Università di Enna «Kore»

Centro La.R.A. Laboratori di Ricerca sulle Acque

(Distributed) Data quality and urban flood modelling uncertainty

Gabriele Freni

(not an hydrologist… so be patient)

CINID CONSORZIO INTERUNIVERSITARIO PER L'IDROLOGIA

UNESCO United Nations Educational, Scientific and Cultural Organization · World Water Assessment Programme

WA RRE DOC

Università per Stranieri di Perugia

2019 Edition Doctoral Winter School DATA RICH HYDROLOGY

PERUGIA (Italy), January 28 - February 1, 2019 - Villa Colombella

Regione Umbria

Pluvial, fluvial or coastal flooding in urban areas?

- Fluvial flooding: Levee breach in Denver County (2016)

Pluvial, fluvial or coastal flooding in urban areas?

- Coastal flooding: Hurricane Sandy, USA (2012)

Università di Enna «Kore»
Centro di ricerca La.R.A.

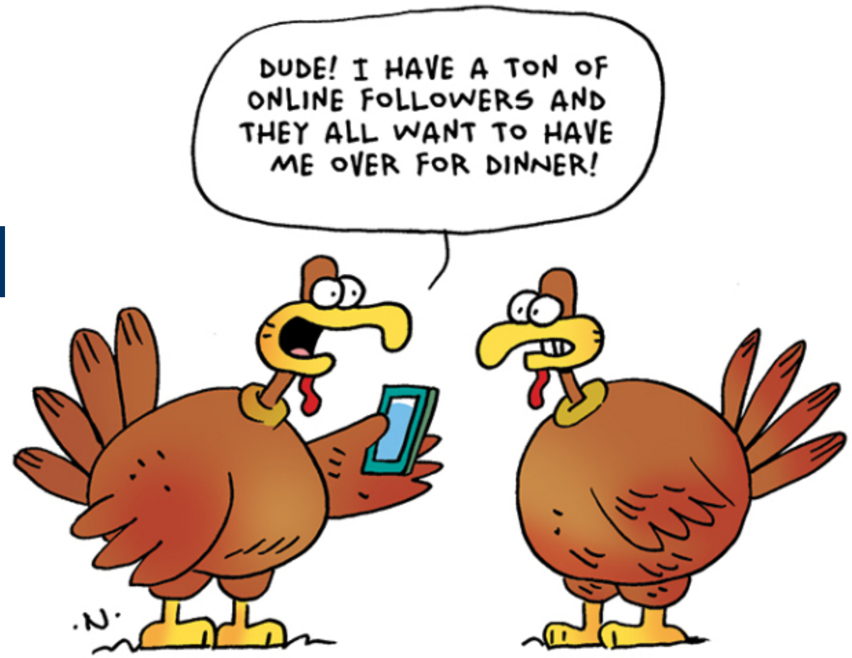
Pluvial, fluvial or coastal flooding in urban areas?

- Pluvial Flooding: Manchester (UK), 2018

Few statements

We use models to transform data to information and information to wisdom (<u>to take decisions</u>)

(and because we **<u>do not believe</u>** to have enough evidences to guide decisions – remember the inductivist turkey)



Few statements

All models are wrong but some are useful

(and some are more useful than others)

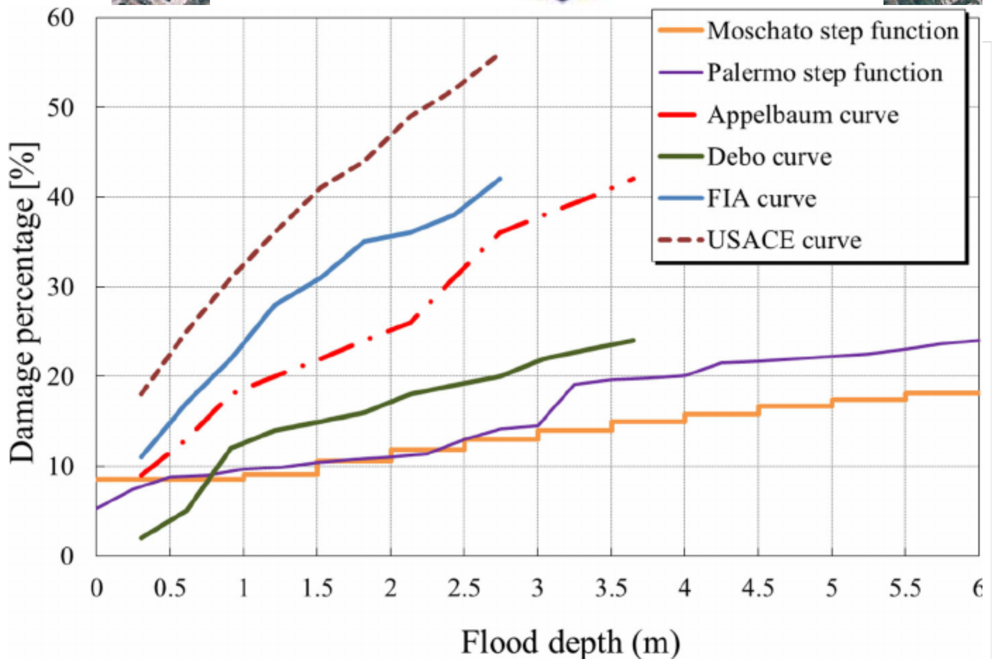
(All data are useful, but some are more uncertain than others)

Common components of a pluvial flooding model

- Rainfall – runoff

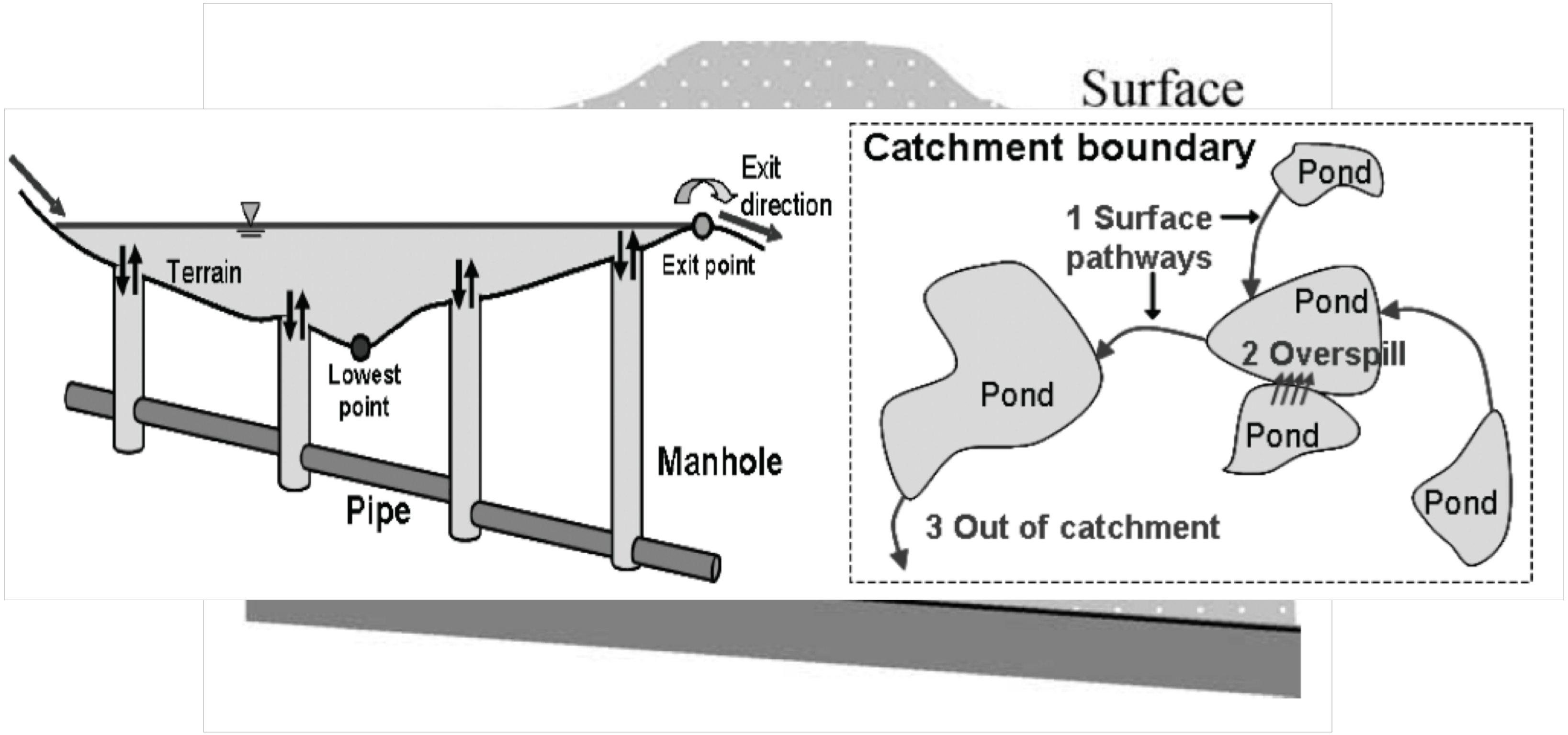
- Hydraulic propagation

- (Damage assessment)



