unsaturated fractured rock (Altman et al., 1996)

Single continuum | Equivalent matrix and fracture continuum | Dual porosity | Dual permeability | Discrete fractures without matrix | Discrete fractures with matrix

Dual porosity: **Matrix Fracture**

Dual permeability: **Matrix Fracture**



Complexity →

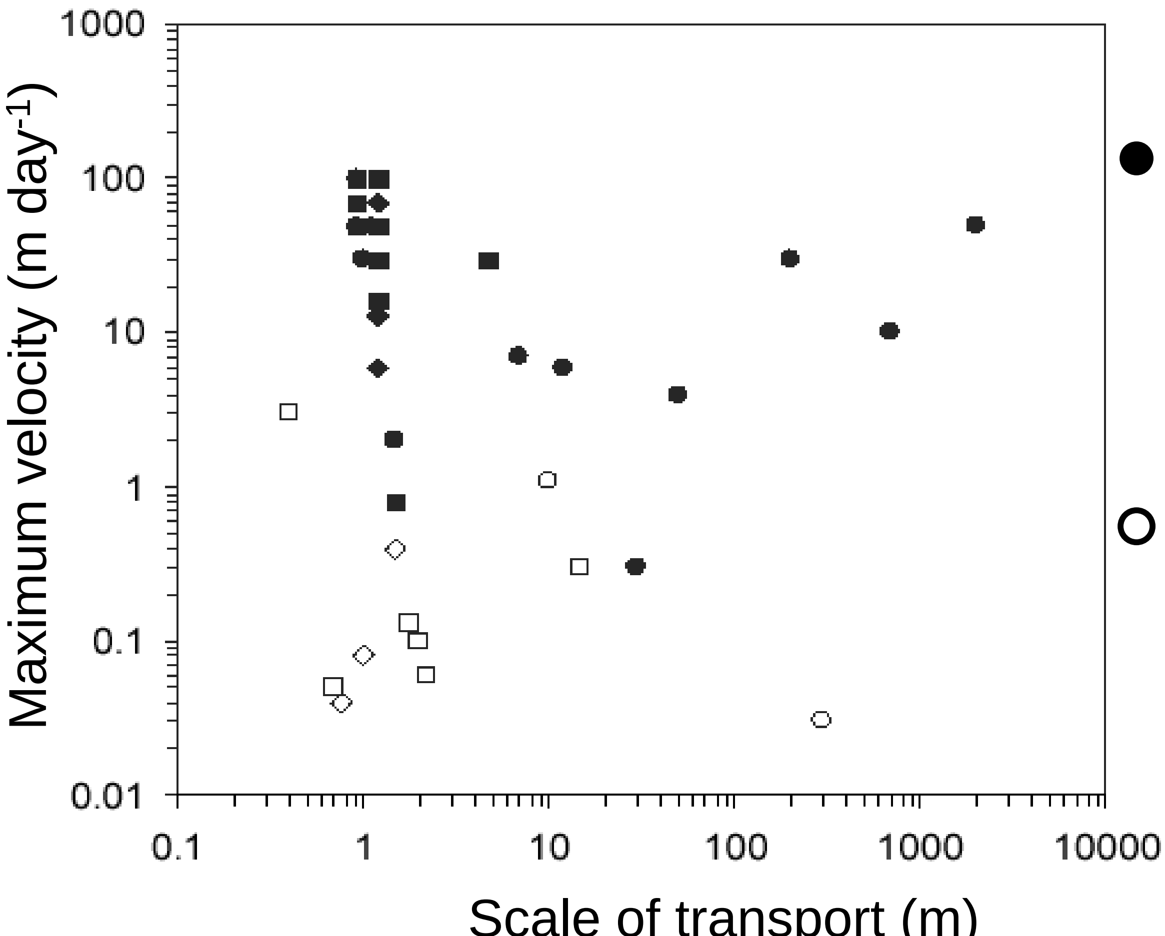
← **Model abstraction**

MODEL ABSTRACTION IN FLOW AND TRANSPORT SIMULATIONS

Boundary	Hierarchy
Behavior	Input space
Form	Approximation

Simplifying process descriptions based on the specific range of input parameters

Maximum tracer velocity in 34 documented experiments (Nimmo, 2002)



● **Continuous water supply = preferential flow**

○ **Intermittent water supply = no preferential flow**

MODEL ABSTRACTION IN FLOW AND TRANSPORT SIMULATIONS

Boundary	Hierarchy
Behavior	Input space
Form	Approximation

Parameter and process elimination based on simulation results.

- **sensitivity analysis**
- **correlation in model outputs**
- **…**

MODEL ABSTRACTION IN FLOW AND TRANSPORT SIMULATIONS

Boundary	
Behavior	**Behavior aggregation**
Form	Causal decomposition
	Aggregation of cycles
	Numerical representation
	Temporal aggregation
	Entity aggregation

Combining system states whose distinctions are irrelevant to the simulation output

- **combining individual stream tubes/flowpaths,**
- **upscaling with stochastic averaging**
- **aggregating soil or sediment layers into one "general" layer**
- **aggregating individual plants into one vegetation layer**
- **…**

MODEL ABSTRACTION IN FLOW AND TRANSPORT SIMULATIONS

Boundary
Behavior
Form

Behavior aggregation
Causal decomposition
Aggregation of cycles
Numerical representation
Temporal aggregation
Entity aggregation

Dividing a model into loosely connected components, executing each component separately, and searching for constraints that execution of one component can impose on other components. Where no causal relationship exists, the components may be executed in parallel.

- **simulating water flow independently on solute transport**
- **individual-based modeling**
- **“smart agent” models**
- **modeling flow with explicit temporal approximations**
- **…**

MODEL ABSTRACTION IN FLOW AND TRANSPORT SIMULATIONS

Boundary	
Behavior	Behavior aggregation
Form	Causal decomposition
	Aggregation of cycles
	Numerical representation
	Temporal aggregation
	Entity aggregation

Combining states that reflect similar sequences; distinctions among the individual sequences are irrelevant to the final outcome.

- **solute transport modeling with continuous-time random walk**
- **solute transport modeling with diffusion-limited aggregation**
- **…**

Boundary	
Behavior	Behavior aggregation
Form	Causal decomposition
	Aggregation of cycles
	Numerical representation
	Temporal aggregation
	Entity aggregation

Replacing continuous variables by class variables. For example, combining states characterized by floating point numbers that round to the same integer value

Class pedotransfer functions - grouping for parameter estimation

Example: Ksat for US soils (Rawls et al., 1998)

Textural class	High porosity			Low porosity		
	25%	50%	75%	25%	50%	75%
Sand	638.4	436.8	230.4	523.2	218.4	153.6
Fine sand	566.4	338.4	283.2	528.0	240.0	163.2
Loamy sand	468.0	295.2	201.6	184.8	98.4	74.4
Loamy fine sand	292.8	148.8	86.4	278.4	28.8	16.8
Sandy loam	312.0	134.4	72.0	74.4	31.2	12.0
…	…	…	…	…	…	…

MODEL ABSTRACTION IN FLOW AND TRANSPORT SIMULATIONS

Boundary
Behavior
Form

Probabilistic input
Look-up tables
Rule-based solution
Metamodeling

Machine-learning
Spatial structure

Neural nets
CART
Projection
 pursuit
GMDH

How to direct the model abstraction

The simpler models is that model which has the smaller number of independent parameters to measure/estimate
and/or the lesser amount of computations.

Simplicity may be related to:

- the number of processes being considered explicitly
- process descriptions
- spatial discretization/scale
- temporal discretization/scale
- number of measurements for parameter estimation
- correlations among parameters
- speed of computations
- data pre-processing and post-processing

How to assess model abstraction results?

Validity of the simu… …e accuracy of the simulations.

Validity can be relat…

- variability in data against which…
- uncertainty in the simulation results
- physically meaningful results with …
- information content of the model output

A model cannot be proven to be more accurate than… against which… model is bein…

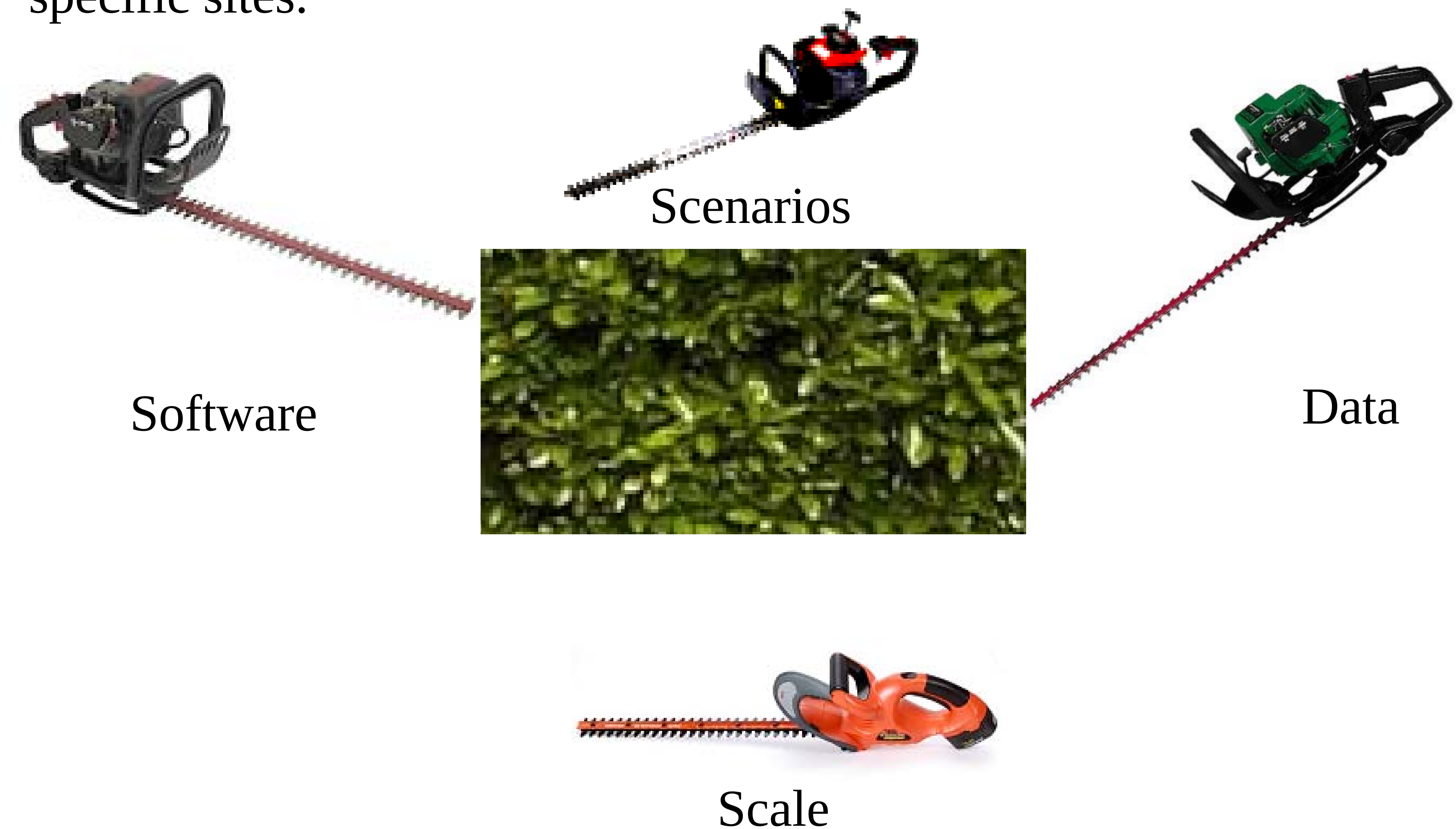
Ensure that the model output uncertain… exceed u… dat…

