

Modelling Advisory Group (MAG)

Hydrological Uncertainty Workshop

Global Sensitivity Analysis using the SAFE toolbox

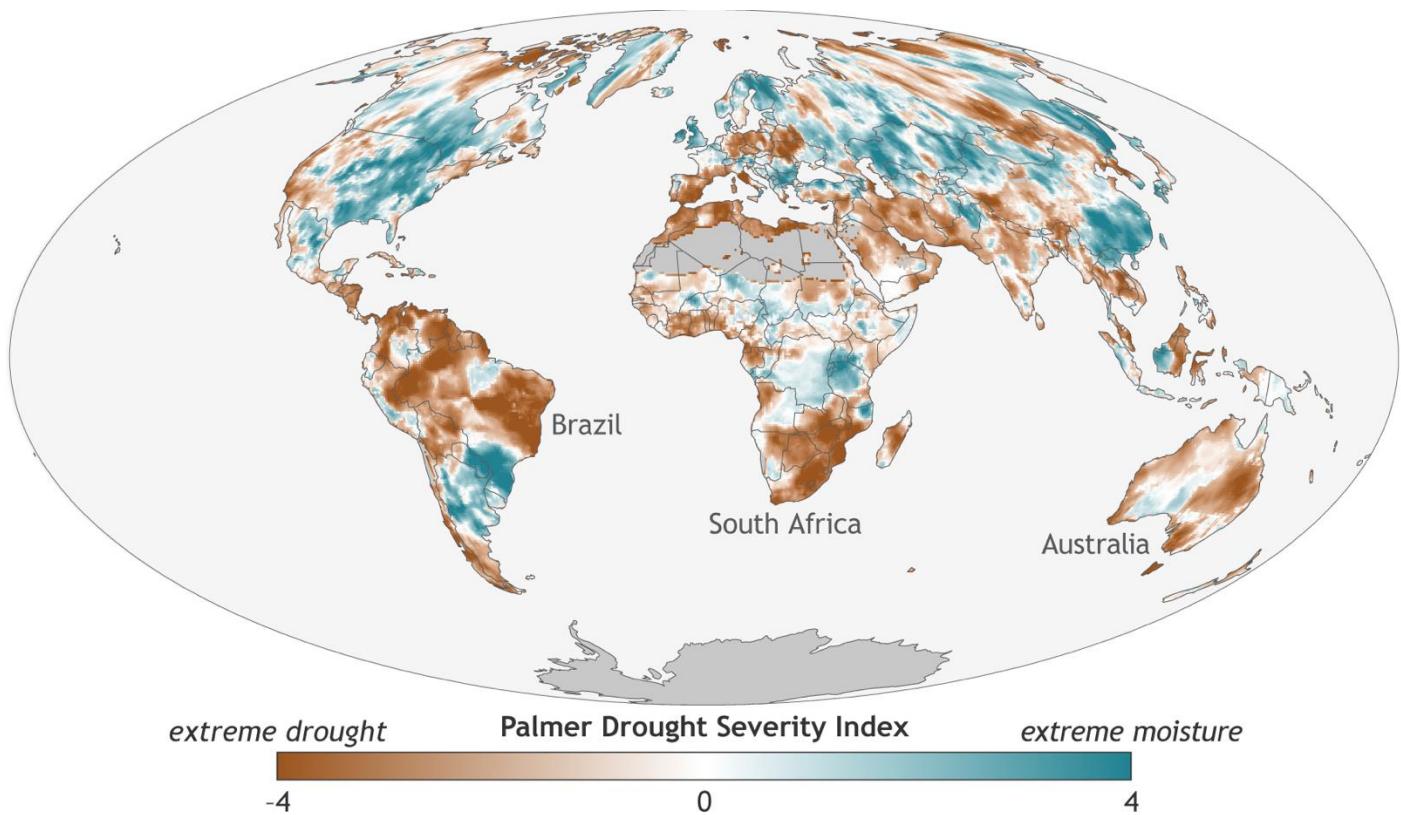
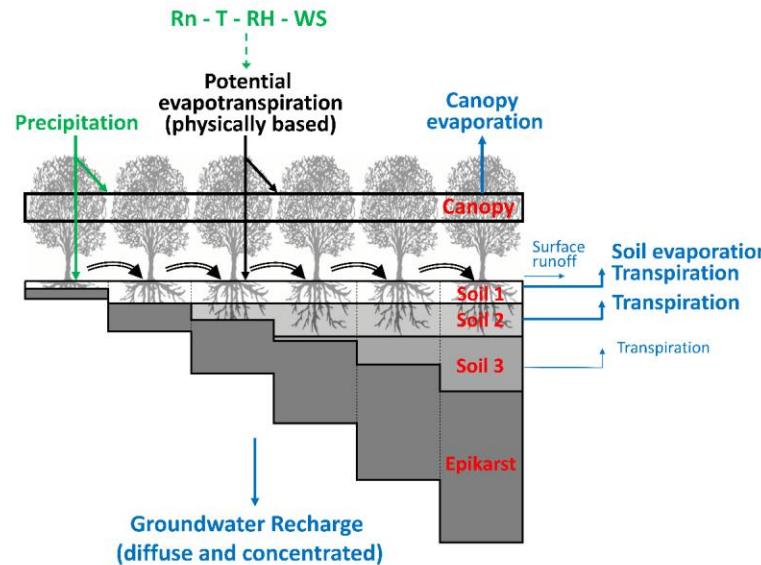
23 September 2025

Francesca Pianosi
University of Bristol



Access to software, papers, example: <https://safetoolbox.github.io/>

Computer models are essential tools in hydrology to advance our science and inform water resources management decisions at different spatial and temporal scales



Despite their differences, all models are built following a similar process

conceptualization

> perceptual model

translation into equations

> mathematical model

implementation into a computer code

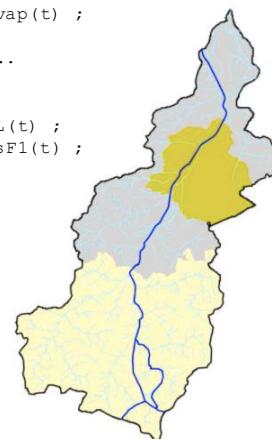
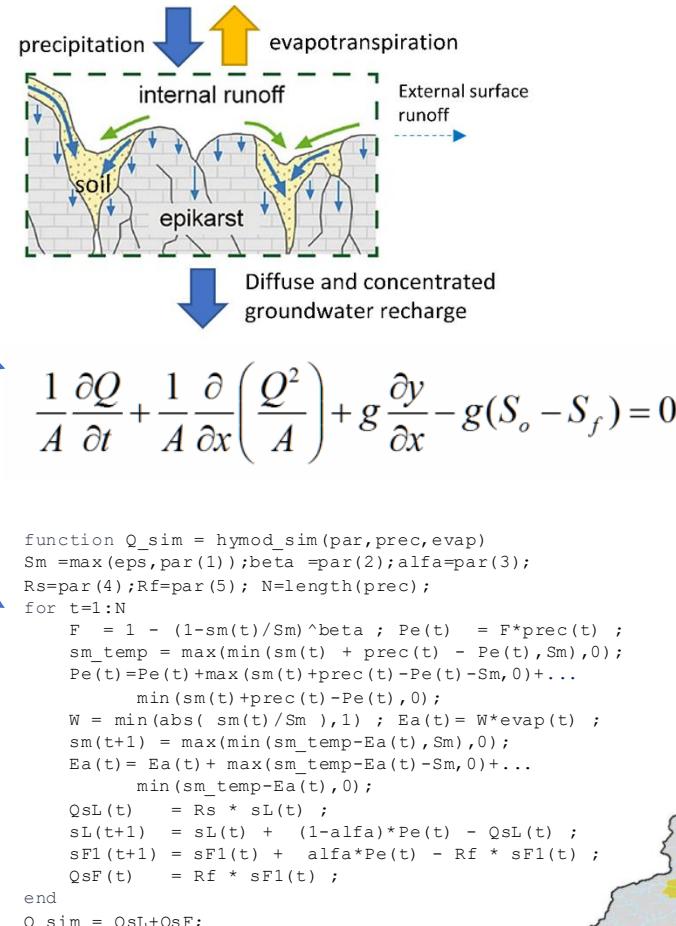
> computer model

calibration

> computer model tailored
to a specific location/system

evaluation

use of the model for prediction
(short-term forecasting, what if analysis, etc.)



Despite their differences, all models are built following a similar process
... which is paved with uncertainties and subjective choices

conceptualization

> perceptual model

translation into equations

> mathematical model

implementation into a computer code

> computer model

calibration

> computer model tailored
to a specific location/system

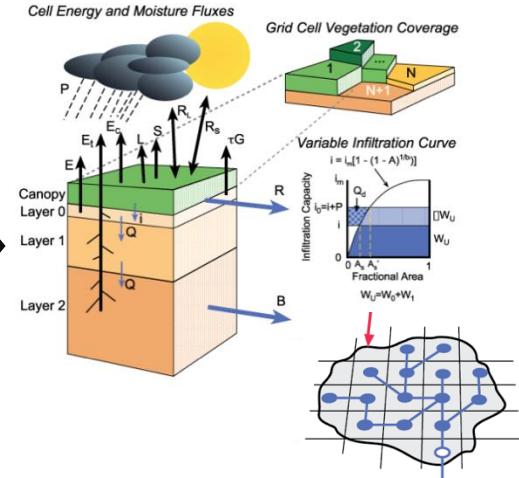
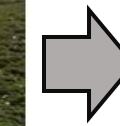
evaluation

use of the model for prediction
(short-term forecasting, what if analysis, etc.)

- simplifying assumptions whose adequacy is uncertain



Water flowing over and through hillslopes



Rainfall-runoff processes
represented in the VIC model

Despite their differences, all models are built following a similar process
... which is paved with uncertainties and subjective choices

conceptualization

> perceptual model

translation into equations

> mathematical model

implementation into a computer code

> computer model

calibration

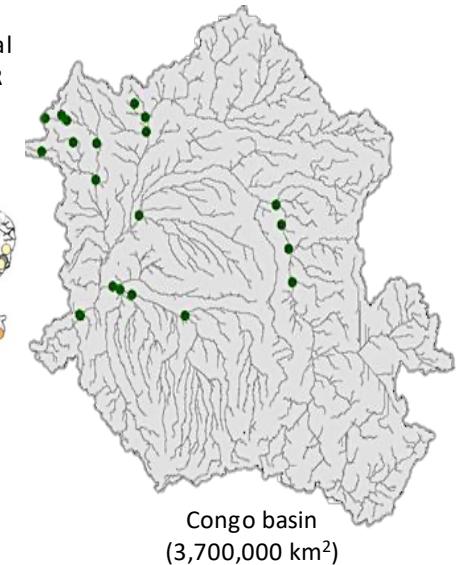
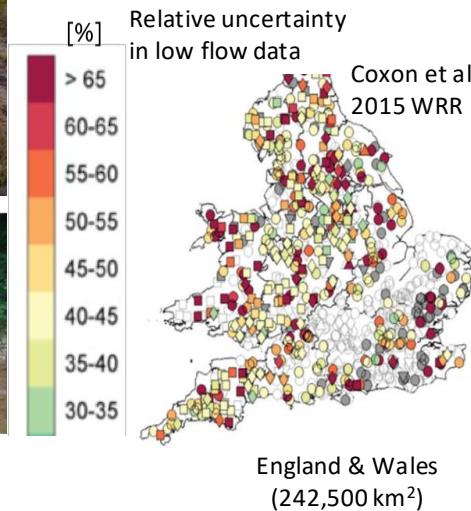
> computer model tailored
to a specific location/system

evaluation

use of the model for prediction
(short-term forecasting, what if analysis, etc.)