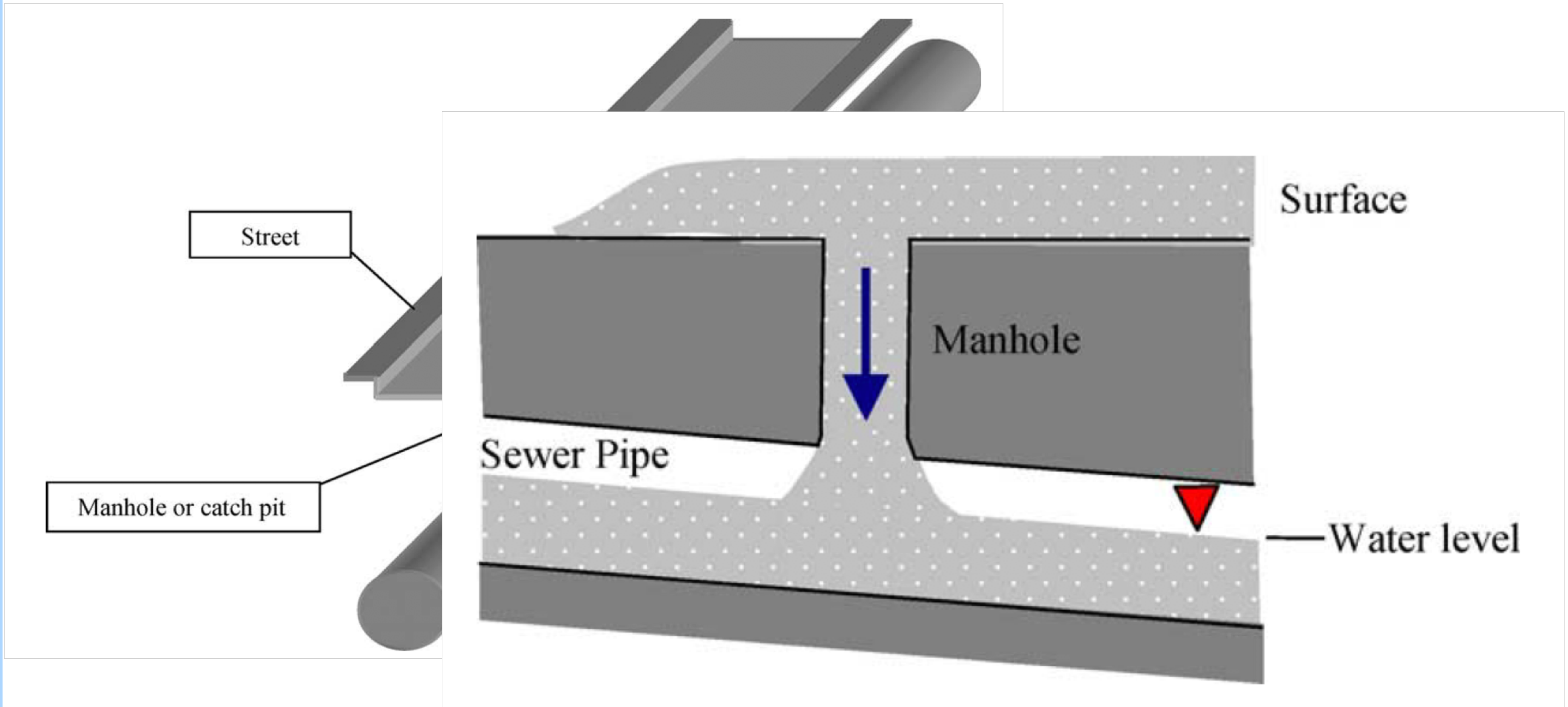
A short evolution of urban flooding propagation models

- Simple surcharge sewer model ('80)



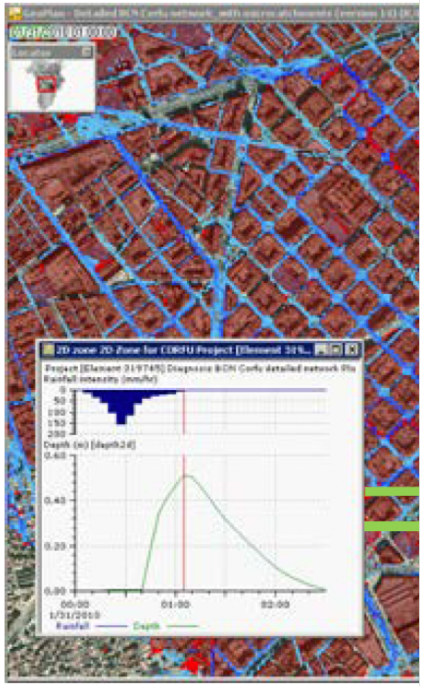
A short evolution of urban flooding propagation models

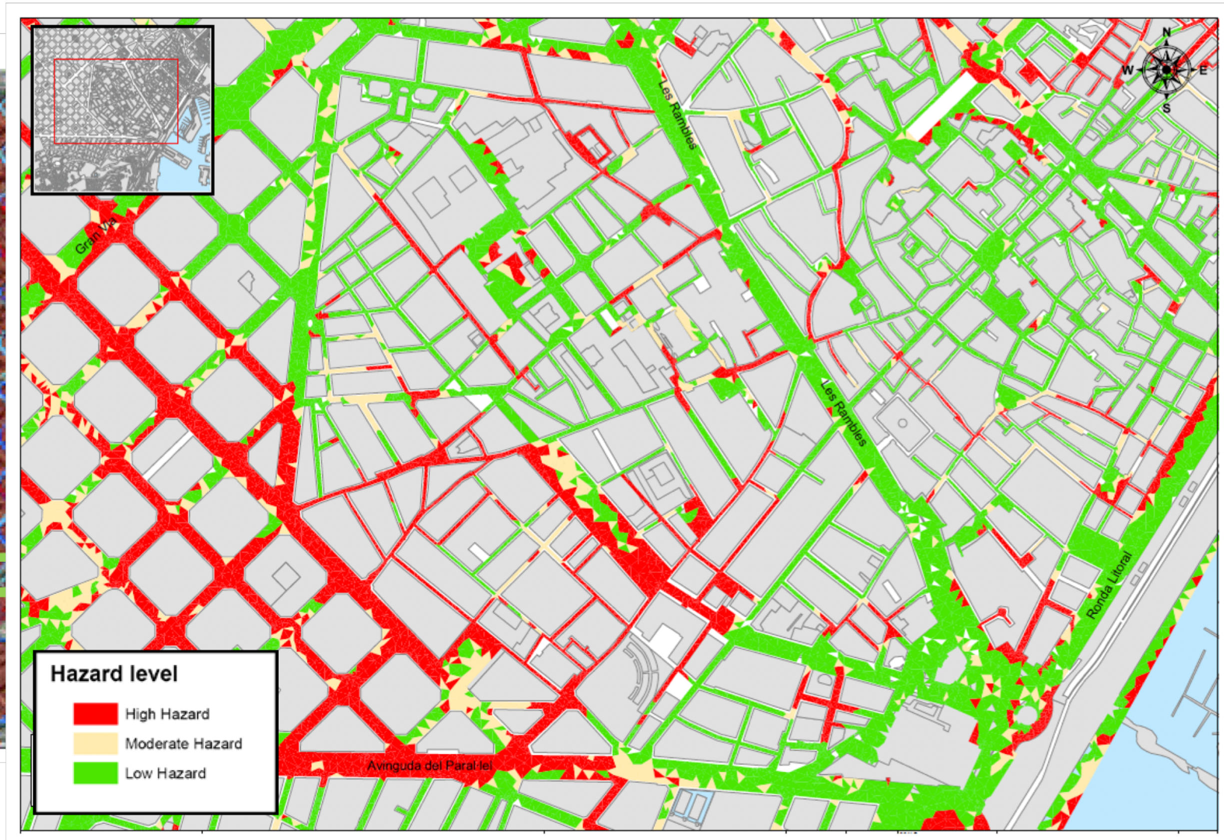
- 1D – 1D channel model ('90 - 2000)