Monte Carlo simulation of…

Accuracy can be the same but the complexity of simulation results can be different

Where to start the model abstraction process?

Model abstraction techniques are always meant to be applied to specific sites.



Case study

Objective: to understand
how model abstraction can affect
performance assessment of contaminant migration
at a relatively humid site
where transport may be affected
by the presence of soil macropores
and related preferential flow phenomena

Experimental setup – Bekkervoort, Belgium

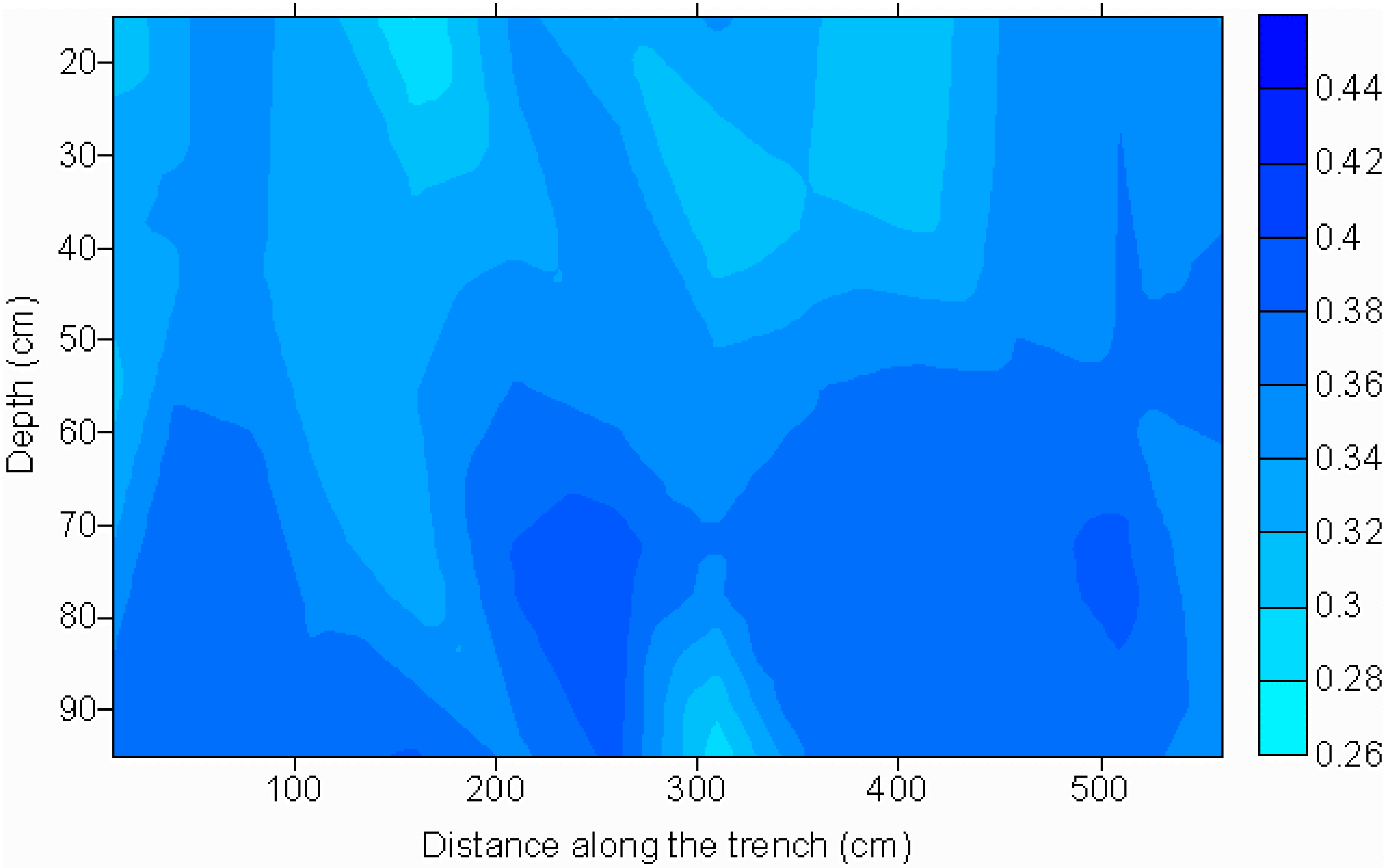
Overview of the database

Transient Flow Conditions, 384 days

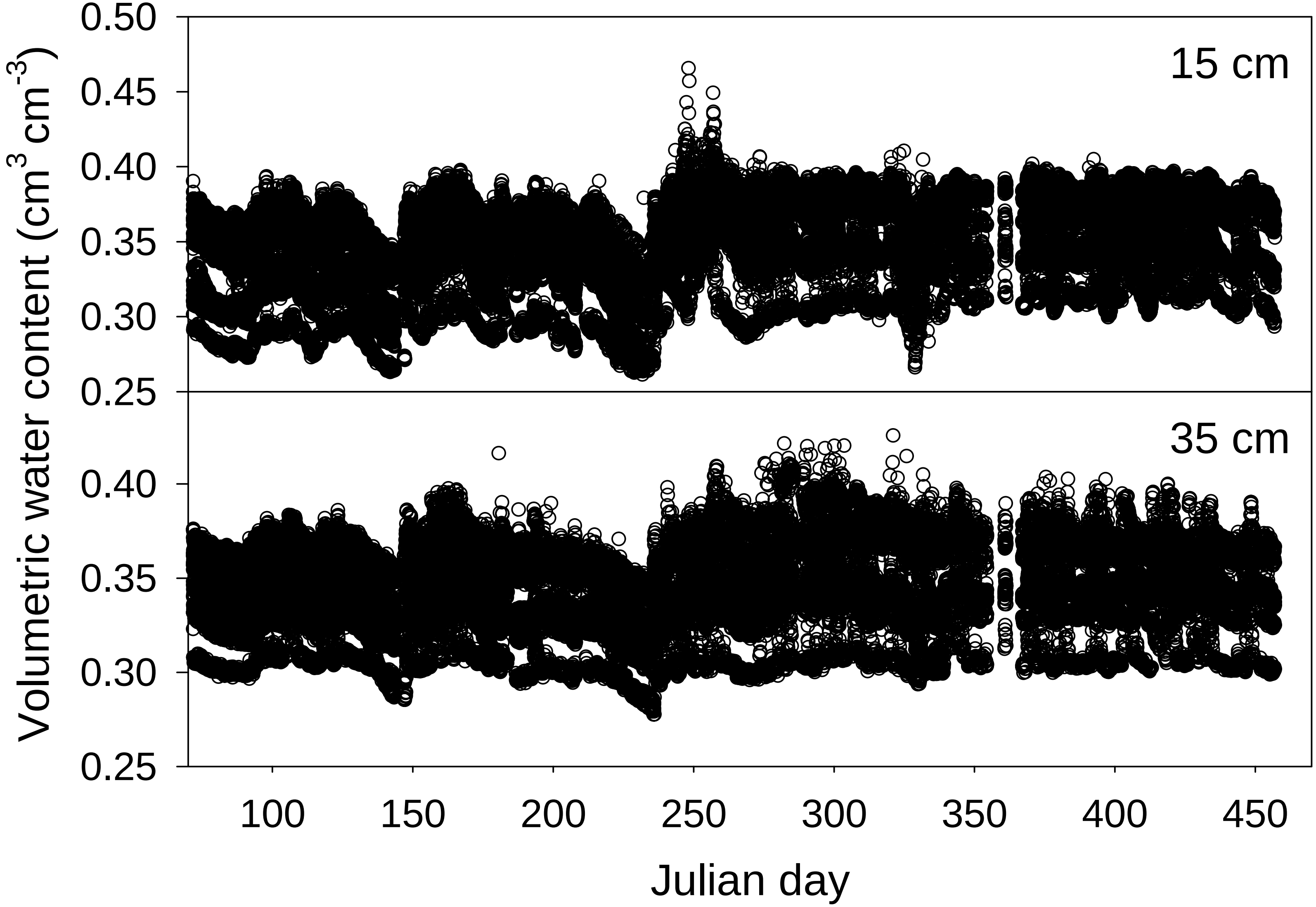
Variable	Device	Frequency	Total positions
θ	TDR-probes	Every 2 hr	5 d x 12p=60
C^r	TDR-probes	Every 2 hr	5 d x 12 p=60
ψ	Tension cups	Every hr	5 d x 12 p=60
T	Temperature probes	Every hr	5 d x 3 p =15
Water fluxes	Passive Cap. Lys.	Every 2/3 days	2 d x 3 p = 6
Solute fluxes	Passive Cap. Lys.	Every 2/3 days	2 d x 3 p = 6
Rainfall	Pluviograph	Continuously	