- simplifying assumptions whose adequacy is uncertain
- errors and gaps in the data we use to build and test models

Example: river flow data



Despite their differences, all models are built following a similar process
... which is paved with uncertainties and subjective choices

conceptualization

> perceptual model

translation into equations

> mathematical model

implementation into a computer code

> computer model

calibration

> computer model tailored
to a specific location/system

evaluation

use of the model for prediction
(short-term forecasting, what if analysis, etc.)

- simplifying assumptions whose adequacy is uncertain
- errors and gaps in the data we use to build and test models

*These uncertainties can be “**aleatory**” (=due to intrinsic randomness, e.g. from noise in measurement devices) or “**epistemic**” (=due to lack of knowledge, e.g. on the mechanics of a particular process)*

Aleatory and epistemic uncertainties are - to some extent - quantifiable and reducible

Despite their differences, all models are built following a similar process
... which is paved with uncertainties and subjective choices

conceptualization

> perceptual model

translation into equations

> mathematical model

implementation into a computer code

> computer model

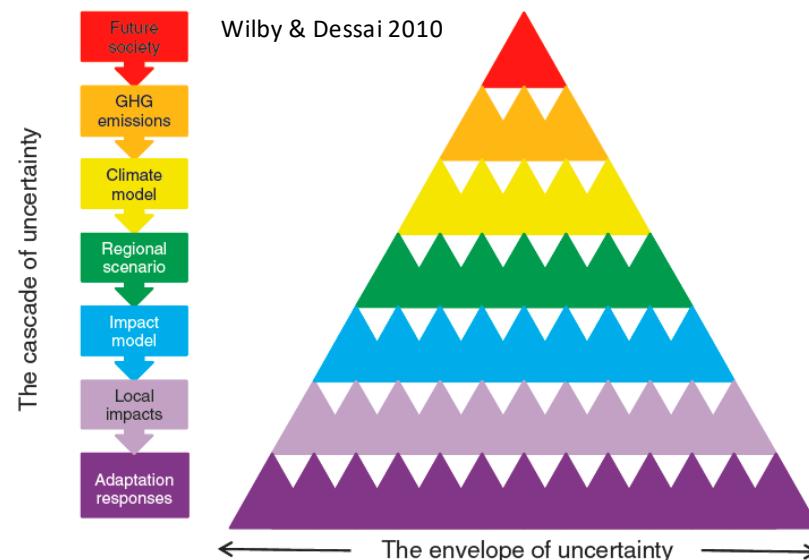
calibration

> computer model tailored
to a specific location/system

evaluation

use of the model for prediction
(short-term forecasting, what if analysis, etc.)

- simplifying assumptions whose adequacy is uncertain
- errors and gaps in the data we use to build and test models
- If we use models to simulate long-term system behaviour, then we need to build scenarios for how the system drivers will evolve in the future

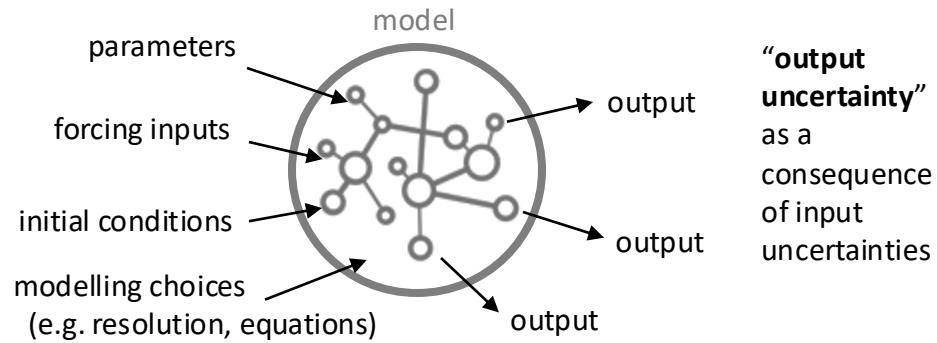


These uncertainties are "deep" (= there is no consensus regarding the appropriate way to represent them)

Deep uncertainties are unquantifiable and irreducible

Uncertainty in >>> Uncertainty out

"input uncertainties"
when the 'right' value or choice for some of the model input factors is uncertain



"output uncertainty"
as a consequence of input uncertainties

Uncertainty in >>> Uncertainty out
So, what do we do with it?

- Ignore it



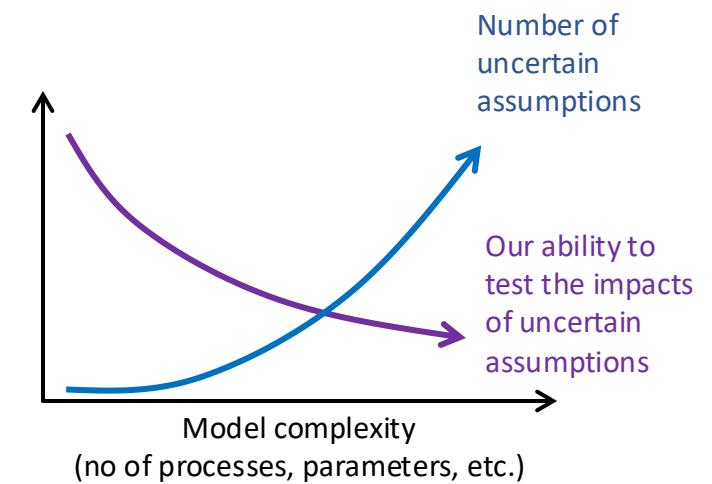
Uncertainty in >>> Uncertainty out

So, what do we do with it?

- Ignore it
- Fight it
 - more monitoring, fieldwork, experiments, thinking, etc.: yes 😊
 - more detailed and “realistic” models: meh 😐

The complexity paradox: “A complex model may be more realistic, yet it is ironic that as we add more factors to a model, the certainty of its predictions may decrease even as our intuitive faith in the model increases (...) The more we strive for realism by incorporating as many as possible of the different processes and parameters that we believe to be operating in the system, the more difficult it is for us to know if our tests of the model are meaningful”

Oreskes 2003 *The role of quantitative models in science*

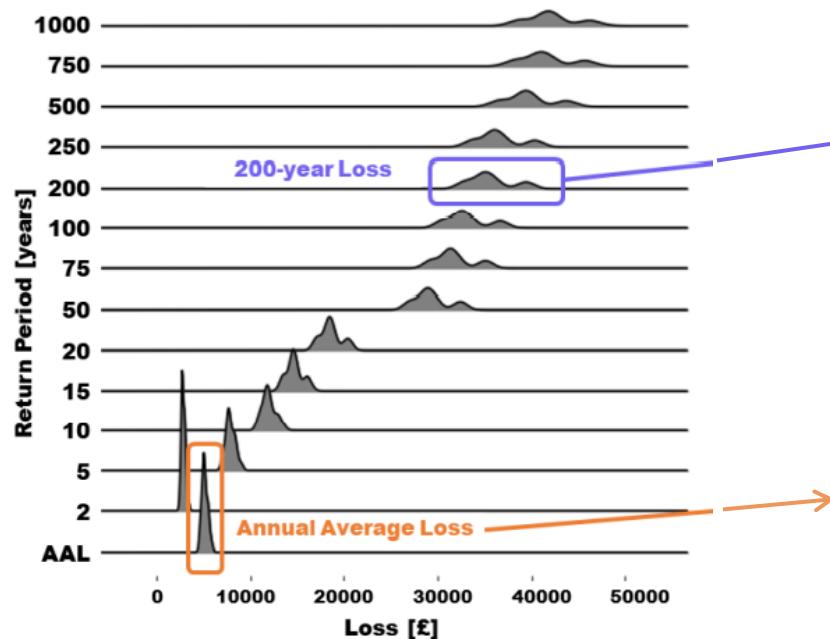


Uncertainty in >>> Uncertainty out

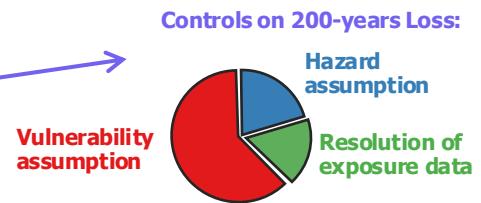
So, what do we do with it?

- Ignore it
- Fight it
- Acknowledge it

Uncertainty analysis:
quantify uncertainty in the model output(s) as a consequence of known (assumed) input uncertainties



Sensitivity analysis:
assess the relative contribution of different input uncertainties to the output(s) uncertainty



What kind of questions can we answer through uncertainty and sensitivity analysis and who are these interesting for?

- Questions to prioritise efforts for model improvement

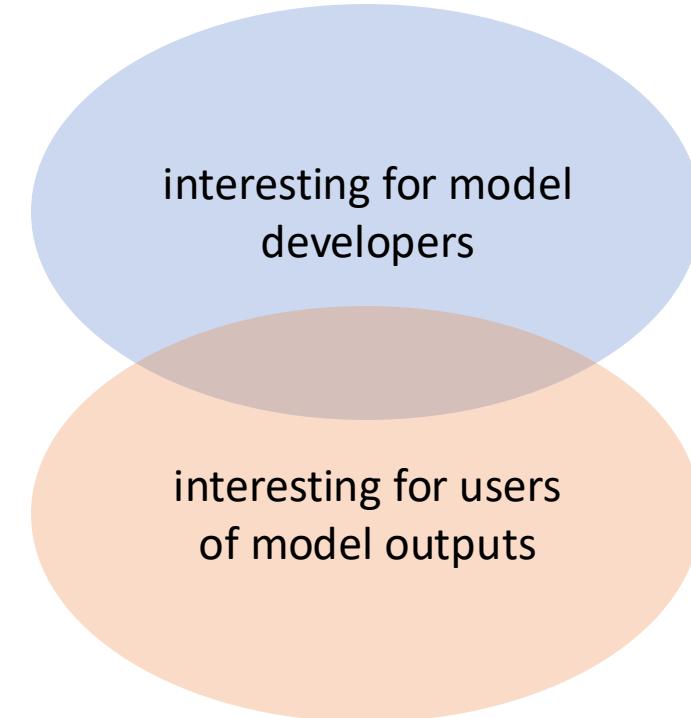
What are the input data, parameters or model components that control the model outputs uncertainty most and where reduction of uncertainty would be most beneficial?