A short evolution of urban flooding propagation models

- 1D sewer model – 2D overland model (2000 -)





Questions we ask about models (in general)

- Is the model valid?
- Are the assumptions reasonable?
- Are data requirements acceptable?
- Is the model credible?
- Do the model predictions match the observed data?
- How uncertain are the results?

What is a good model?

*Simple, realistic, data efficient, useful, reliable, **valid** etc*

Università di Enna «Kore»
Centro di ricerca La.R.A.

Question we ask about urban flood models

- Is the model able to represent the complex geometry of the system?
- Is the model resource intensive?
- Are relevant pr… neglected…
- Is the mod… evalu…
- …esses
- …e to go real time?
- …re available data sufficient?

…or modeller?

…nplex to be inapplicable, not too simple to be …able

All the questions can be answered in a Bayesian framework.

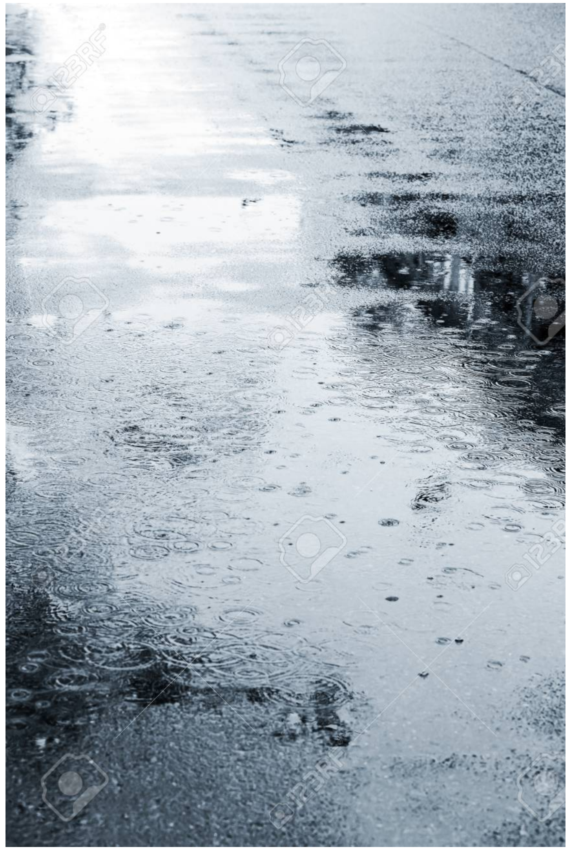
For Bayesians, the concept of probability is extended to cover degrees of certainty about statements.

My personal perspective to urban flooding uncertainty sources

- Temporal and spatial variability of rainfall

My personal perspective to urban flooding uncertainty sources

- Dry or wet initial conditions



My personal perspective to urban flooding uncertainty sources

- Inlet clogging

My personal perspective to urban flooding uncertainty sources

- Sewer clogging

My personal perspective to urban flooding uncertainty sources

- Goods and furniture potentially exposed to flooding

Families of uncertainty sources

- Uncertainty in model quantities/parameters/ inputs
- Uncertainty about model form
- Uncertainty about model completeness
- Lack of observations contribute to
  - uncertainties in input data
  - parameter uncertainties
- Conflicting evidence contributes to
  - uncertainty about model form
  - Uncertainty about validity of assumptions

Making it difficult to judge how good a model is!!

Tools to cope with uncertainties

- Data assimilation is a mathematical discipline that seeks to optimally combine theory (usually in the form of a numerical model) with observations.
- Sensitivity Analysis studies how much each individual source of uncertainty contributes to the output variance
- Uncertainty Analysis focuses on how uncertainty in the input factors propagates through the structure of the model and affects the values of the output

To understand SA and UA: backward reasoning

Most people, if you describe a train of events to them will tell you what the result will be. There will be few people however, that if you told them a result, would be able to evolve from their own consciousness what the steps were that led to that result. This is what I mean when I talk about reasoning backward.
— Sherlock Holmes, A Study in Scarlet, Sir Arthur Conan Doyle (1887)

Modellers conduct SA to determine

(a) if a model resembles the system or processes under study,

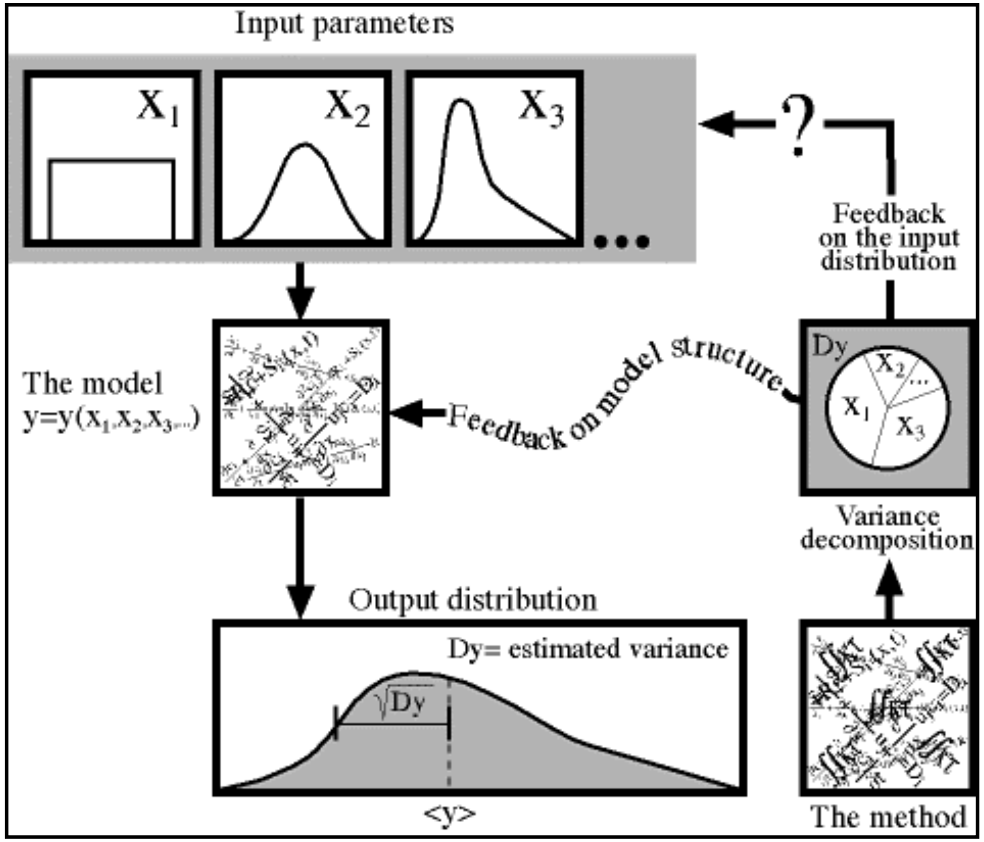
(b) the factors that mostly contribute to the output variability,

(c) the model parameters (or parts of the model itself) that are insignificant,

(d) if there is some region in the space of input factors for which the model variation is maximum,

(e) if and which (group of) factors interact with each other.