400,000 measurements

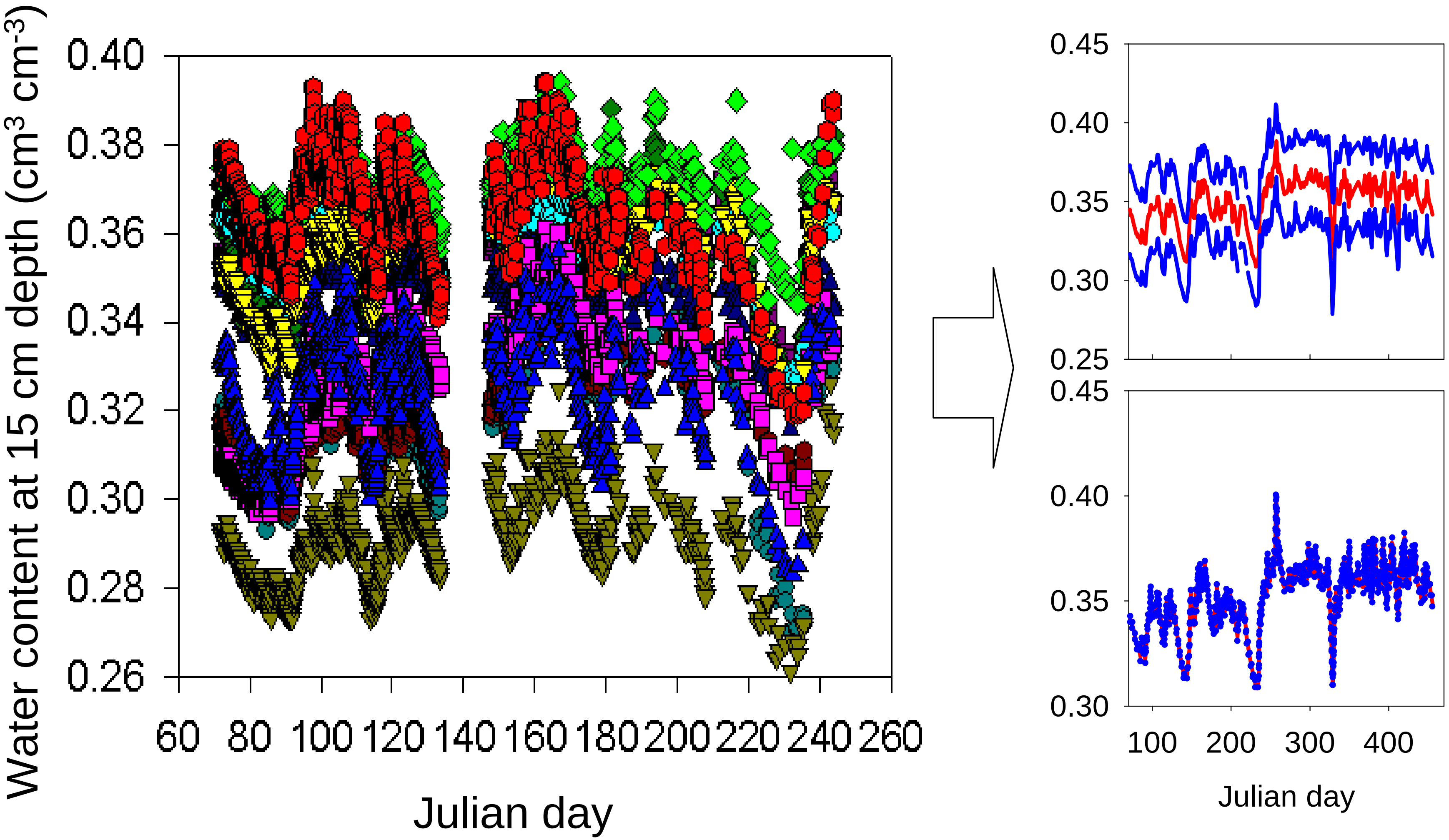
Example of one month of water contents observations



Soil water contents



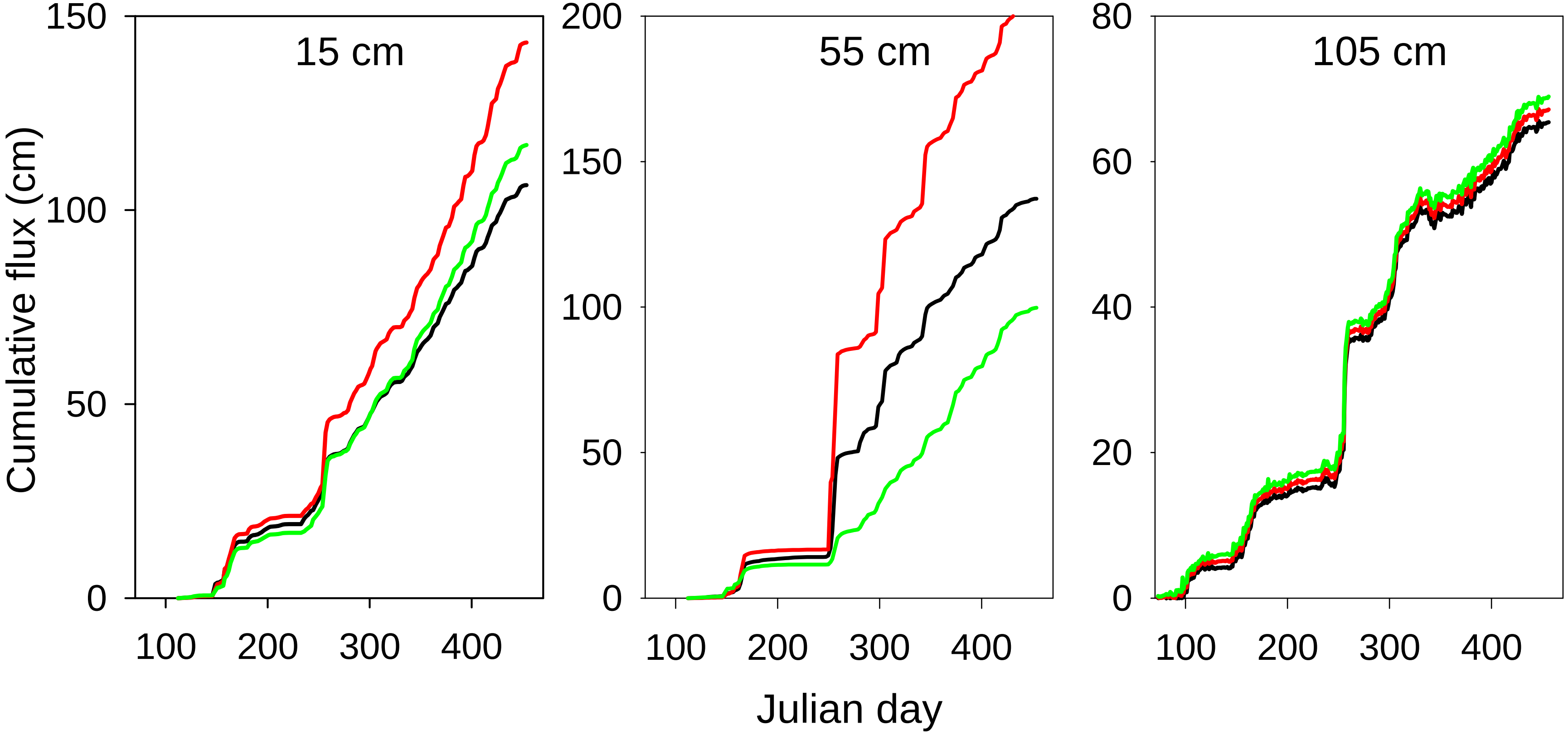
Temporal persistence in soil water contents



Soil water fluxes

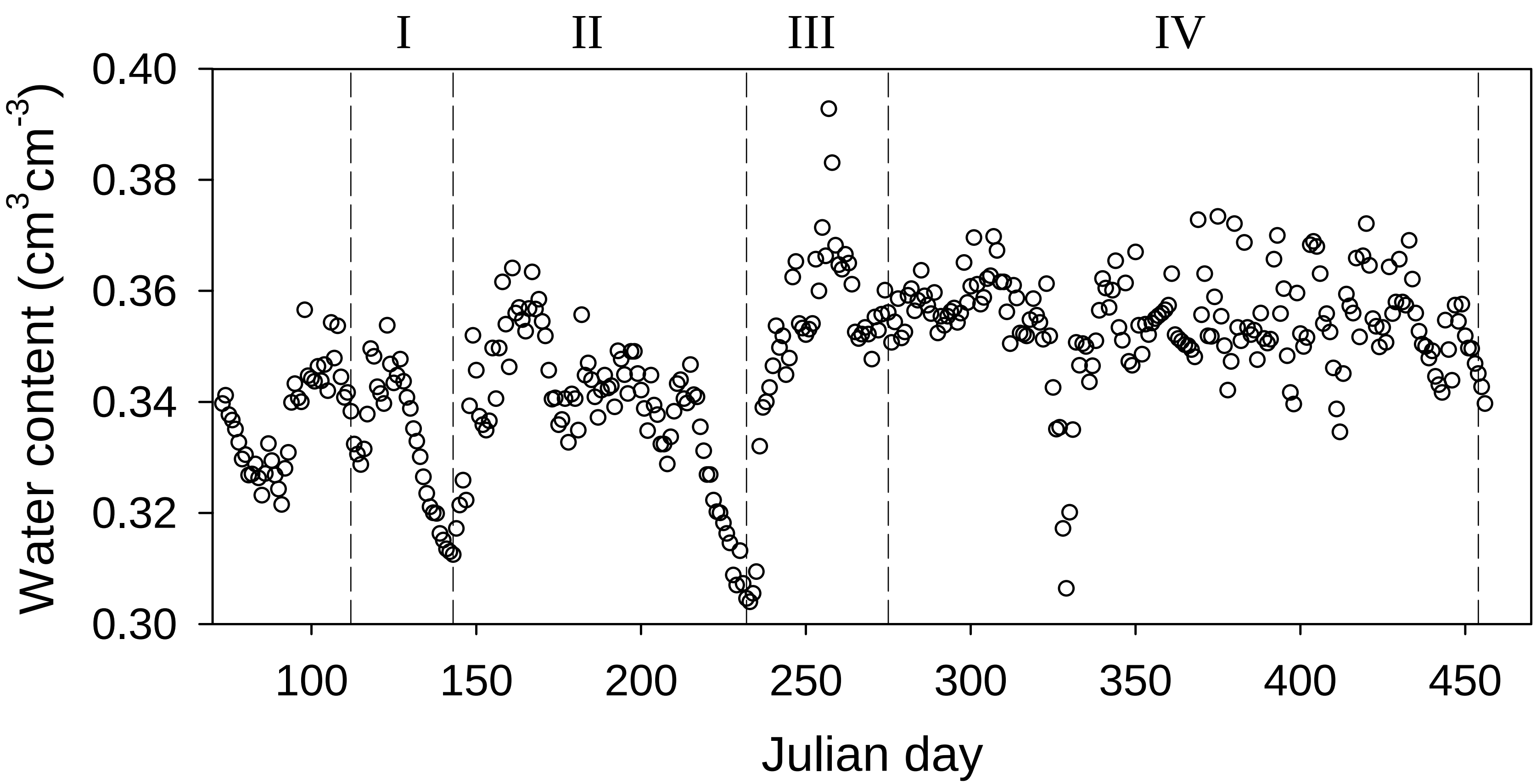
Passive capillary lysimeters
(~ Glendon Gee, PNNL)