- Questions to evaluate models and establish they are fit for purpose

Does the “right” inputs/components control the “right” outputs? Are model outputs sufficiently controlled by the “decision levers” relative to other input uncertainties?

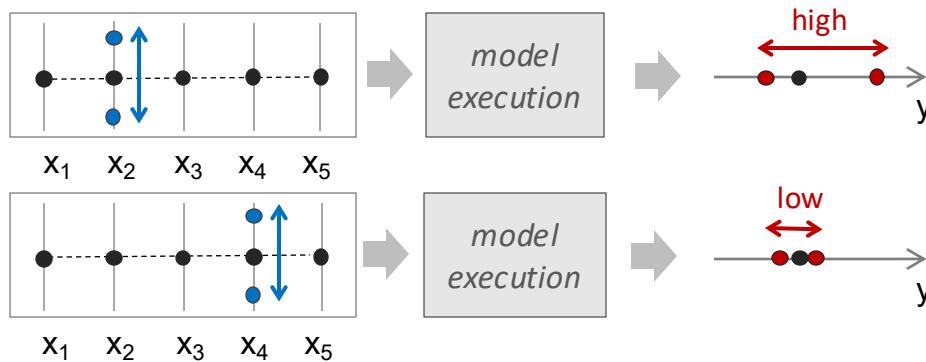
- Questions about the systems behaviour (under the assumption:
system=model)

What are the key drivers of the system in the face of deep uncertainty? Are there “robust” decisions that work sufficiently well across a range of uncertainties?



Two ways of approaching Uncertainty and Sensitivity Analysis

Local approaches: investigate the sensitivity of the output (y) to varying the uncertain inputs (x_1, x_2, \dots) **one at the time** around a **baseline**



Limitation:

- ignore interactions
- results only valid for the chosen baseline: output uncertainty and sensitivity may be very different for a different baseline

Global approaches: investigate the sensitivity of the output (y) to varying all uncertain inputs **simultaneously** across their entire variability space (no baseline!)

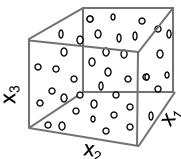
Global approaches are based on repeated executions of the model against different inputs' combinations and a statistical analysis of the resulting input-output dataset

[1] characterize uncertainty in input factors

RUNNING MONTE-CARLO SIMULATIONS

characterize uncertainty in the inputs

(e.g. via a list of possible values, variability range, probability distribution.)



sample N combinations of inputs

(e.g. via random sampling, Latin Hypercube sampling)

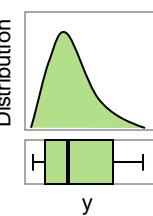
execute the model against each inputs' combination

[2] propagate input uncertainties through the model

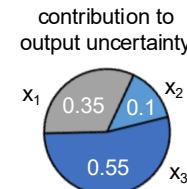
[3] quantify output uncertainty

ANALYSING INPUT-OUTPUT DATASET

derive the output(s)' distribution or CIs



calculate sensitivity indices



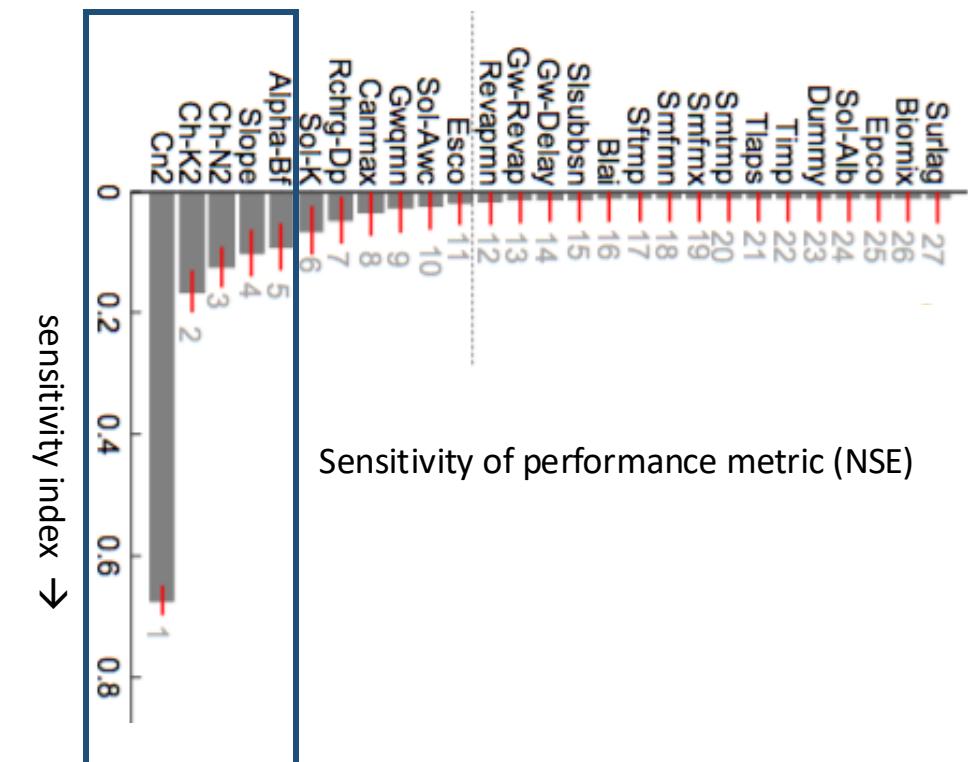
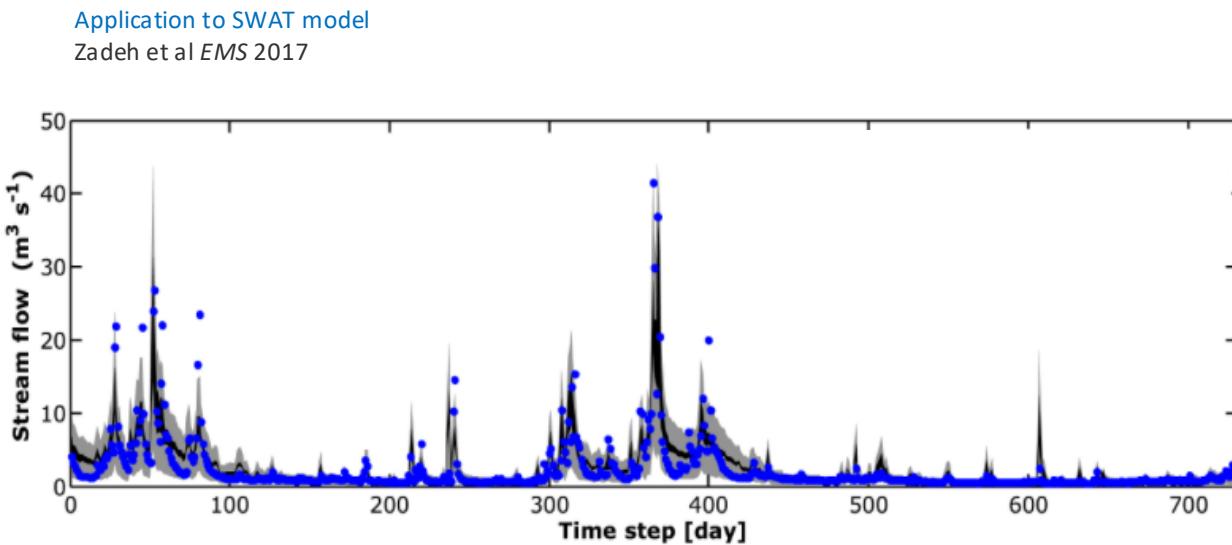
- [4] attribute output uncertainty
- Elementary Effects method
 - Regional Sensitivity Analysis
 - Variance-based (Sobol') method
 - Distribution-based methods (for example, PAWN) etc.

WHY

doing UA/SA?

Guiding model calibration

If we have output observations to compare with, which model parameters control the predictions accuracy, and thus should be the focus of model calibration?