SA flow chart (Saltelli, Chan and Scott, 2000)



Design of the SA experiment

- Simple factorial designs (one at a time)
- Factorial designs (including potential interaction terms)
- Fractional factorial designs

SA techniques

- Screening techniques
  - *O(ne) A(t) T(ime), factorial, fractional factorial designs used to isolate a set of important factors*
- Local/differential analysis
- Sampling-based (Monte Carlo) methods
- Variance based methods
  - *variance decomposition of output to compute sensitivity indices*

Screening

- **screening experiments** can be used to identify the parameter subset that controls most of the output variability with low computational effort.

Screening methods

- Vary one factor at a time (NOT particularly recommended)
- Morris OAT design (global)
  - Estimate the *main effect* of a factor by computing a number r of local measures at different points $x_1,\ldots,x_r$ in the input space and then average them.
  - Order the input factors

Local SA

- **Local SA** concentrates on the local impact of the factors on the model. Local SA is usually carried out by computing partial derivatives of the output functions with respect to the input variables.
- The input parameters are varied in a small interval around a nominal value. The interval is usually the same for all of the variables and is not related to the degree of knowledge of the variables.

Global SA

- **Global SA** apportions the output uncertainty to the uncertainty in the input factors, covering their entire range space.
- A global method evaluates the effect of x_j while all other $x_i, i \neq j$ are varied as well.

On the other hand - Uncertainty analysis

- Parameter uncertainty
  - *usually quantified in form of a distribution.*
- Model structural uncertainty
  - *more than one model may be fit, expressed as a prior on model structure.*
- Scenario uncertainty
  - *uncertainty on future conditions.*

Tools for handling uncertainty

- Parameter uncertainty
  - *Probability distributions and Sensitivity analysis*
- Structural uncertainty
  - *Bayesian framework*
  - *one possibility to define a discrete set of models, other possibility to use a Gaussian process*

Sources of uncertainty

- Errors in the input and boundary condition data
- Errors in the model structure
- Errors in estimates of parameter values
- Commensurability of modelled and observed variables and parameters
- Errors in the observations used to calibrate or evaluate models
- Errors of omission (not always the unknown unknowns)

Difficult (*impossible*) to disentangle different sources of error without making strong assumptions (Beven, 2005)

Università di Enna «Kore»
Centro di ricerca La.R.A.

Types of Uncertainty

Aleatory Uncertainty

Epistemic Uncertainty

System Dynamics

Forcing and Response Data

Disinformation

Semantic/Linguistic Uncertainty

Ontological Uncertainty

Type of Uncertainty	Description
Aleatory Uncertainty	Uncertainty with stationary statistical characteristics. May be structured (bias, autocorrelation, long term persistence) but can be reduced to residual stationary random component
Epistemic Uncertainty (system dynamics)	Uncertainty arsing from a lack of knowledge about how to represent the catchment system in terms of both model structure and parameters. Note that this may include things that are included in the perceptual model of the catchment processes but which are not included in the model. They may also include things that have not yet been perceived as being important but which might result in reduced model performance.
Epistemic Uncertainty (forcing and response data)	Uncertainty arising from lack of knowledge about the forcing data or the response data with which model outputs can be evaluated. This may be because of commensurability or interpolation issues when not enough information is provided by the observational techniques to adequately describe variables required in the modelling process.
Epistemic Uncertainty (disinformation)	Analogous to known unknowns (in either system representation or forcing data that are *known* to be inconsistent or wrong. Will have the expectation of introducing disinformation into the modelling processes resulting in biased or incorrect inference (including false positives and false negatives in testing models as hypotheses)
Semantic / Linguistic Uncertainty	Uncertainty about what statements or quantities in the relevant domain actually mean (there are many examples in hydrology including storm runoff, baseflow, hydraulic conductivity, stationarity etc). This can partly result from commensurability issues that quantities with the same name have different meanings in different contexts or scales.
Ontological Uncertainty	Uncertainty associated with different belief systems. Relevant example here might be beliefs about whether formal probability is an appropriate framework for the representation of mode errors. Different beliefs about the appropriate assumptions could lead to very different uncertainty estimates.

Beven, Leonardo Lecture: Facets of Uncertainty, Hyd.Sci.J. 2014

Aleatory Uncertainty

Uncertainty with stationary statistical characteristics. May be structured (bias, autocorrelation, long term persistence) but can be reduced to residual stationary random component

Important because:
Full power of statistical theory can be used to estimate the probability of matching a new sample or observation conditional on the model

Epistemic Uncertainty (System Dynamics)

Epistemic uncertainty arising from a lack of knowledge about how to represent the catchment system in terms of both model structure and parameters. This may include things that are included in the perceptual model of the catchment processes but which are not included in the model. They may also include things that have not yet been perceived as being important but which might result in reduced model performance.

Important because:
May result in non-stationarity in residual characteristics that if it cannot be represented explicitly might lead to overconfidence in inference

Epistemic Uncertainty (Forcing and Response Date)

Epistemic Uncertainty arising from lack of knowledge about the forcing data or the response data with which model outputs can be evaluated. This may be because of commensurability or interpolation issues when not enough information is provided by the observational techniques to adequately describe variables required in the modelling process.

Important because:

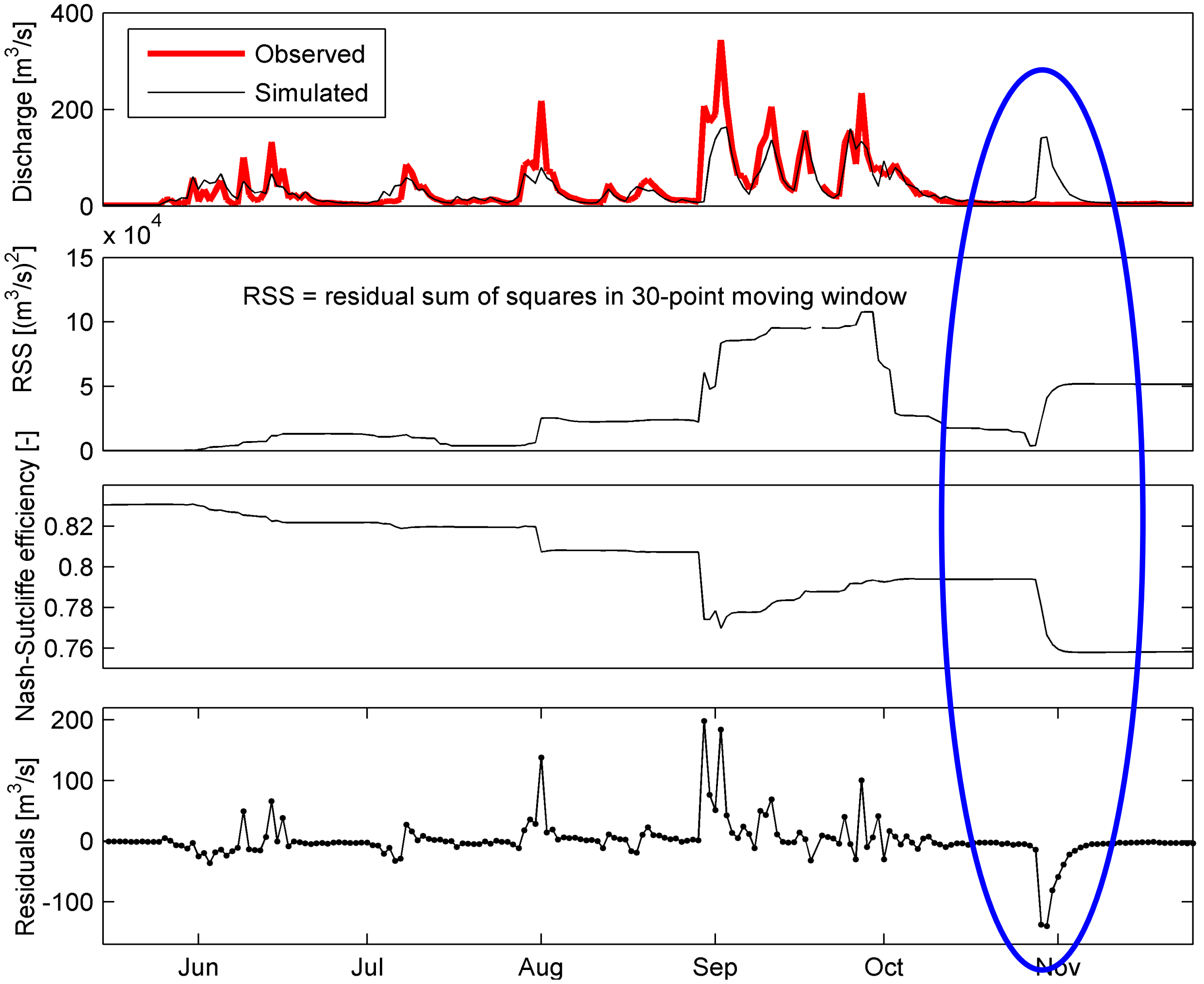
May result in non-stationarity in residual characteristics that if it cannot be represented explicitly might lead to overconfidence in inference