Mass balance approach
(~NUREG/CR 6653)

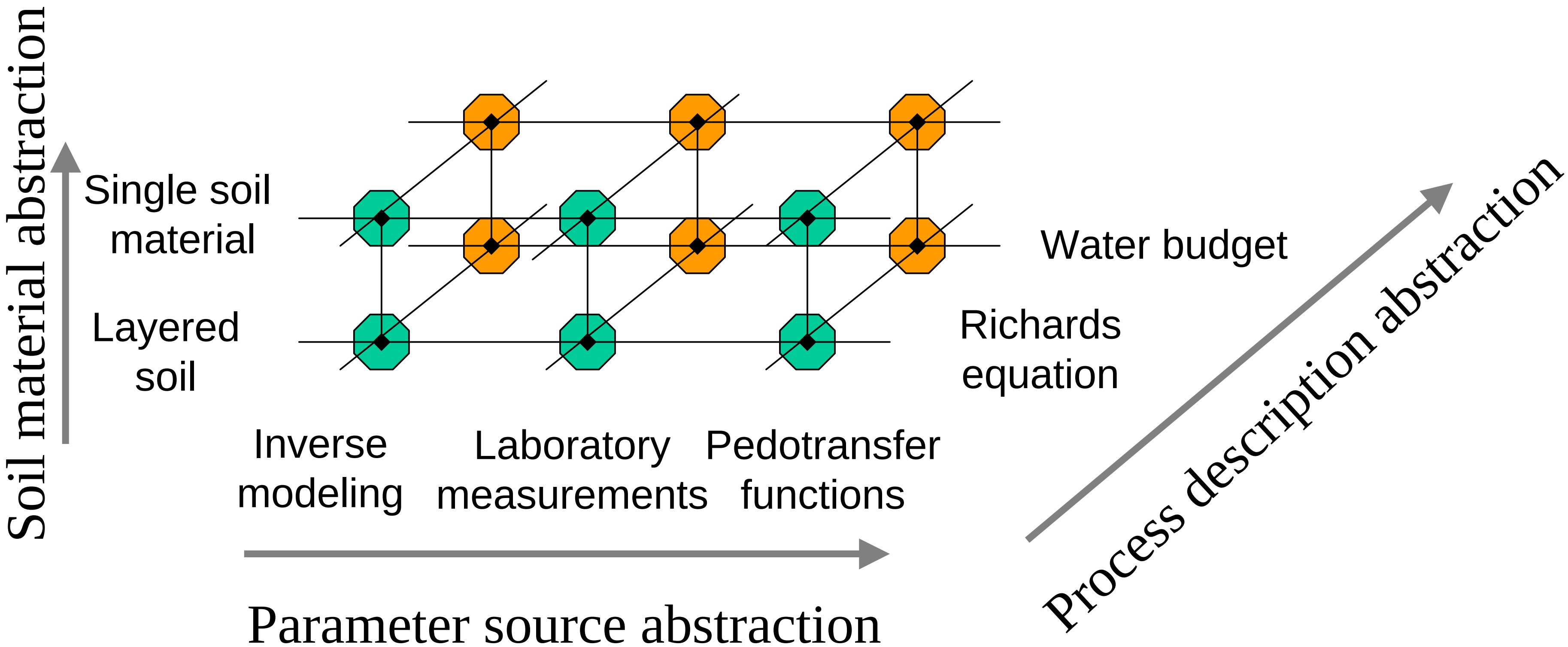


Question to be addressed

Estimate the cumulative water fluxes at capillary sampler depths and through the bottom of soil profile over four drying-wetting cycles



Design of model abstraction via model boundary and behavior modification



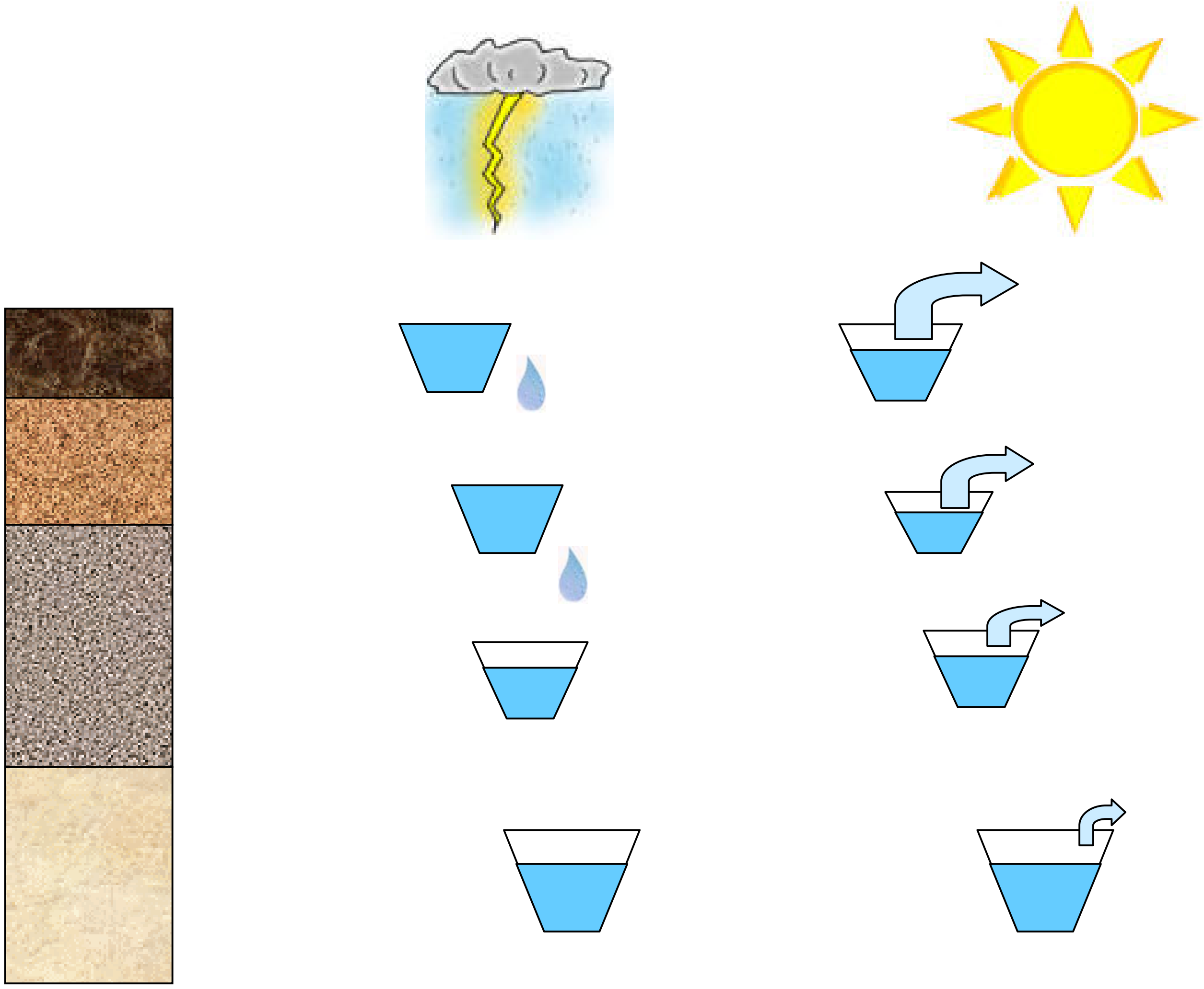
HYDRUS

MWBUS (Model of Water Budget in Unsaturated Soil)

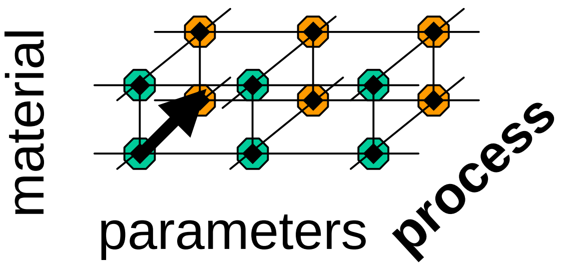
50 Monte Carlo simulations to estimate uncertainty

The MWBUS model

Extension of the to the PNNL water budget model (NUREG/GR 6565)