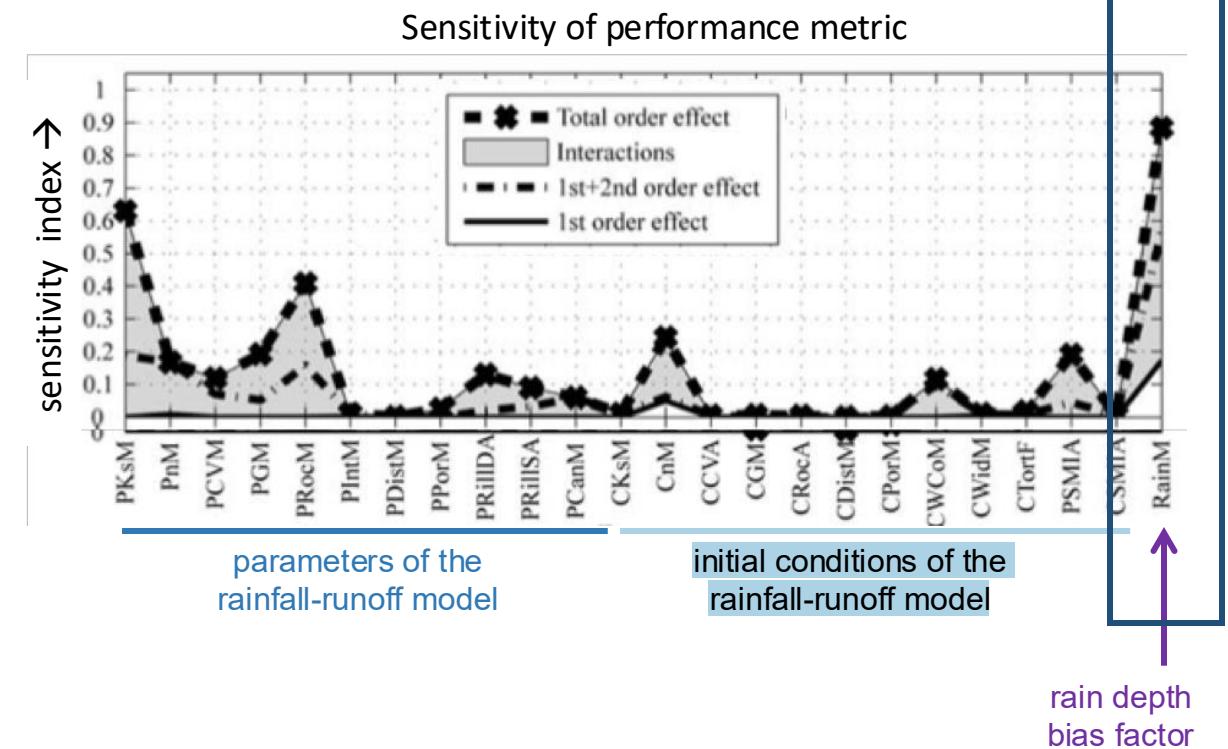
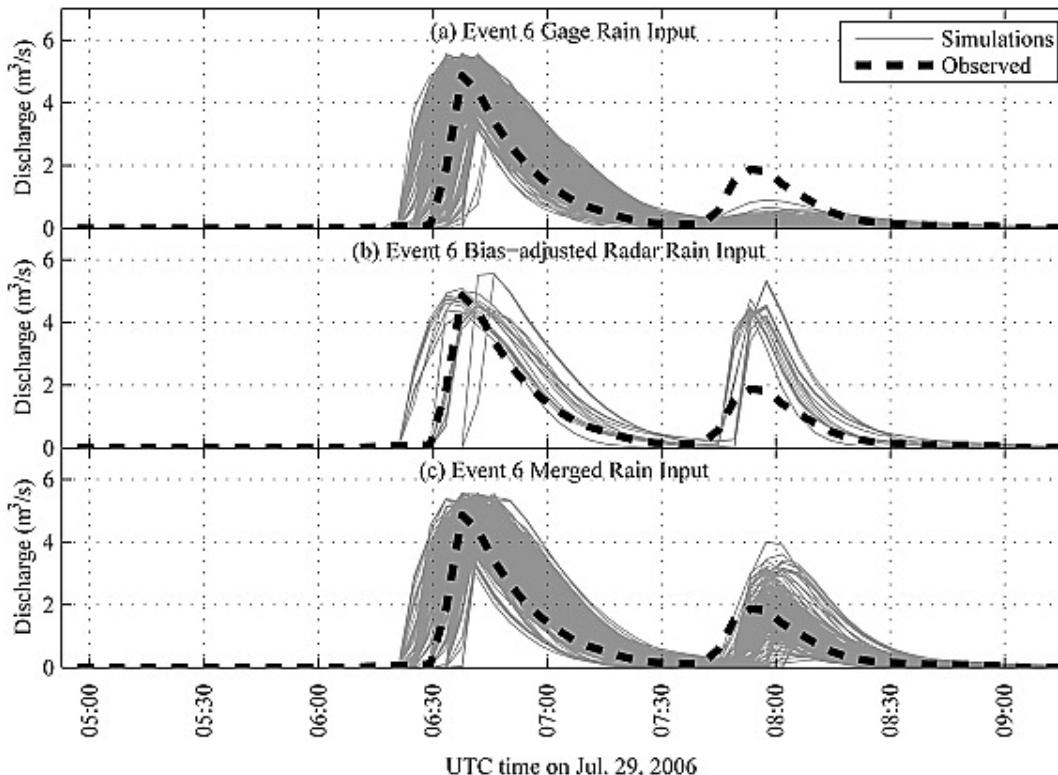


Guiding model calibration

How much of the model performance is controlled by the uncertainty in parameters of the hydrological model vs uncertainty in parameters used for pre-processing forcing inputs (for example, precipitation)?

Application to spatially-distributed rainfall runoff model

Yatheendradas et al *WRR* 2008



Supporting model evaluation

If we do not have output observations, can we at least ensure that the 'right' parameters control the model response at the right place?

"Data-based" evaluation:



Fit-to-data: are model outputs consistent with observations?

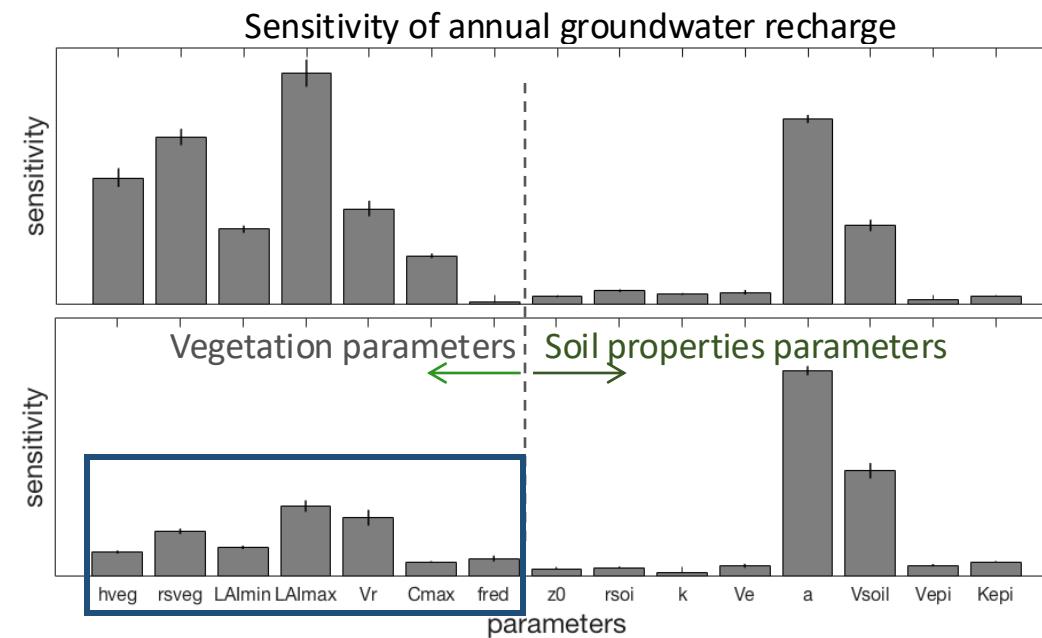
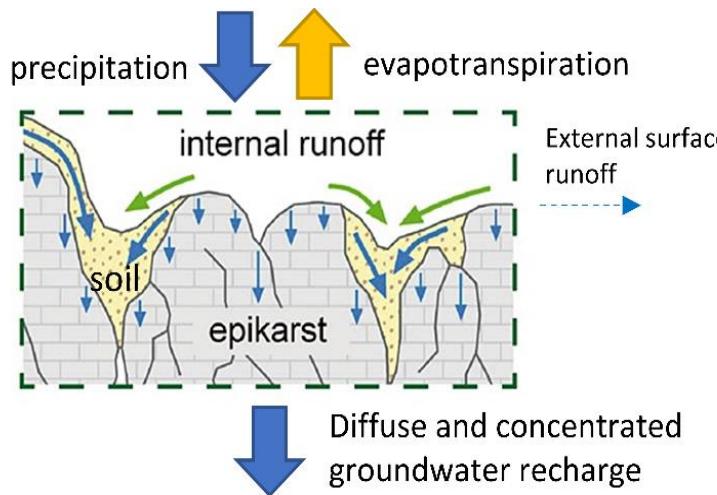
"Response-based" evaluation:



Is the model input-output response consistent with our understanding of the systems functioning?

Wagener et al WIREs-CC 2022
doi:10.1002/wcc.772

Application to karst groundwater recharge model
Sarrazin et al GMD 2018



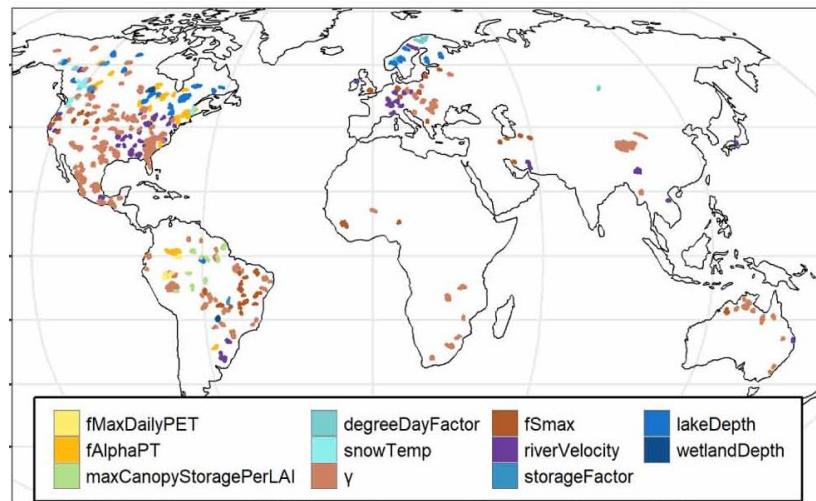
Guiding model calibration and evaluation

Application to global hydrological model WaterGAP3

Kupzig et al ERL 2023

doi:10.1088/1748-9326/acdae8

Identifying most influential parameter on NSE of daily streamflow across 347 gauged catchments



Identifying most influential parameters on different output metrics

	observations-free										fit-to-observations			
	Q10	maxTiming	Q50	Q90	minTiming	FDC.slope	sd	mean	KGE	logNSE	NSE	r		
wetlandDepth	13	13	13	12	13	13	13	14	13	14	12	13		
lakeDepth	10	9	10	8	8	4	9	10	9	10	9	9		
storageFactor	12	11	11	10	10	9	11	12	11	11	11	10		
evapoReductionExp	13	15.5	13	12	13	12	14	11	13	13	13	13		
wetOutflowExp	17	17	17	17	17	17	17	17	17	17	17	17		
lakeOutflowExp	14	15	14	14	14	14	14	13	15	14	14	15		
riverVelocity	7	4	5	8	9	8	3	12	5	6	3	1		
k_g	7	8	6	2	3	2	7	9	8	8	8	7		
runoffFracBuiltUp	14	16	14	14	14	14	14	13	14	14	14	14		
fSmax	2	3	6	6	4	5	2	3	2	3	3	3		
Y	1	2	1	1	1	2	2	1	1	1	2	3		
snowTemp	10	7	10	10.5	9	11	9	9	10	10	10	9		
degreeDayFactor	11	10	12	13	13	13	11	11	12	12	11	11		
canopyEvapoExp	13	16	13	14	15	14	13	13	14	14	13	14		
axCanopyStoragePerLAI	5	7	3	4	6	8	8	4	6	4	5	8		
fAlphaPT	3	6	4	6	6	6	4	3	4	4	6	7		
fMaxDailyPET	5	7	6	7	7	8	7	5	6	5	7	7		

Ranks
 ≥7 [less important]
 4 - 6
 ≤ 3 [more important]

Some parameters are consistently less important across metrics and basins → exclude from calibration

Some parameters are consistently very important across metrics and basins → prioritise for calibration

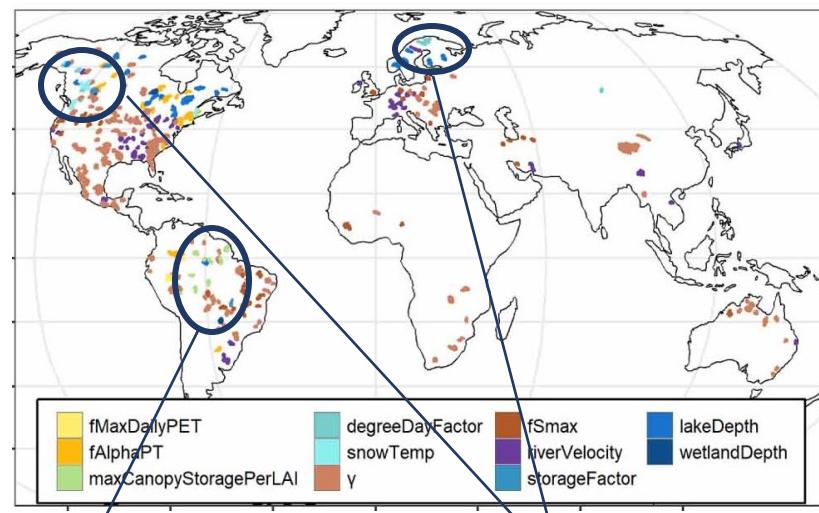
Guiding model calibration and evaluation

Application to global hydrological model WaterGAP3

Kupzig et al ERL 2023

doi:10.1088/1748-9326/acdae8

Is the most influential parameter the “right one” at the “right place”?