Epistemic Uncertainty (Disinformation)

Analogous to known unknowns (in either system representation or forcing data that are *known* to be inconsistent or wrong.
Will have the expectation of introducing disinformation into the modelling processes resulting in biased or incorrect inference (including false positives and false negatives in testing models as hypotheses)

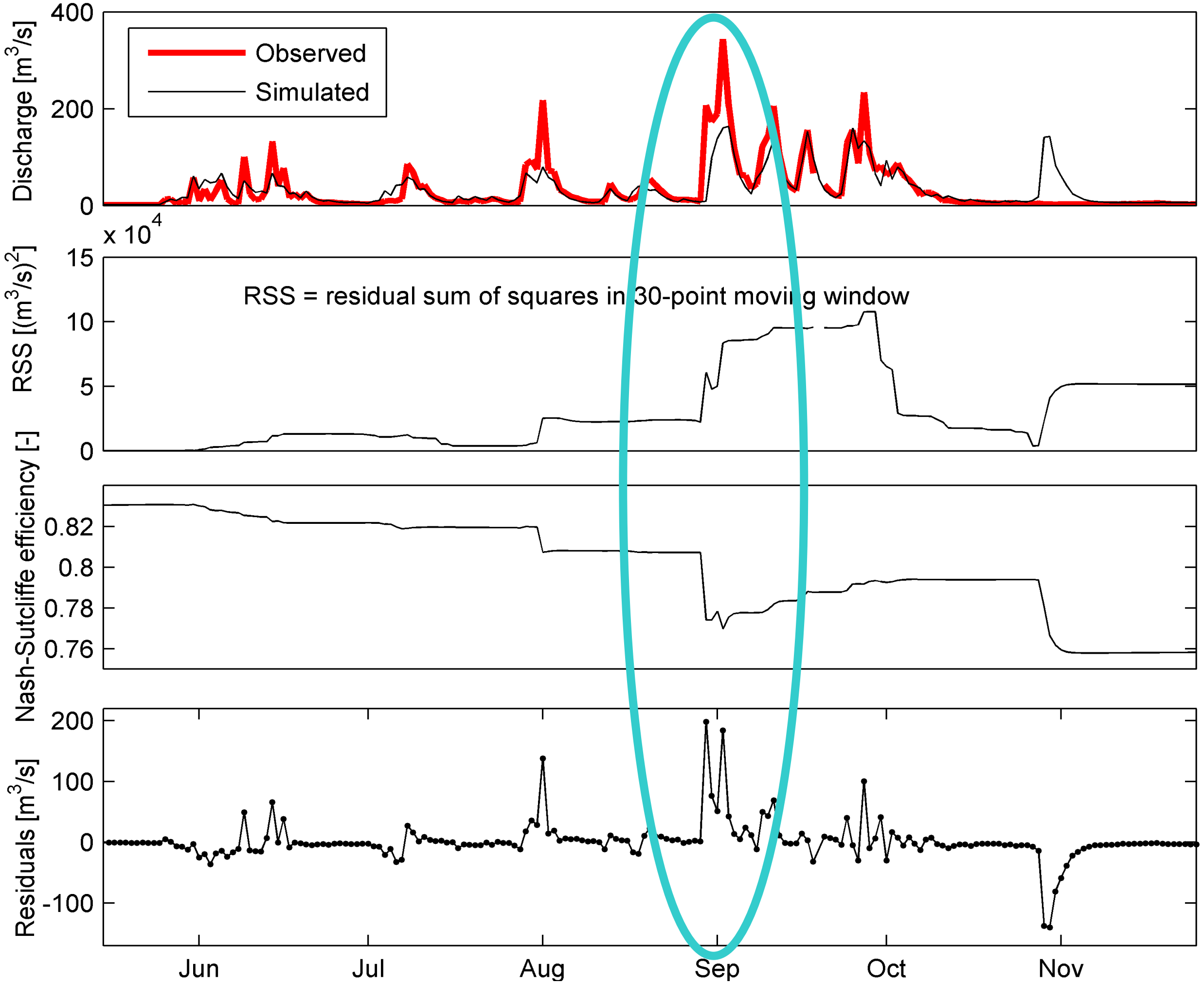
Important because:
May feed disinformation into the model identification process

Disinformation in calibration data



Università di Enna «Kore»
Centro di ricerca La.R.A.

Disinformation in calibration data



Semantic/Linguistic Uncertainty

Uncertainty about what statements or quantities in the relevant domain actually mean (there are many examples in hydrology including storm runoff, baseflow, hydraulic conductivity, stationarity etc). This can partly result from commensurability issues that quantities with the same name have different meanings in different contexts or scales.

Important because:

May limit effective information content in inference if compare variables that do not have the same meaning

Ontological Uncertainty

Uncertainty associated with different belief systems. Relevant example here might be beliefs about whether formal probability is an appropriate framework for the representation of mode errors. Different beliefs about the appropriate assumptions could lead to very different uncertainty estimates.

Important because:
All the previous definitions may be subject to ontological uncertainty.

Types of error and why they are important
Aleatory (A), Epistemic (E) or Disinformative (D)

- Errors in the input and boundary condition data (A/E/D)
- Errors in the model structure (E/D?)
- Errors in estimates of parameter values (A/E)
- Commensurability of modelled and observed variables and parameters (A/E/D)
- Errors in the observations used to calibrate or evaluate models (A/E/D)
- Errors of omission (sometimes known omissions) (E/D?)
- The unknown unknowns (D?, becoming E/D)

Uncertainty about uncertainty estimation

- Many sources of uncertainty in the modelling process but can generally only evaluate the differences (residuals) between some observed and predicted variables (e.g. water levels, discharges, ……)

- Leaves lots of scope for different interpretations and assumptions about the nature of different sources of uncertainty

- Model structural error particularly difficult to assess (not easily separated from input and other uncertainties without making strong and difficult to justify assumptions) – often assessed AS IF model is correct

- Therefore lots of uncertainty estimation methods

Statistical Uncertainty Estimation

- Treat the optimal model as if it were the "true" model
- Fit a model to the residuals using appropriate assumptions (e.g. residuals are of zero mean and constant variance and uncorrelated in time/space - or something more realistic, with bias, non-constant variance {heteroscedasticity}, and correlated residuals)
- Nature of error model defines a likelihood function
- Sum of model + error distribution can be used to estimate likelihood (probability) of predicting an observation given the model
- Problem that treating multiple sources of error as if all "measurement error"

But what if residual series shows complex (non-stationary?) structure?

Many reasons why error structures might be complex

- Inconsistent observations
- Effects of processing input error through model structural error in time and space (might lead to non-stationary correlation…)
- Incommensurable observed and predicted variables
- Sample from long term stationary stochastic behaviour
- ….