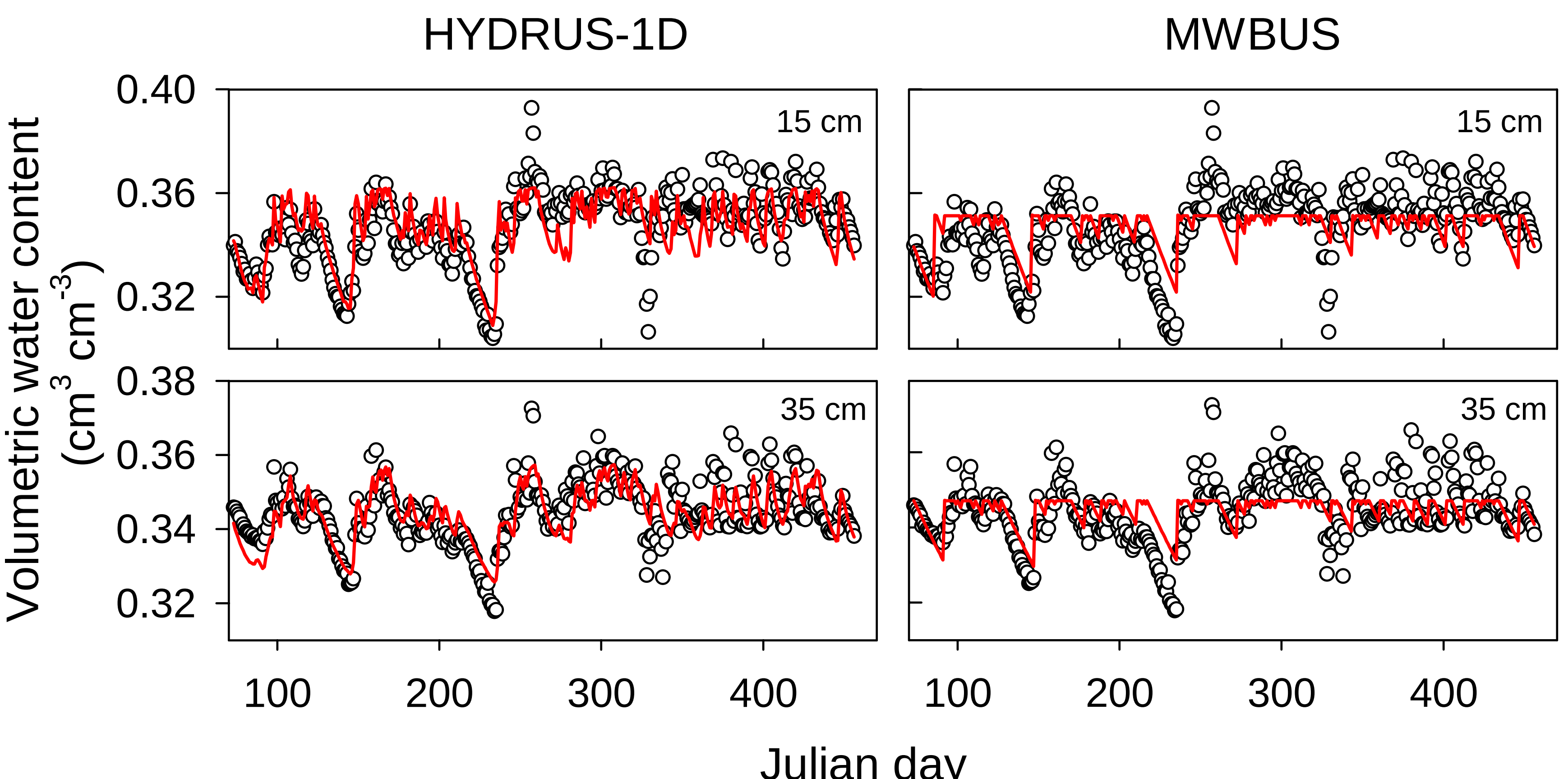


Process description abstraction

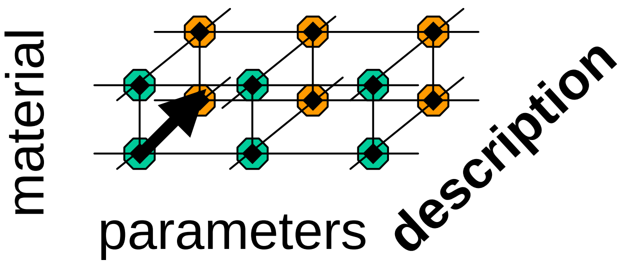


Richards equation ➜ water budget
(inverse modeling, layered soil)

Soil water contents

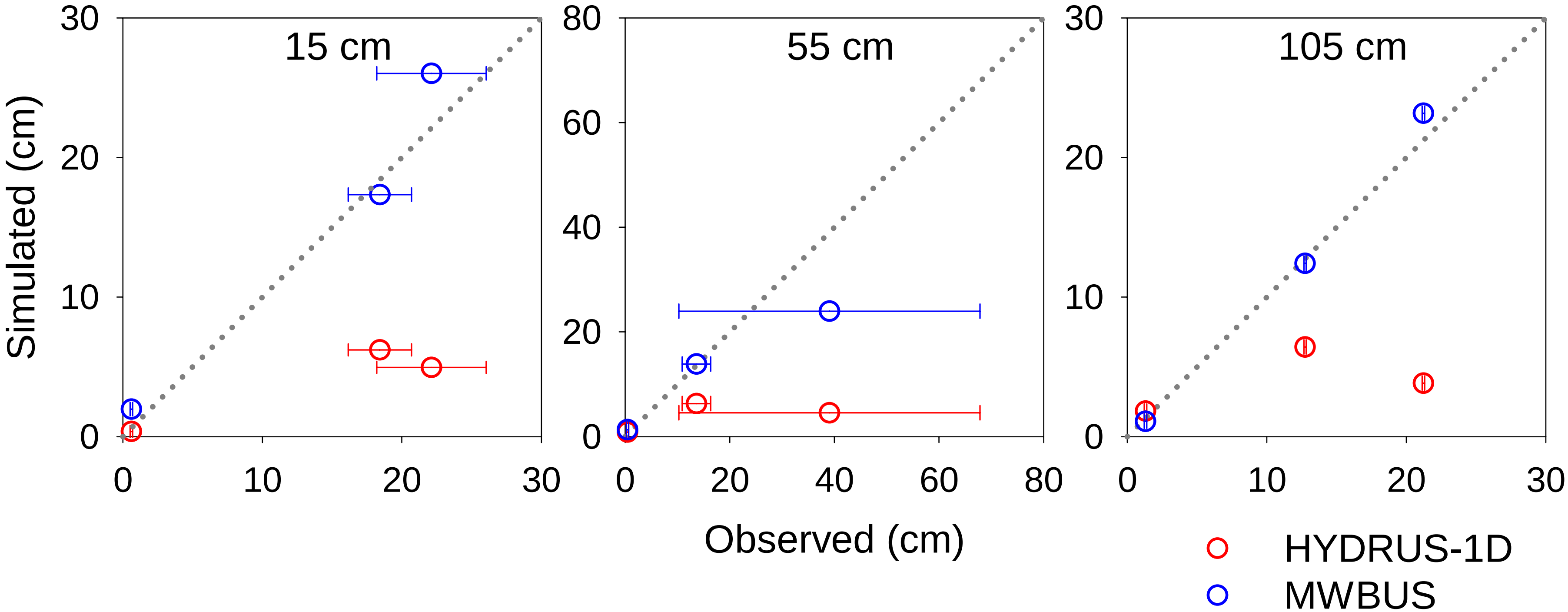


Process description abstraction

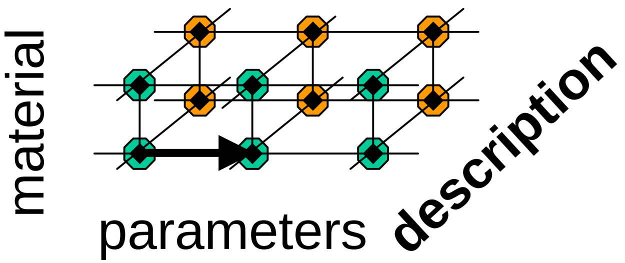


Richards equation ➜ water budget
(inverse modeling, layered soil)