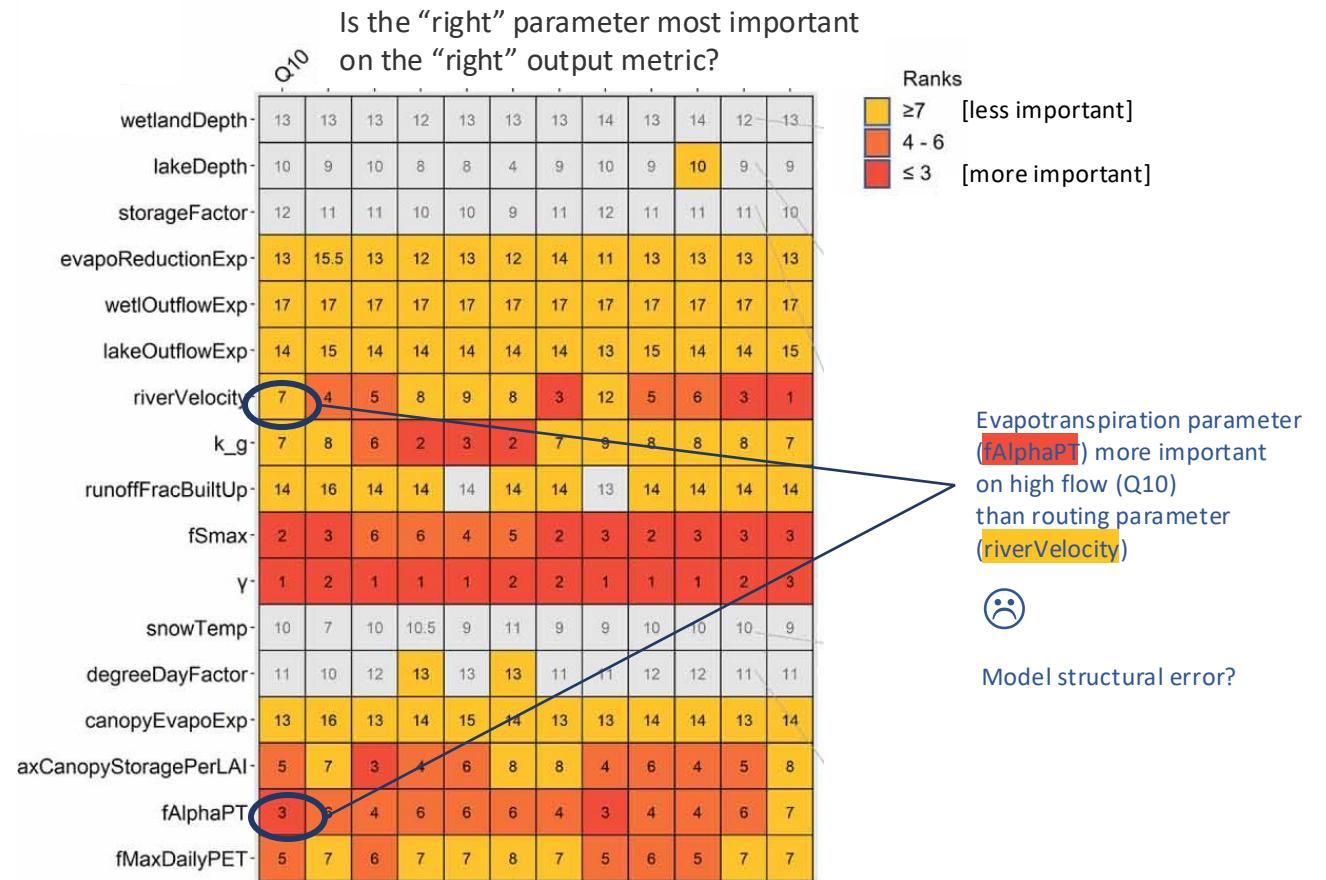


CanopyStorage
key in forested
catchments



SnowTemp
key in snowmelt
dominated catchments

Is parameter importance explained by catchment attributes and/or climate zones?



Finding controls of system behaviour

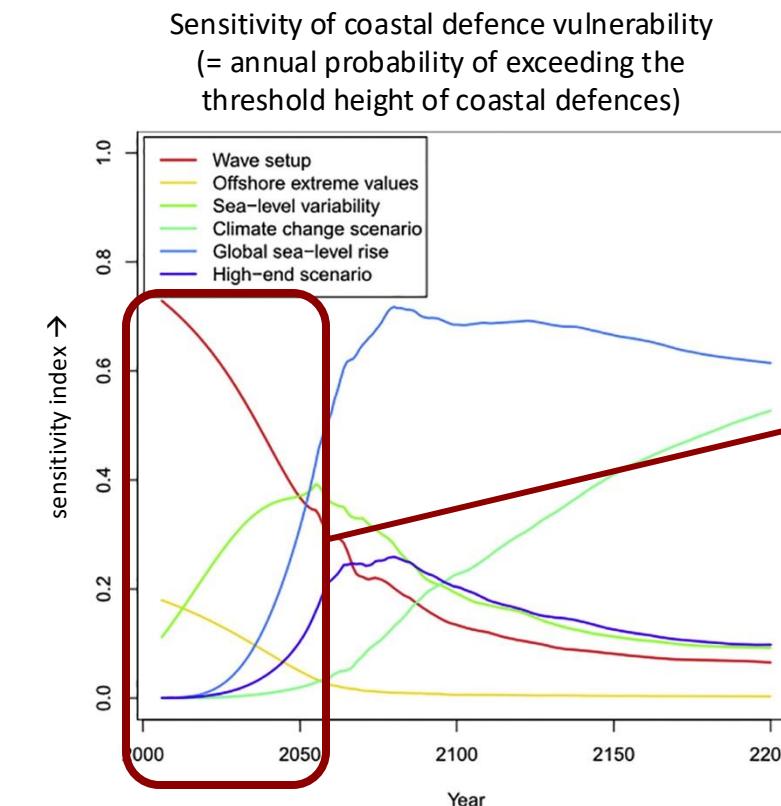
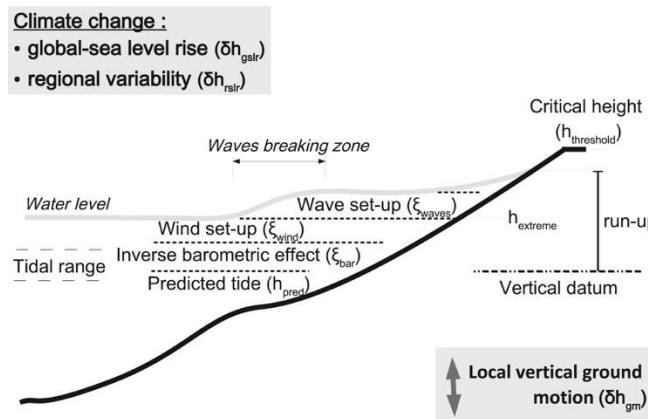
What are the dominant control of the model output?
Are model outputs sensitive to decision-relevant inputs?

If outputs are more strongly controlled by uncertain assumptions/parameters than by the policy/scenario inputs, then the model will tell us more about the consequences of the assumptions embedded in it than it will tell us about the different policy options/scenarios

Application to flood defence assessment model

Le Cozannet et al EMS 2015

doi: 10.1016/j.envsoft.2015.07.021



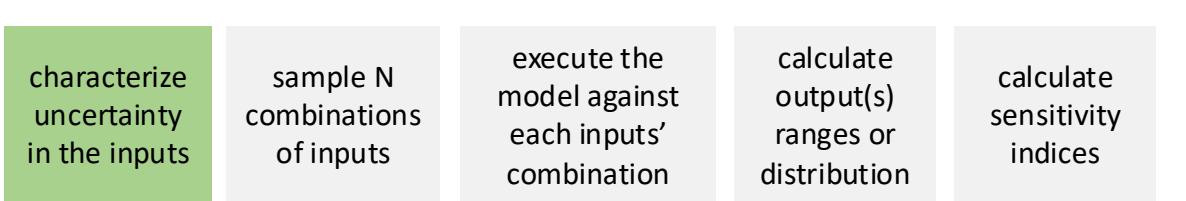
In the 'mid-term', the model predictions are controlled by the modeller's choice of the (very uncertain) 'wave setup' parameter way more than they are by climate change and sea level rise scenarios

→ the model should not be used for impact assessment on such temporal scales

How to do UA/SA?

Characterizing uncertainty in input factors

What is the appropriate distribution / range for the uncertain inputs?



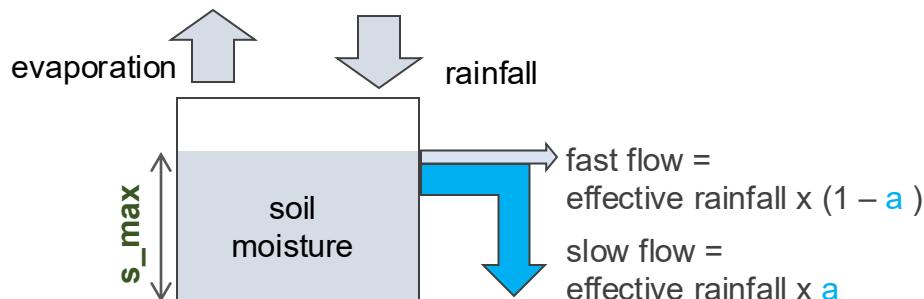
Characterizing uncertainty in input factors

Depending on the type of input (scalar parameter, time series, discrete modelling choice, etc.) and on our level of uncertainty about it, we can use:

- a list of possible values
 - a uniform distribution within an uncertainty range
 - a probability distribution

... and define them using literature sources, historical observations, experts' judgment, etc.