Alternative Frameworks

Allowing that epistemic uncertainties may not be represented as simple statistical functions

- Informal likelihoods (no longer P(O|M) but can be used to estimate P(M|O))
- Fuzzy methods
- Dempster-Shafer Evidence Theory
- ....

May be inefficient and qualitative – HANDLE WITH CARE!!!

Uncertainty as a likelihood surface in the model space

Basic requirements of a likelihood as belief

- Should be higher for models that are "better"
- Should be zero for models that do not give useful results
- Scaling as relative belief in a hypothesis rather than probability

But how then best to determine weights from evidence given epistemic uncertainties??

A general framework for model evaluation

1. Eliminate obviously disinformative data

2. Set up limits of acceptability (reflecting observation error, commensurability error and input error) ***prior*** to running the model.

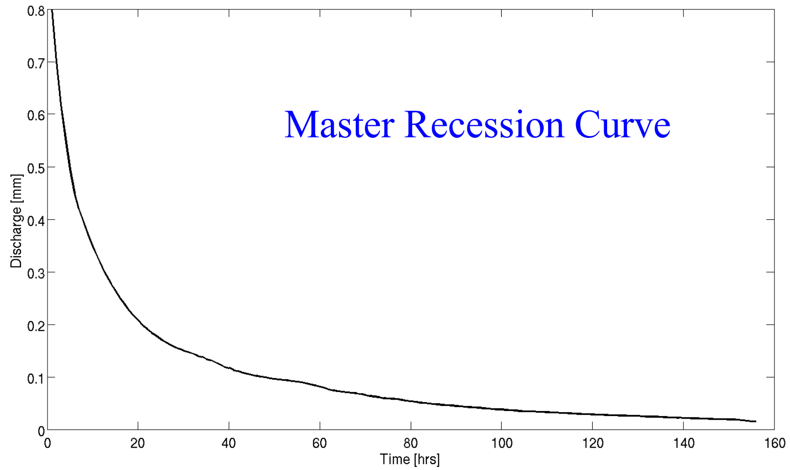
3. For each model run, evaluate performance against limits of acceptability

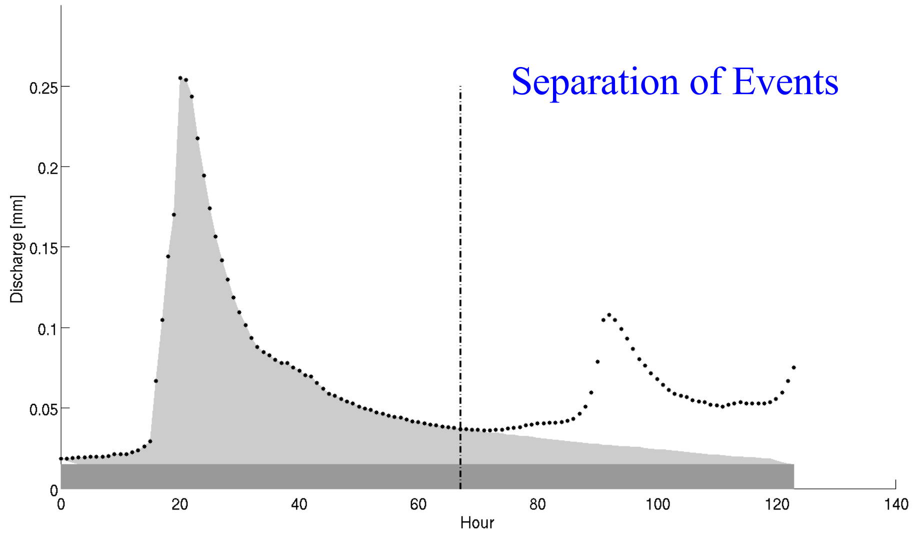
4. Check for error reconstruction to improve predictions / calculate distributions of errors.

Identifying disinformative data

First criterion: Event mass balance consistency (expectation that event runoff coefficient Q / R will be less than one)

But…difficulty of separating events



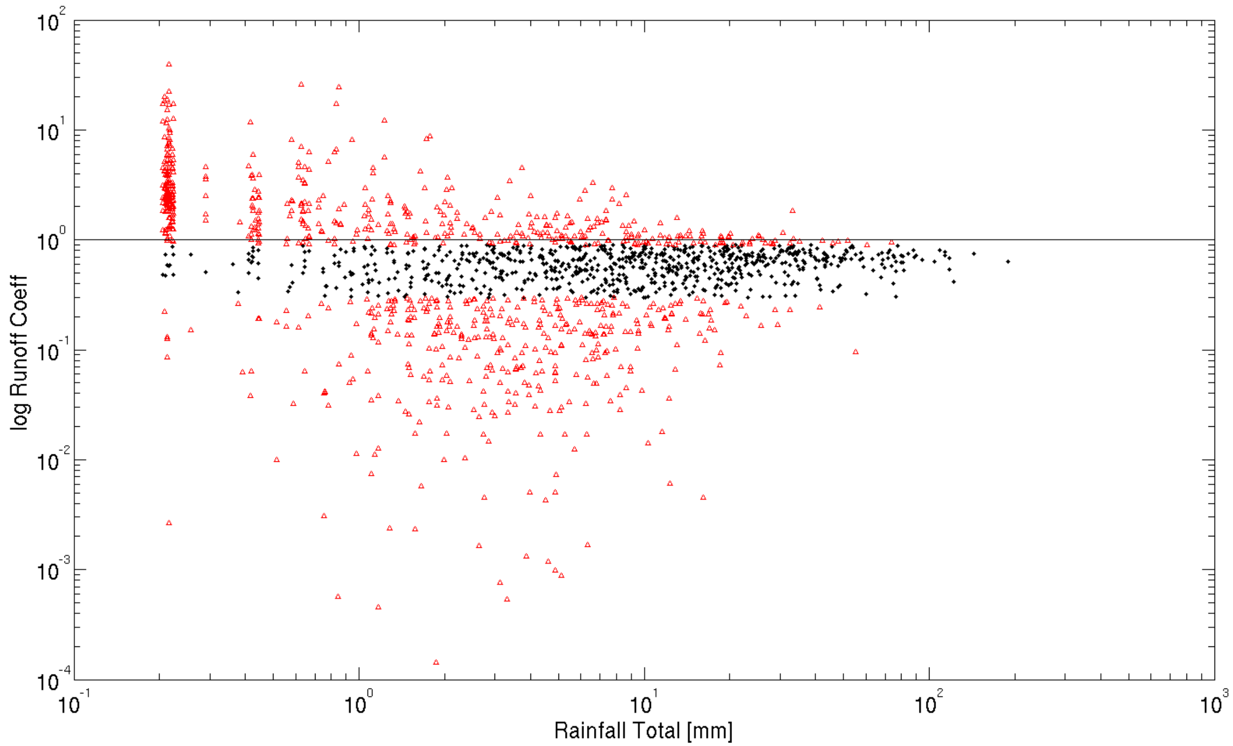


and impact of an inconsistent event on model results might persist for following events, gradually decaying

Università di Enna «Kore»
Centro di ricerca La.R.A.