Soil water cumulative fluxes



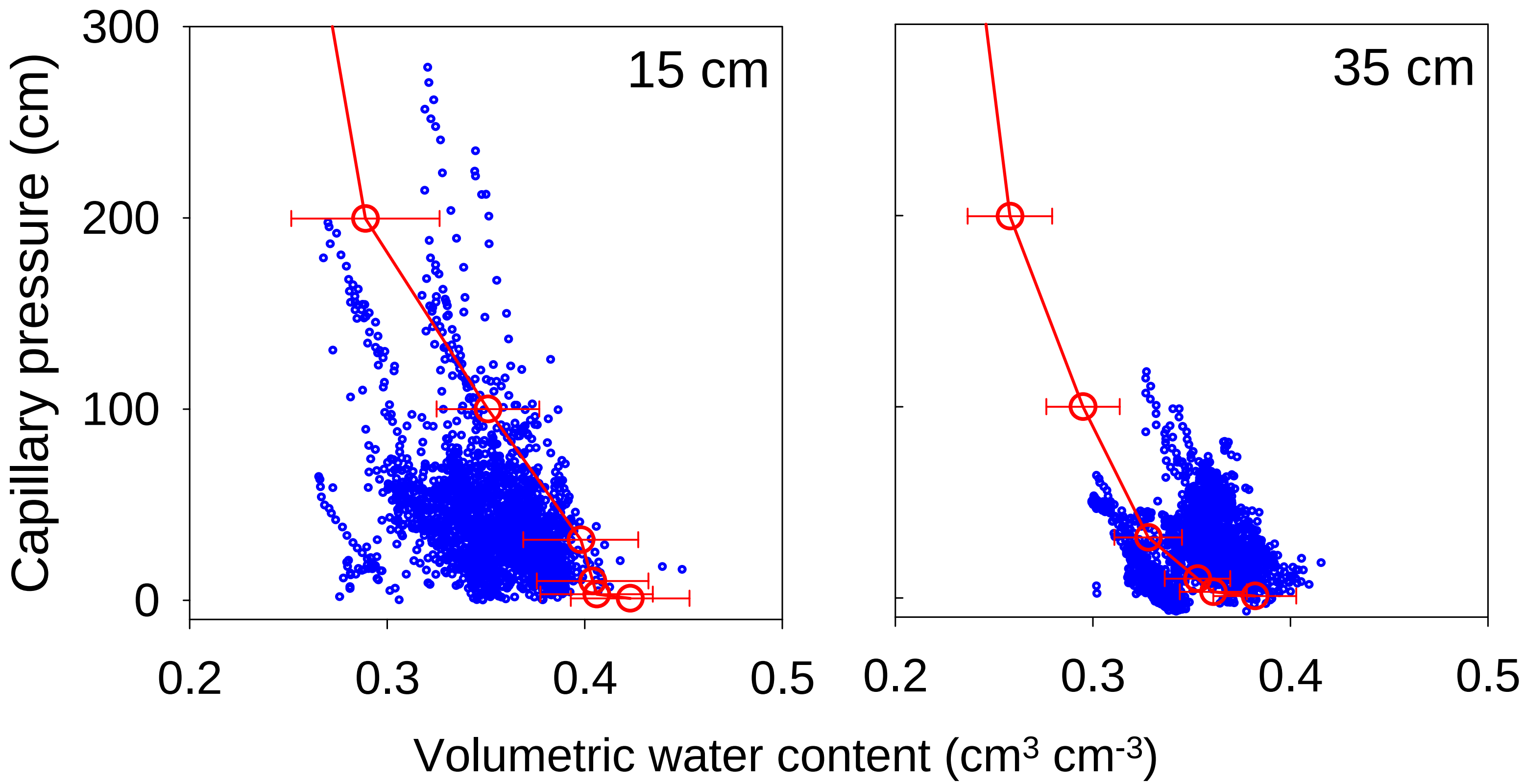
Parameter source abstraction – lab data



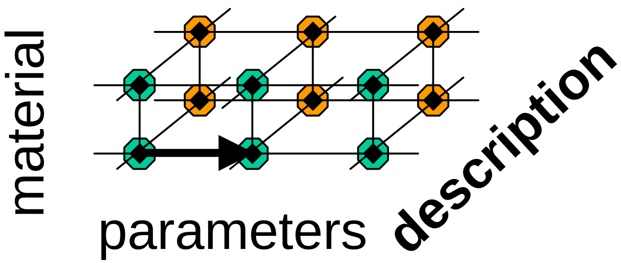
Inverse modeling ➜ lab data

(Richards equation, layered soil)

Field and lab water retention

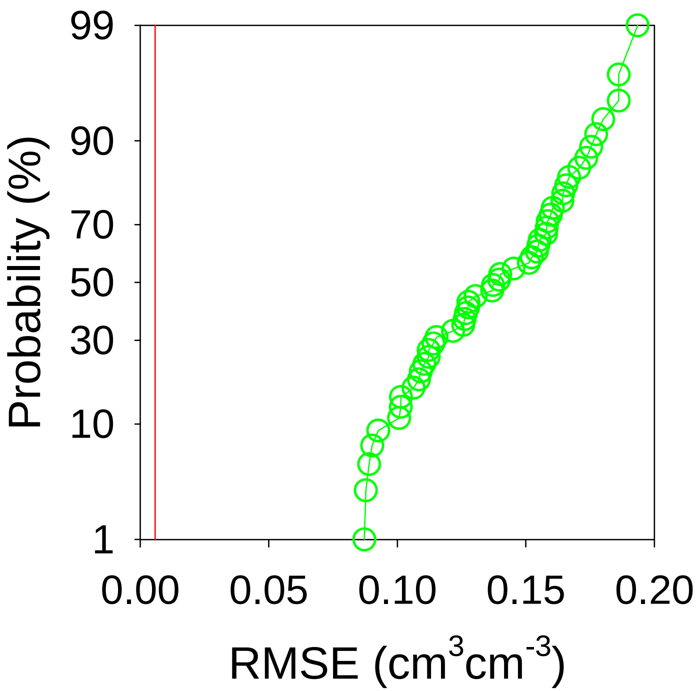


Parameter source abstraction – lab data

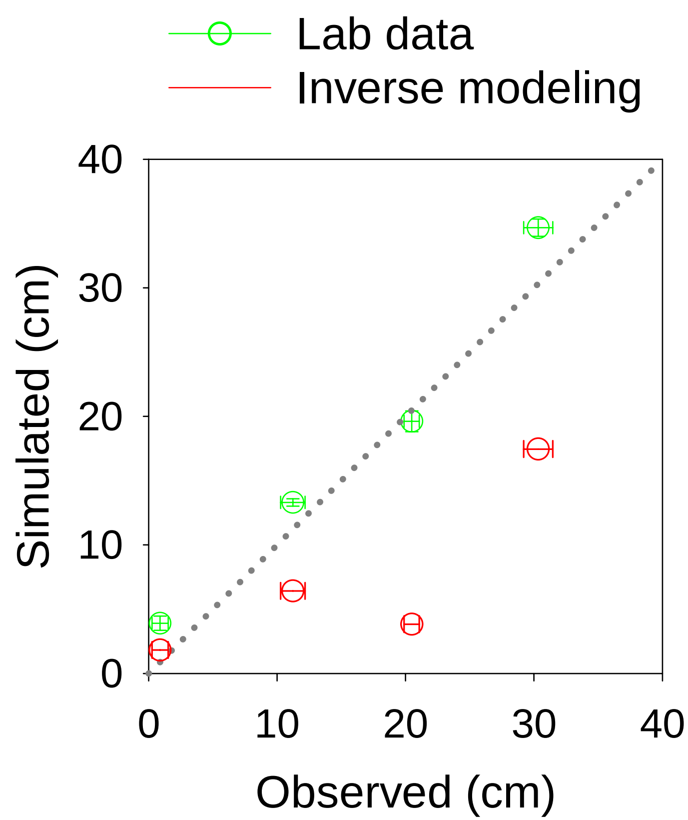


Inverse modeling ➜ lab data
(Richards equation, layered soil)

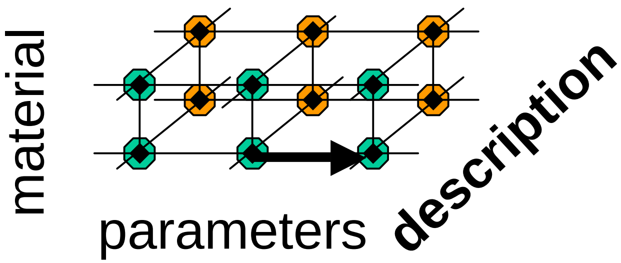
Statistical distributions of
root-mean square errors



Cumulative soil water fluxes

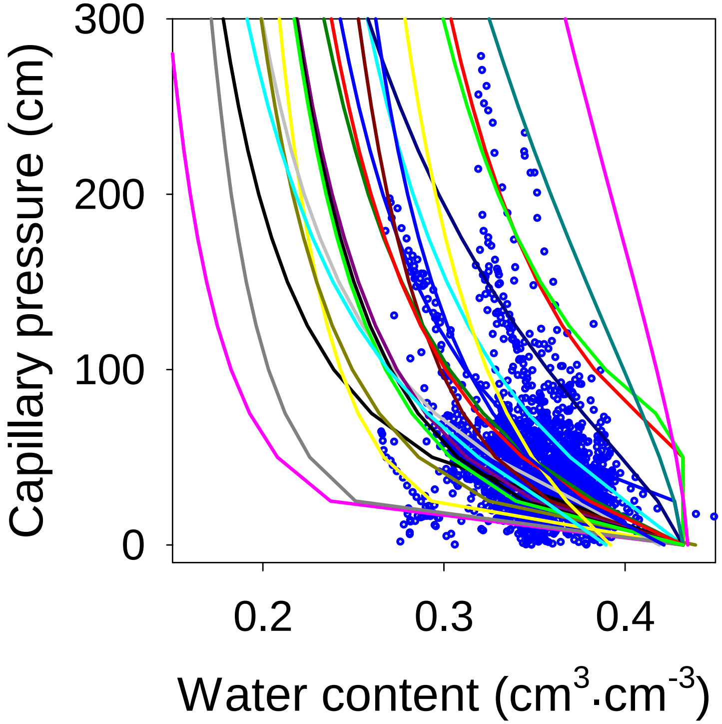


Parameter source abstraction – pedotransfer functions



Lab data ➔ pedotransfer functions
(Richards equation, layered soil)

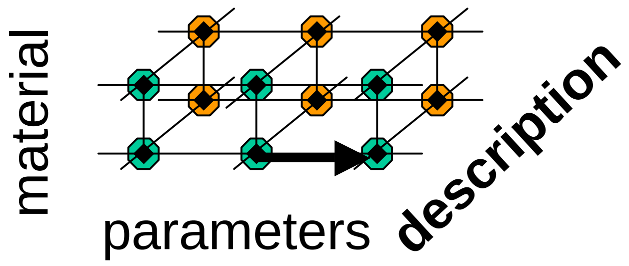
Field and pedotransfer water retention



Pedotransfer sources
(Pachepsky and Rawls, 2005)