Sometimes the range (distribution) is univocally defined by the physical meaning of the input, but most often different definitions are possible

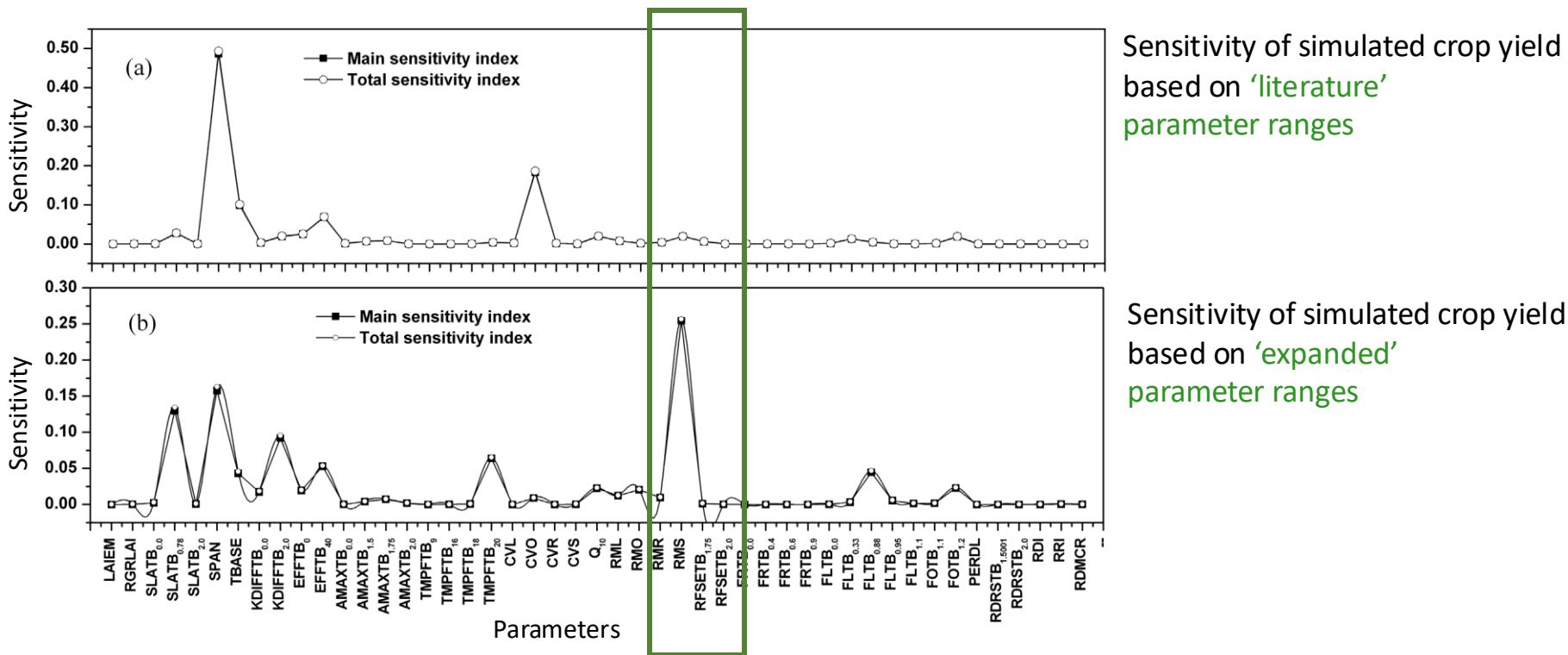


The maximum soil capacity (s_{max}) varies between 0 and an upper bound that may be difficult to define

The repartition coefficient (a) varies between 0 and 1 by definition

When different definitions of the inputs ranges (distributions) are possible, the choice can significantly condition UA/SA results

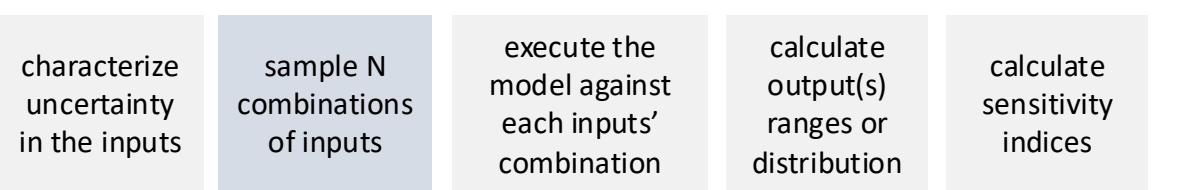
Example from SA of a crop growth model (Wang et al EMS 2013)



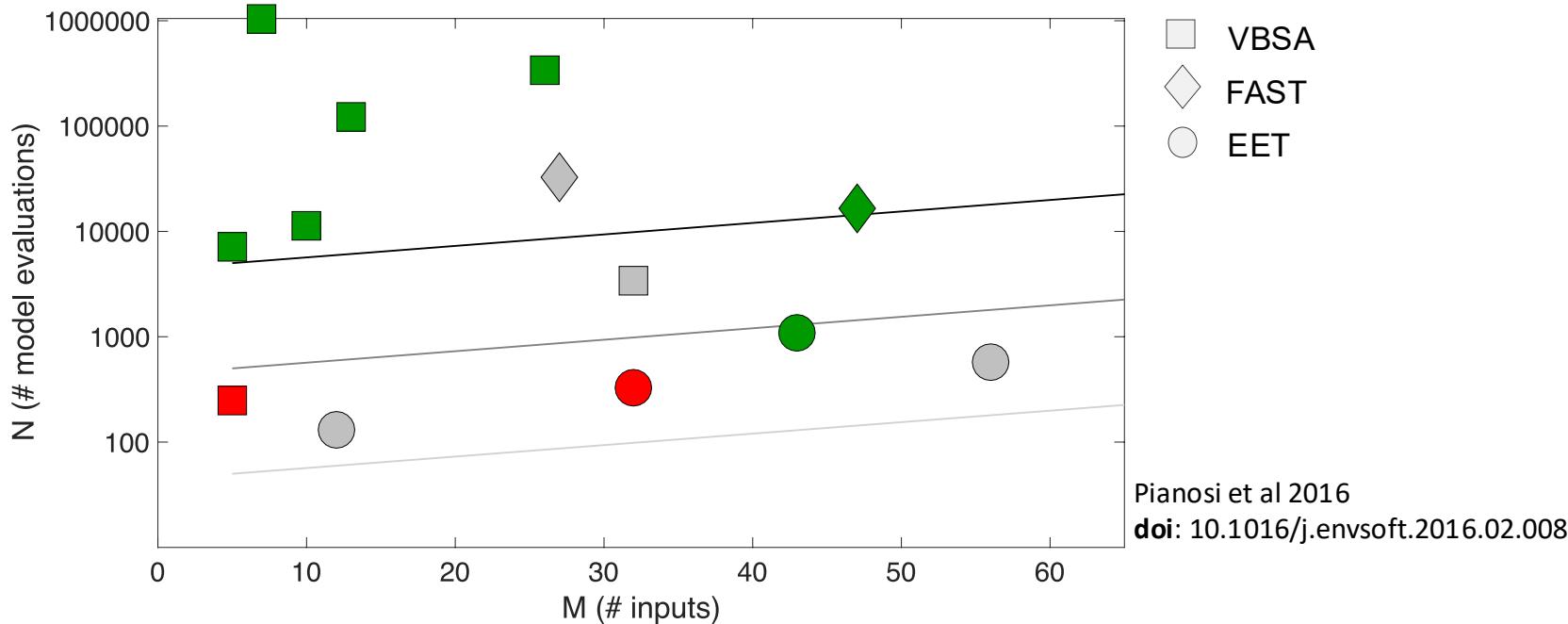
Choosing the sampling strategy and size

Which sampling technique to use (e.g. random sampling, Latin Hypercube, quasi-random sequences)?

How many samples are needed?

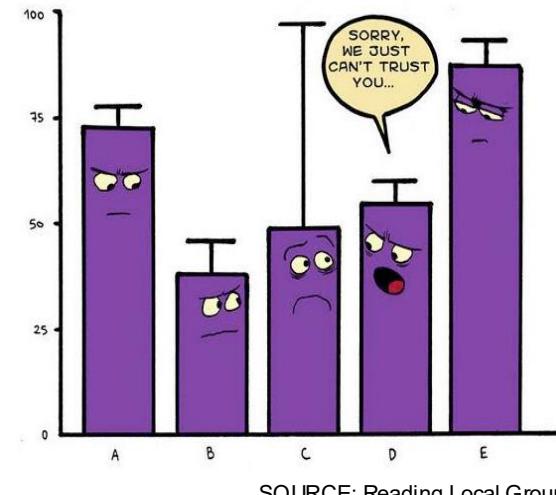
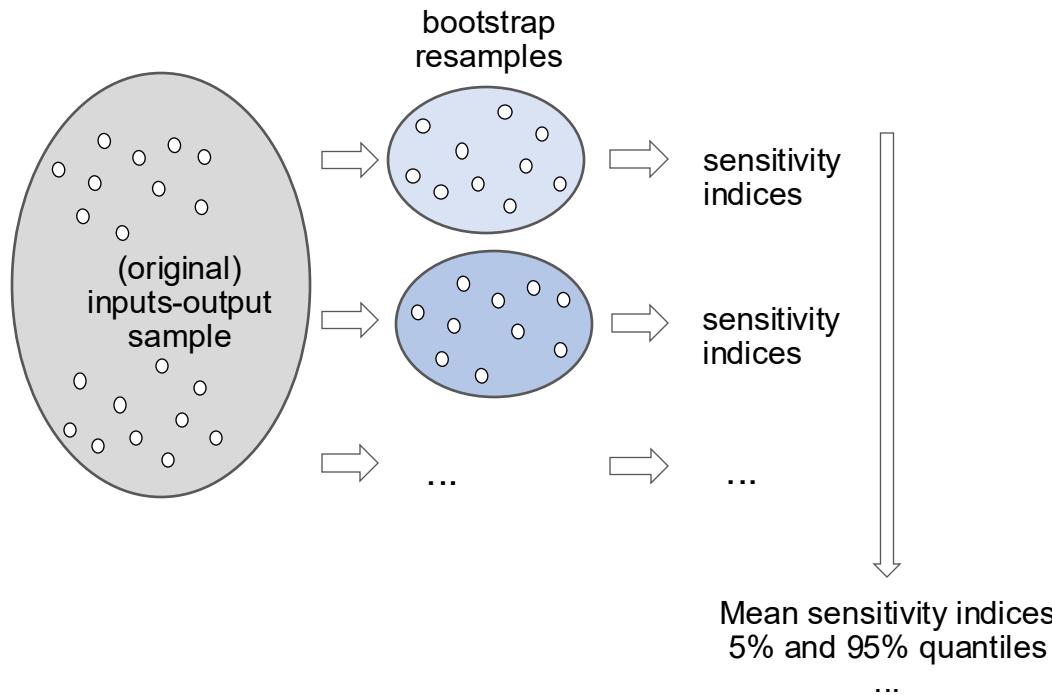