Setting Limits of Acceptability prior to running a model

Results of runoff coefficient determination for River Tyne at Station 23006 – plotted against rainfall totals over catchment area as estimated from 5 gauges (black – range 0.3 to 0.9)



Beven et al, HESS, 2011

Limits of acceptability

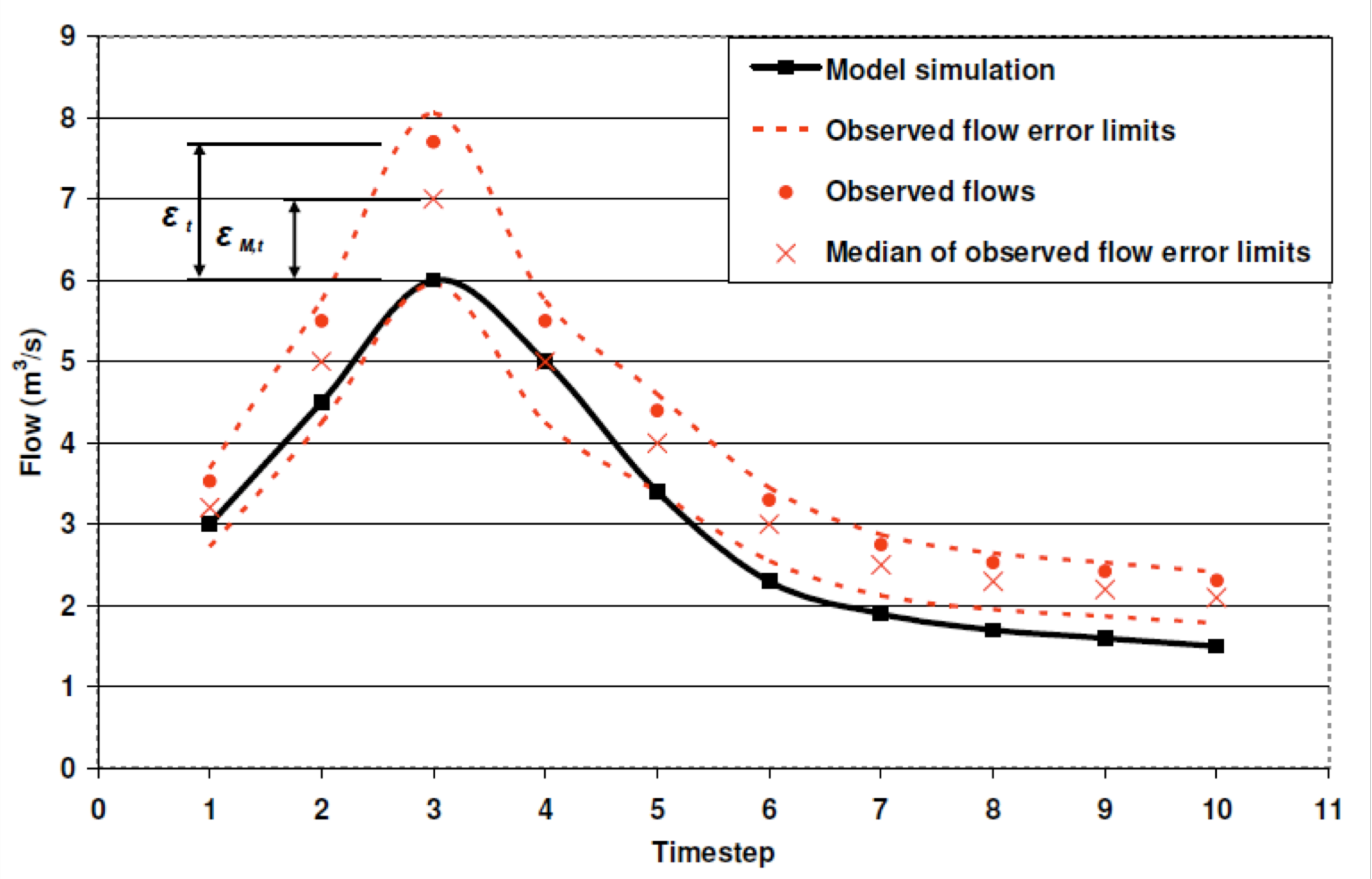
- The question that then arises within this framework is whether, for an particular realisation of the inputs and boundary conditions, $\varepsilon_M(\theta, I, \varepsilon_I, x, t)$ is acceptable in relation to the terms $\varepsilon_O(x,t) + \varepsilon_C(\Delta x,\Delta t, x,t)$. This is equivalent to asking if the following inequality holds:

$$O_{min}(x,t) < M(\theta, I, \varepsilon_I, x, t) < O_{max}(x,t) \text{ for all } O(x,t)$$

  where $O_{min}(x,t)$ and $O_{max}(x,t)$ are acceptable limits for the prediction of the output variables given $\varepsilon_O(x,t)$ and $\varepsilon_C(\Delta x,\Delta t, x,t)$
- Idseally, limits of acceptability should be evaluated prior to running the model (but note I,ε_I in $M(\theta, I, \varepsilon_I, x, t)$)

Model Evaluation using Limits of Acceptability

Likelihood can be developed based on scaled deviation away from observation, with zero value at any time step that prediction lies outside limits.

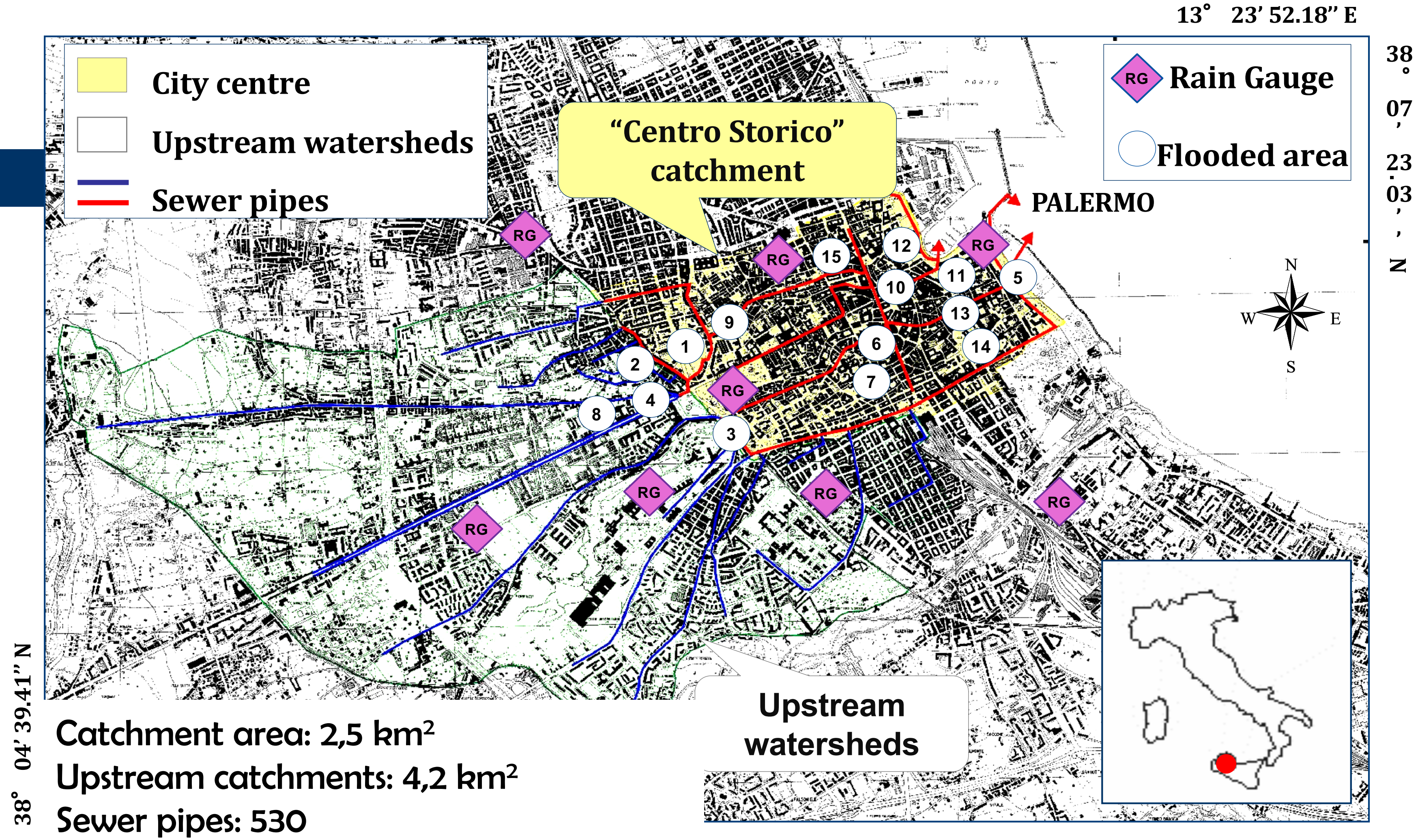


Our "sand-box" case: Palermo (IT)

- Example 1 (Distributed Input Uncertainty - Rainfall)
  - *Aleatory: random metering error*
  - *Epistemic: not considering a raingauge*
  - *Disinformative: shifting series in time*
- Example 2 (Lumped Parameter uncertainty)
  - *Aleatory: variability due to random factors (accidental events, initial conditions)*
  - *Epistemic: not directly measurable*

Università di Enna «Kore»
Centro di ricerca La.R.A.