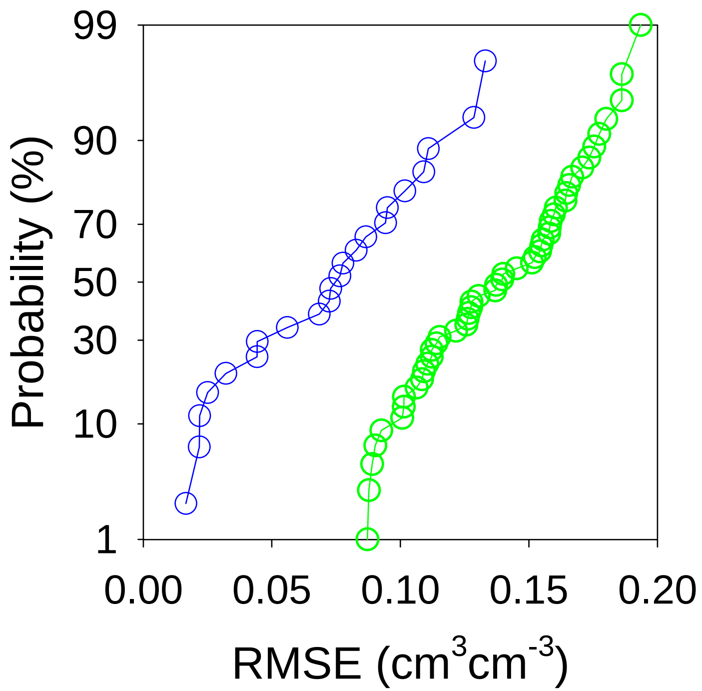
Rawls&Brakensiek	**Saxton et al.**
Williams 1	**Williams 2**
Campbell	**Oosterveldt**
Rawls et al.	**Verekeen et al.**
Gupta&Larson	**Baumer**
Pachepsky et al.	**Canarache**
Peterson	**Rajkai**
Bruand	**Tomasella**
Rawls	**Hall**
Wosten 1	**Wosten 2**
Mayr-Jarvis	**Rosetta**

Parameter source abstraction – pedotransfer functions

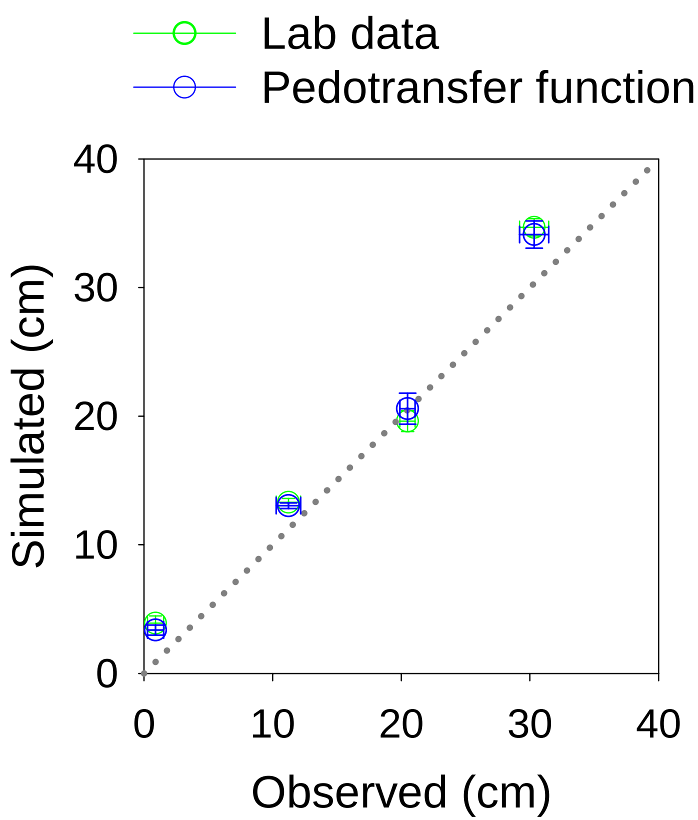


Lab data ➜ pedotransfer functions
(Richards equation, layered soil)

Statistical distributions of root-mean square errors



Cumulative soil water fluxes



Model form abstraction – using neural network

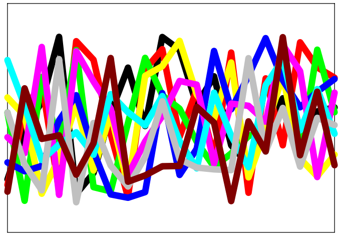
HYDRUS 1D
MWBUS → artificial neural network

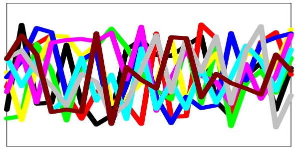
Use weather generator to generate probable weather patterns

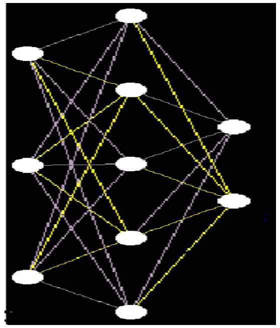
Run Monte Carlo simulations of soil water flow

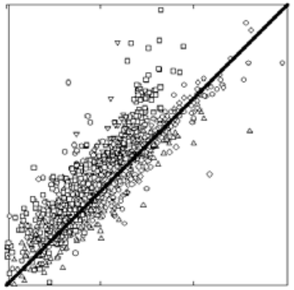
Train a neural network to mimic the effect of weather on soil water fluxes