In general, the required sample size (N) increases with the number of uncertain inputs (M). However, the proportionality rate varies significantly from one method to another, and from one application of the same method to another

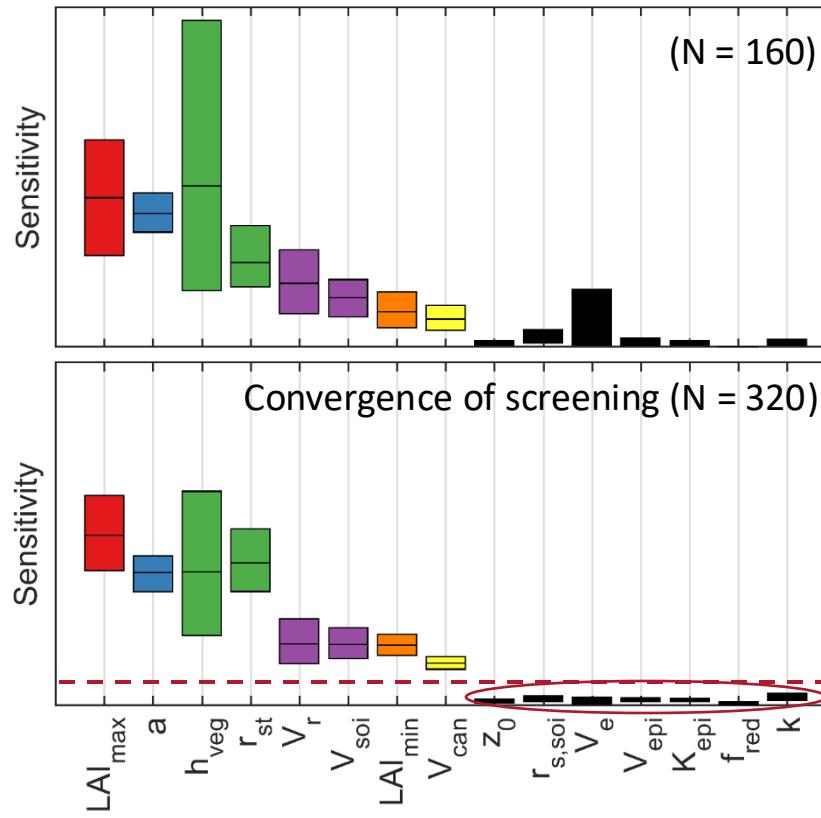


Sensitivity indices are calculated from a sample, so if the sample size is small, their values may be poorly approximated

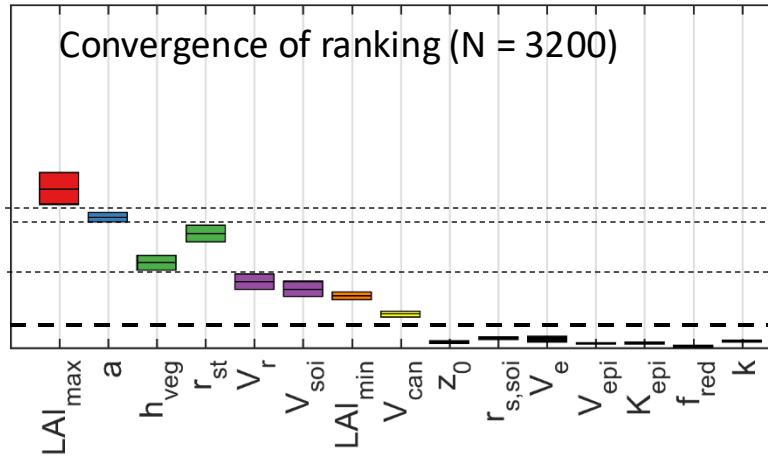
In order to assess the robustness of our sensitivity estimates to the chosen sample, *without re-running the model*, we can use bootstrapping



If the confidence intervals of our sensitivity indices are not “small enough” we must increase the sample size (what is “small enough” depends on the goal of our GSA)



black line: mean sensitivity index
bar: 90% confidence interval



Defining scalar output metric(s)

Which output metrics
should we look at?

characterize
uncertainty
in the inputs

sample N
combinations
of inputs

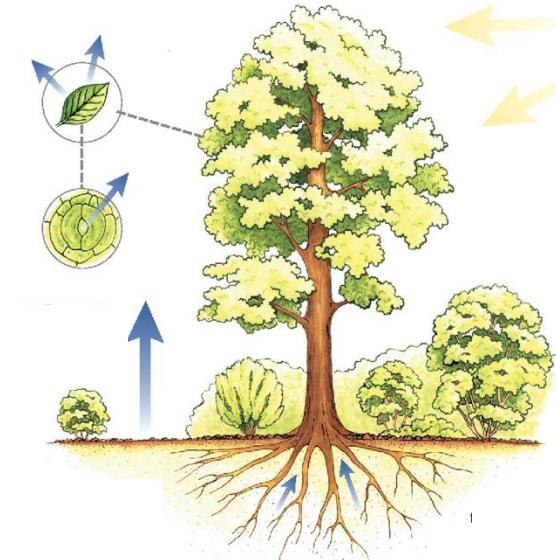
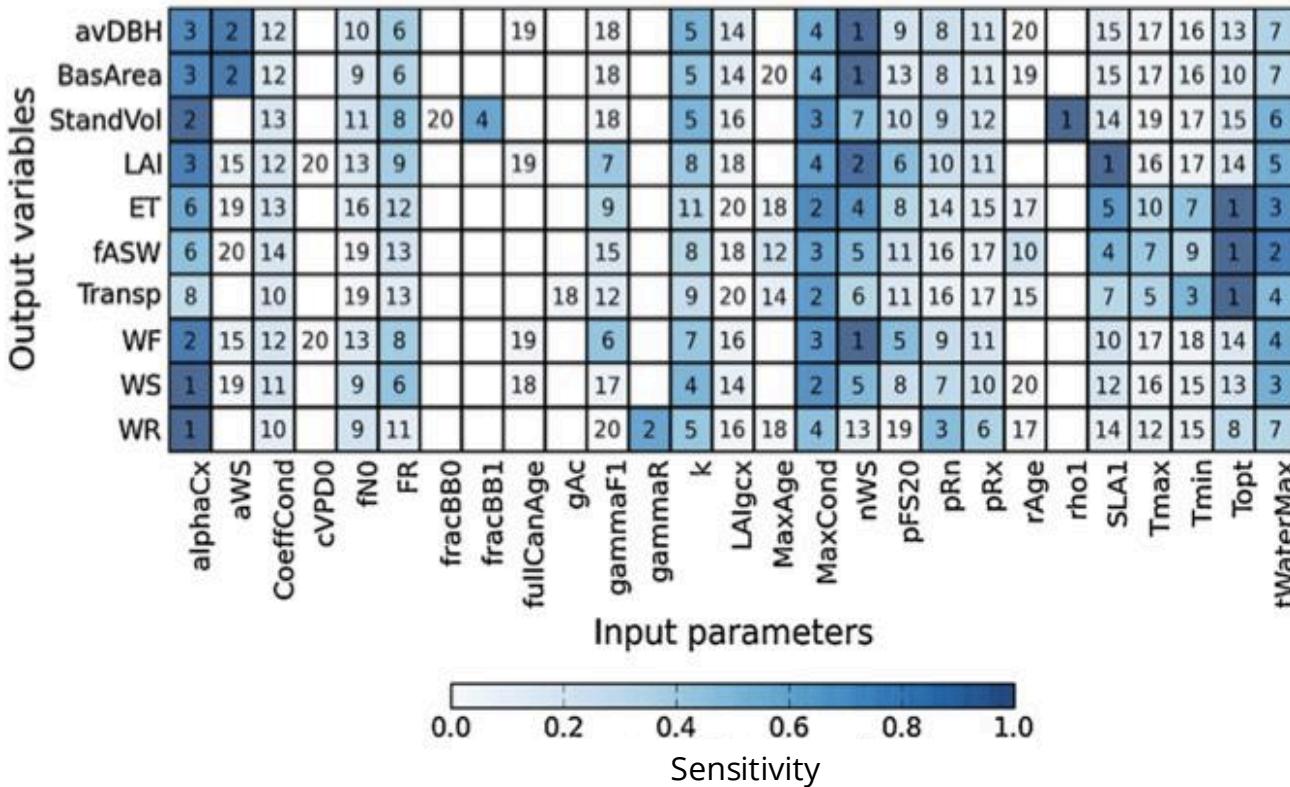
execute the
model against
each inputs'
combination

calculate
output(s)
ranges or
distribution

calculate
sensitivity
indices

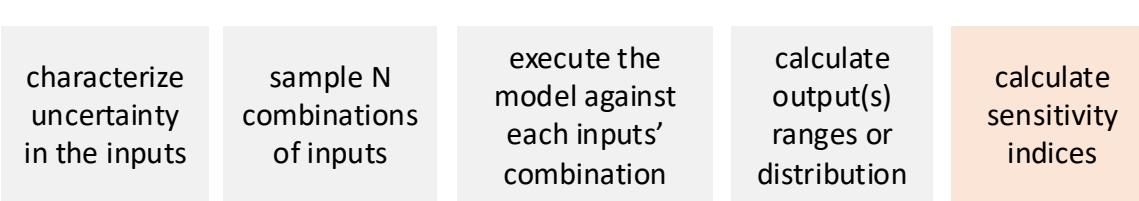
Each output metric is typically sensitive to a small subset of inputs, but which are those inputs will differ from one metric to another

Application to a forest growth model (Song et al EcM 2012)



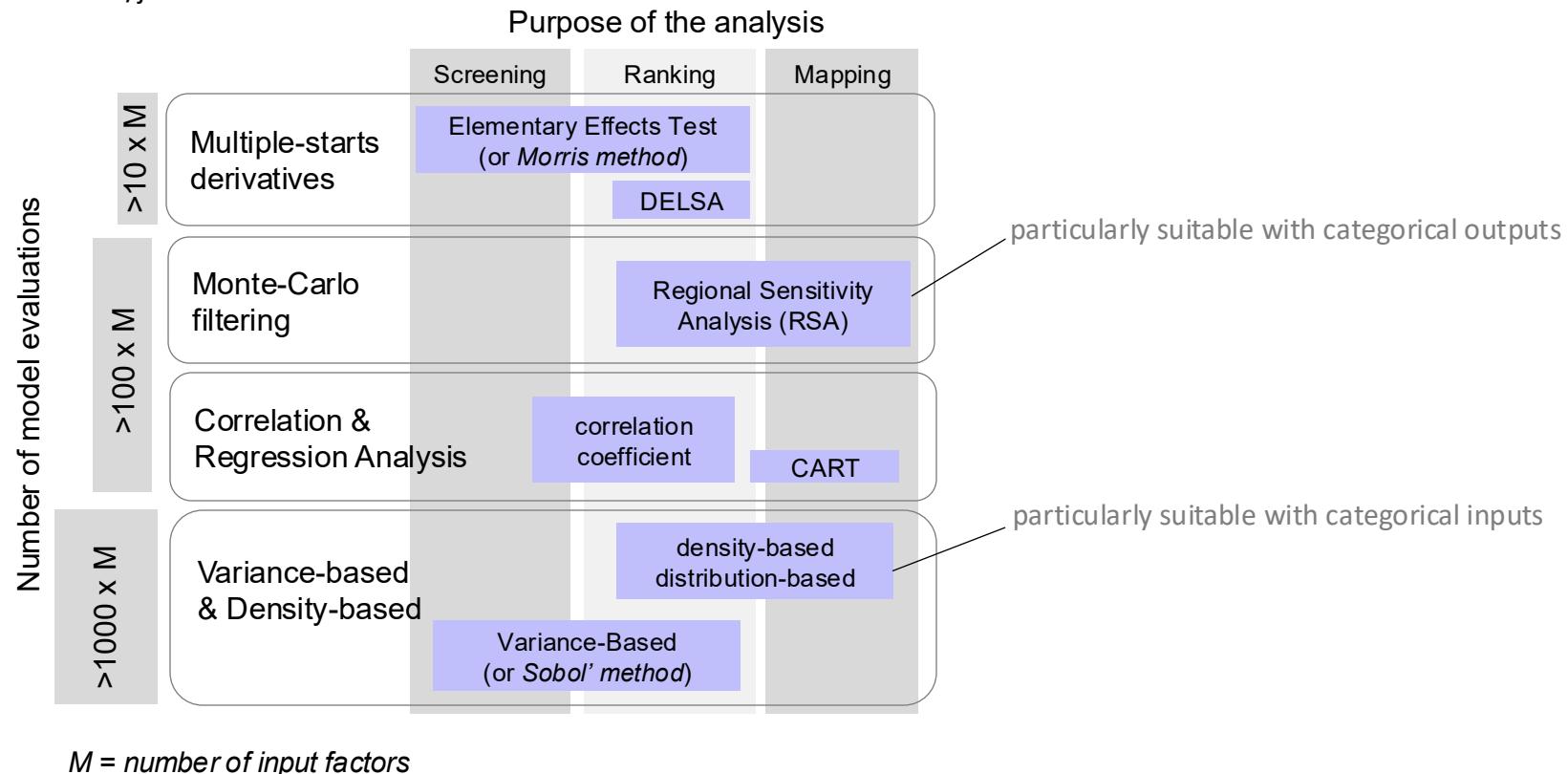
Choosing a method for calculating sensitivity indices

Which global sensitivity analysis method to use (e.g. variance-based, elementary effects test, regional sensitivity analysis, etc.)?



Different methods defines “sensitivity” in different ways and are more or less suitable for specific purposes or problems

Pianosi et al 2016
doi: 10.1016/j.envsoft.2016.02.008



Elementary Effects Test (Morris, 1991)

Number of model evaluations

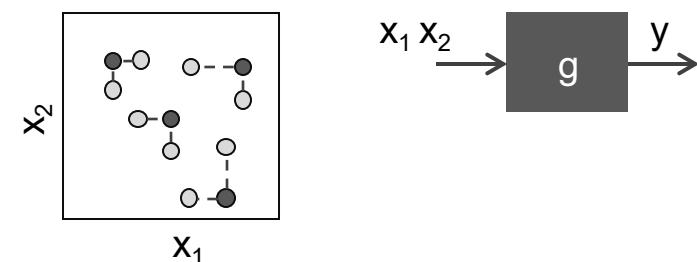
Purpose of the analysis		
	Screening	Ranking
>10 x M	Multiple-starts derivatives	Elementary Effects Test (or Morris method)
>100 x M	Monte-Carlo filtering	Regional Sensitivity Analysis (RSA)
>1000 x M	Correlation & Regression Analysis	correlation coefficient
	Variance-based & Density-based	density-based distribution-based
		Variance-Based (or Sobol' method)

Sensitivity is proportional to...

the mean finite differences of the output across the input space

$$S_i = \frac{1}{r} \sum_{j=1}^r EE^j$$

$$EE^j = \frac{g(\bar{x}_1^j, \dots, \bar{x}_i^j + \Delta_i^j, \dots, \bar{x}_M^j) - g(\bar{x}_1^j, \dots, \bar{x}_i^j, \dots, \bar{x}_M^j)}{\Delta_i^j} c_i$$



Regional Sensitivity Analysis (Hornberger & Spear, 1980)

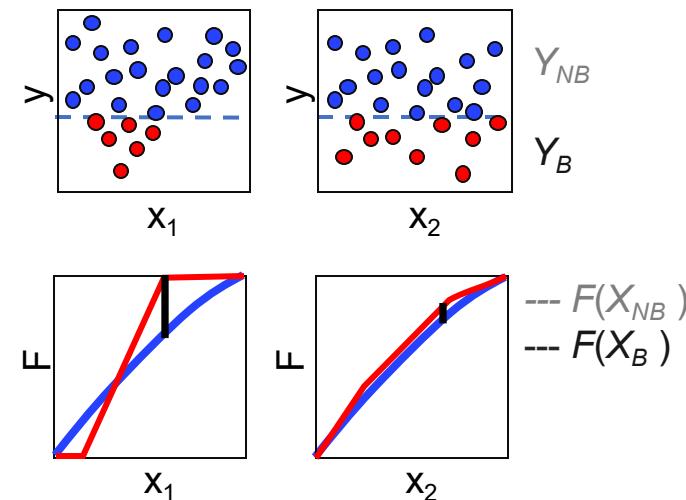
Number of model evaluations

	Purpose of the analysis		
	Screening	Ranking	Mapping
>10 x M	Multiple-starts derivatives	Elementary Effects Test (or <i>Morris method</i>)	DELSA
>100 x M	Monte-Carlo filtering		Regional Sensitivity Analysis (RSA)
>1000 x M	Correlation & Regression Analysis	correlation coefficient	CART
	Variance-based & Density-based	density-based distribution-based	Variance-Based (or <i>Sobol' method</i>)

Sensitivity is proportional to...

the variation induced in the distribution of an input by conditioning the output

$$S_i = \max_q |F_{x_i}(x|x \in X_B) - F_{x_i}(x|x \in X_{NB})|$$



Variance-based Sensitivity Analysis (Homma & Saltelli, 1996)

		Purpose of the analysis		
		Screening	Ranking	Mapping
Number of model evaluations	>10 x M	Multiple-starts derivatives	Elementary Effects Test (or <i>Morris method</i>)	
	>100 x M	Monte-Carlo filtering		Regional Sensitivity Analysis (RSA)
	>1000 x M	Correlation & Regression Analysis	correlation coefficient	CART
	>1000 x M	Variance-based & Density-based	density-based distribution-based	Variance-Based (or <i>Sobol' method</i>)

Sensitivity is proportional to...

variation induced in the variance of the output by conditioning an input

$$S_i = \frac{V_i}{V} = \frac{V_{x_i}[E(y|x_i)]}{V(y)} = \frac{V(y) - E_{x_i}[V(y|x_i)]}{V(y)}$$

$$S_i^T = 1 - \frac{V_{x_{\sim i}}}{V} = 1 - \frac{V_{x_{\sim i}}[E(y|x_{\sim i})]}{V(y)}$$

$$V = \sum_i V_i + \sum_{i < j} V_{ij} + \sum_{i < j < k} V_{ijk} + \dots + V_{12\dots M}$$

Distribution-based Sensitivity Analysis (PAWN)

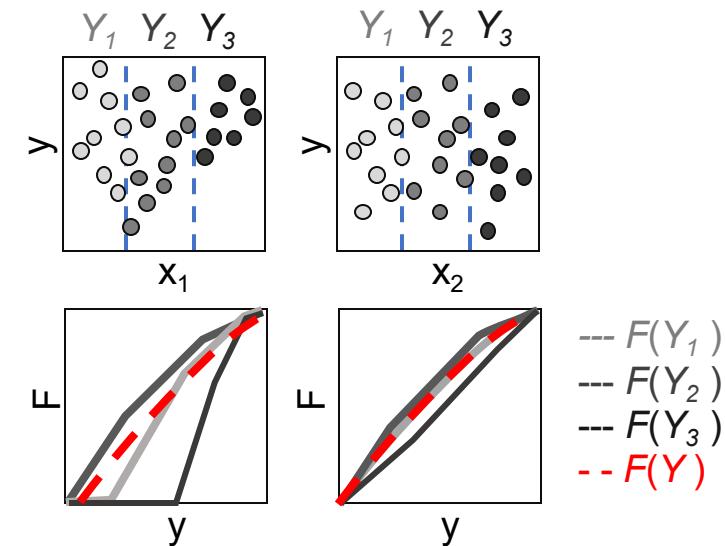
(Pianosi & Wagener 2015)

Purpose of the analysis			
	Screening	Ranking	Mapping
>10 x M	Multiple-starts derivatives	Elementary Effects Test (or <i>Morris method</i>) DELSA	
>100 x M	Monte-Carlo filtering		Regional Sensitivity Analysis (RSA)
>1000 x M	Correlation & Regression Analysis	correlation coefficient	CART
	Variance-based & Density-based	density-based distribution-based	Variance-Based (or <i>Sobol' method</i>)

Sensitivity is proportional to...

variation induced in the distribution of the output by conditioning an input

$$S_i = \text{stat} \max_k \max_q |F_y(q) - F_{y|x_i}(q|x_i \in I_k)|$$



References

- Reinecke, R., Pianosi, F., & Wagener, T. (2024), How to use the impossible map – Considerations for a rigorous exploration of Digital Twins of the Earth, *Socio-Environmental Systems Modelling*, 6, 18786.
- Kupzig, J., Reinecke, R., Pianosi, F., Flörke, M., Wagener, T. (2023), Towards parameter estimation in global hydrological models, *Environmental Research Letters*, 18(7).
- Wagener, T., Reinecke, R. Pianosi, F. (2022), On the evaluation of climate change impact models, *WIRES Climate Change*. 13, 3, e772.
- Wagener, T., Pianosi, F. (2019) What has Global Sensitivity Analysis ever done for us? A systematic review to support scientific advancement and to inform policy-making in earth system modelling. *Earth-science reviews*. 194. 1-18.
- Sarrazin, F., Pianosi, F., Wagener, T. (2016) Global Sensitivity Analysis of environmental models: Convergence and validation . *Environmental Modelling and Software*, 79, pp. 135-152.
- Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T. (2016) Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling and Software*, 79, pp. 214-232.