Test the neural network on independent data



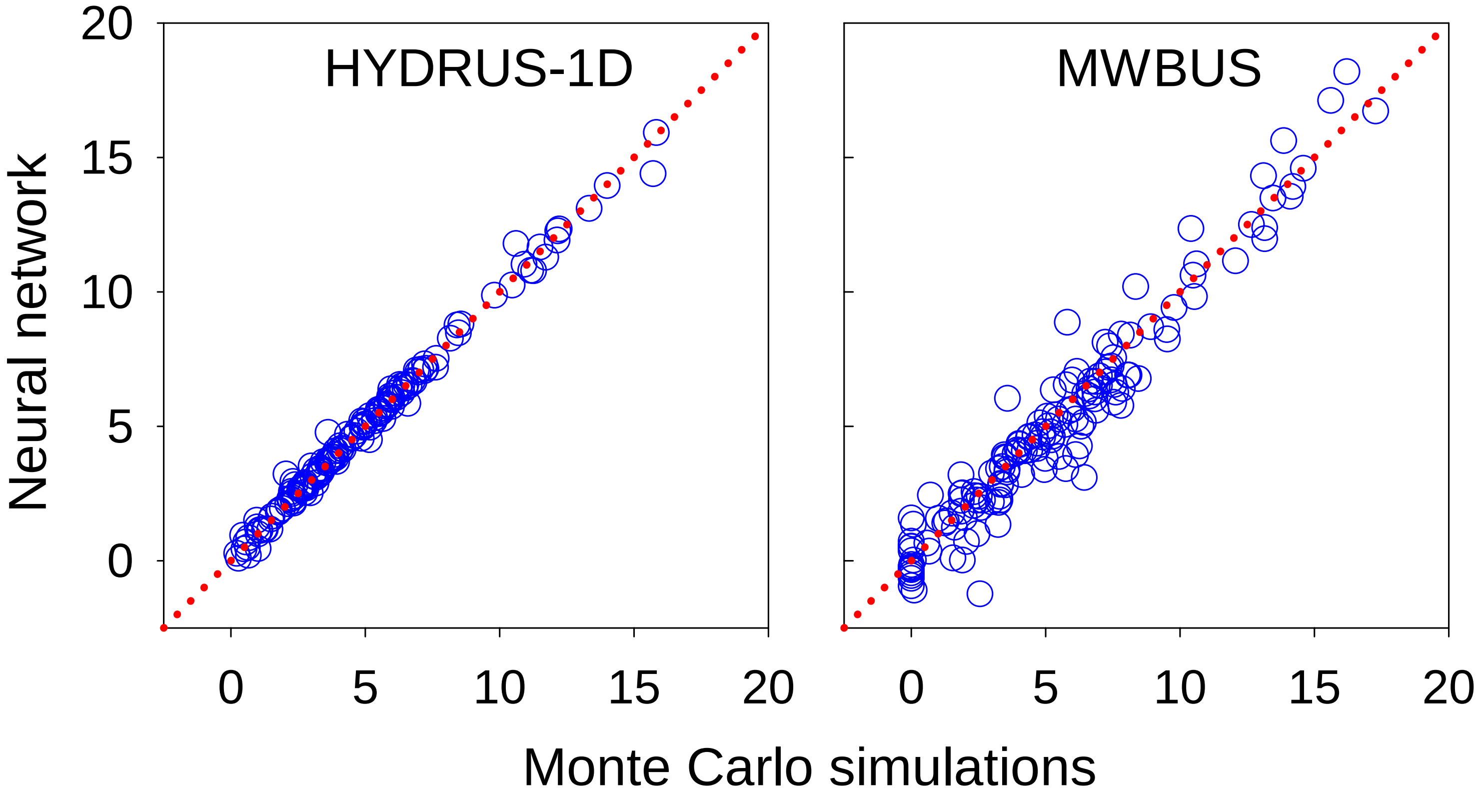






Model form abstraction

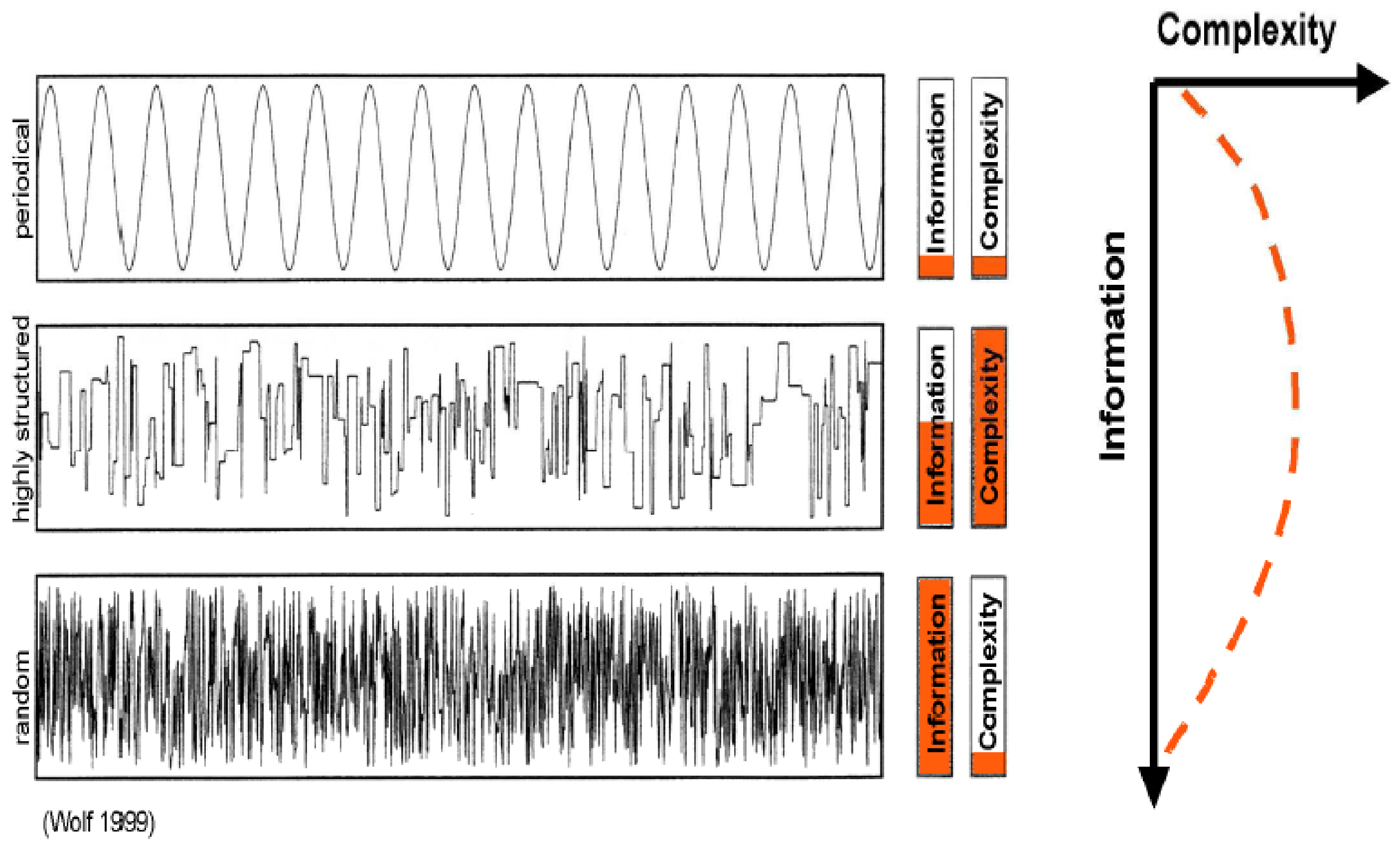
Example of the one–month outflow predicted with daily weather



ANN is 10^5 times faster

Model output complexity

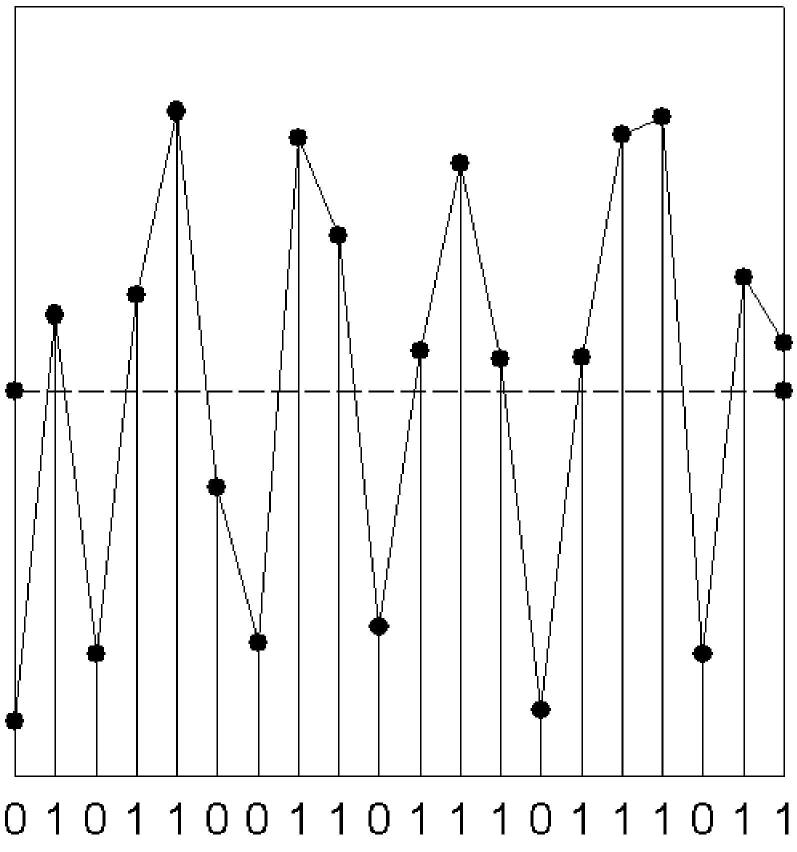
Information theory and complexity measures



intuitive relationship between information and complexity

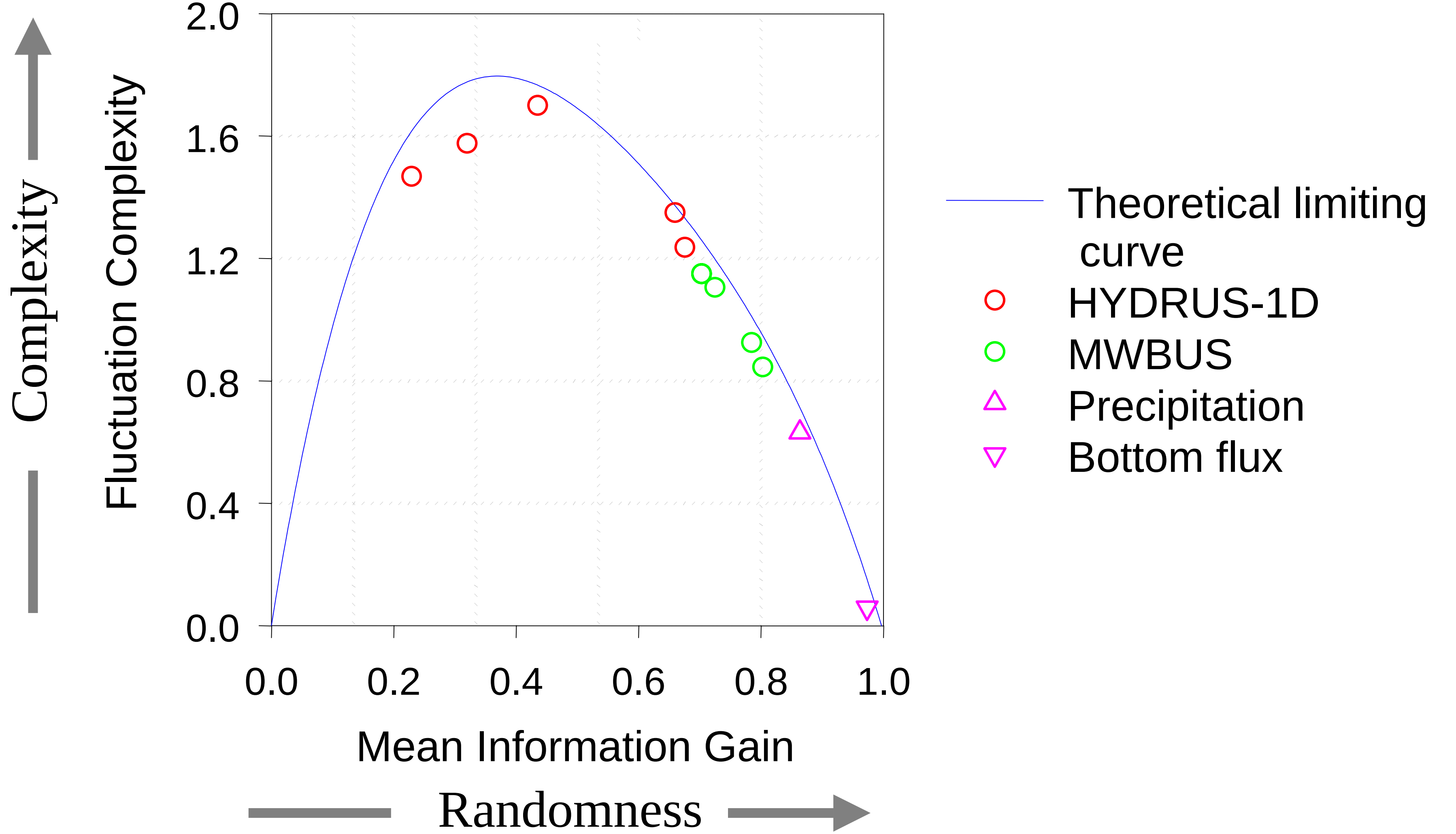
Model output complexity

Binary encoding of the soil water fluxes



Model output complexity – daily fluxes

Complexity and randomness – HYDRUS vs. MWBUS



What did we learn so far?

- Contaminant hydrology has developed a wide variety of efficient model abstraction techniques.
- Model abstraction has to be performed in the uncertainty context.
- Temporal persistence provides an opportunity to decrease uncertainty in upscaled soil water contents.
- At coarse time scales, passive capillary lysimeters measure soil water fluxes that correspond well to water budget computations.
- Results underscore the importance of the “question to be addressed” as the component of the model abstraction process.
- Model abstraction shows pitfalls that may occur when the inverse modeling is used.

What did we learn so far?

- The simple water budget model worked not worse than mechanistic water flow model with respect to water fluxes at coarse time scale.

- Lab measured hydraulic properties did not provide an advantage compared with pedotransfer functions in the uncertainty context.

- A spectrum of pedotransfer functions gave a good representation of uncertainty in hydraulic properties.

- Using neural networks to mimic performance of the complex models is a promising direction of model abstraction.

- Measuring complexity of model output presented a way to rank the potential of flow models to reflect complexity of flow pathways.

- Model abstraction requires specific software utilities to work based on the uncertainty paradigm.

What did we learn so far?

- We now have flux measurement capabilities at coarse temporal scales. But we are lacking flux measurement capabilities at fine temporal scales. Fine-scale water content measurements cannot substitute fine-scale flux measurements.

- Solute concentration measurements have a potential to reveal fine scale fluxes.

- Model abstraction can be an important component in justification of modeling and data collection for performance assessment.

Dependence of the model error on temporal scale