Our “sand-box” case study: Palermo city centre (Italy)



Catchment area: 2,5 km^2
Upstream catchments: 4,2 km^2
Sewer pipes: 530
Main sewers lenght: 56 km
Flooded locations: 15
Monitored flooding events: 36 in the period 1994 - 2011

The available database: rainfall and flooding data

Rainfall data were collected in 7 rain gauges located inside and outside the analysed area for all the events with a temporal resolution of 1-5 minutes.

Flooding data were collected by the Fire Brigades via remote cameras and field reports (more than 700 reports on damaged properties).

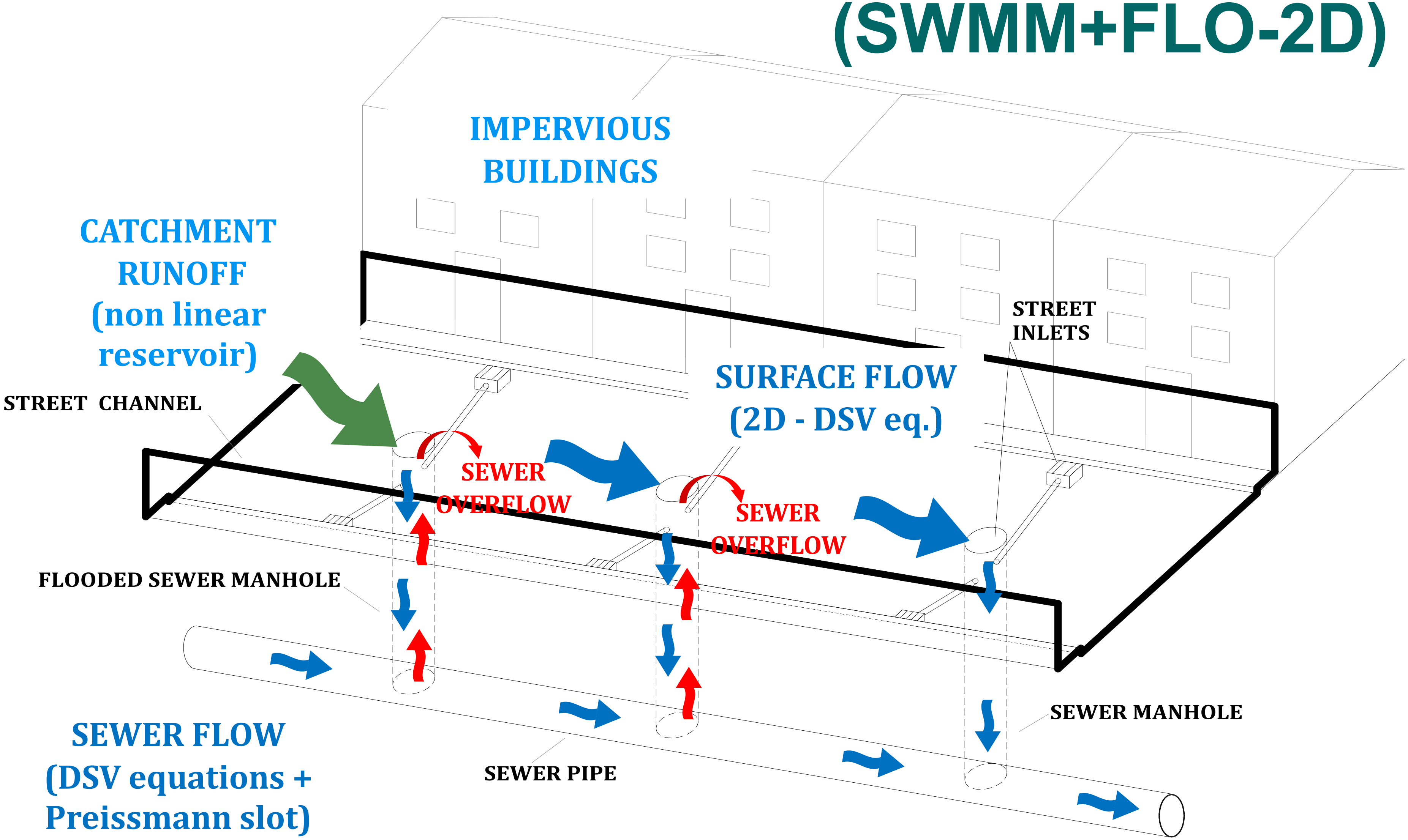
Flooded areas	Mean flooding depth [cm]	Average return period [yrs]	Flooded areas	Mean flooding depth [cm]	Average return period [yrs]
1	144,9	0,54	9	44,4	0,56
2	63,4	0,56	10	25,7	0,71
3	50,7	0,68	11	28,7	0,56
4	45,9	0,56	12	54,6	0,54
5	59,9	0,56	13	43,2	0,54
6	32,2	0,56	14	38,4	0,56
7	28,3	0,68	15	24,6	1,5
8	47,0	0,59			

Our “sand-box” case: Palermo (IT)

- Example 3 (Damage model structure)
  - *Epistemic: some damage processes are not considered or investigated*
  - *Disinformative: some initial assumptions may be false*

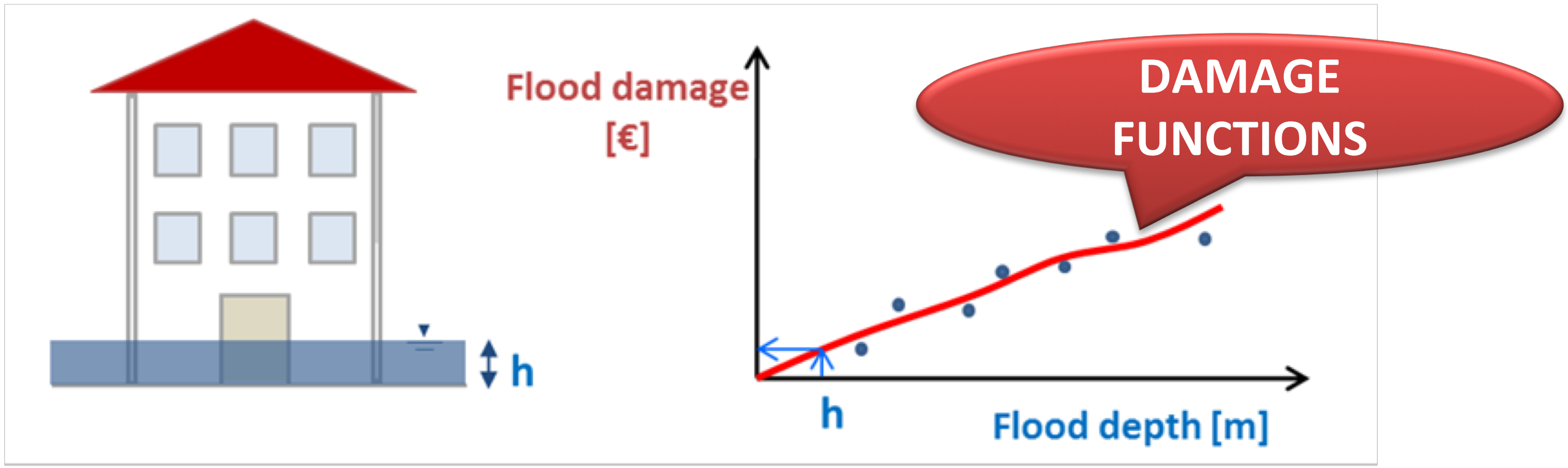
Università di Enna «Kore»
Centro di ricerca La.R.A.

Tools: Flooding 2D – 1D mathematical model (SWMM+FLO-2D)



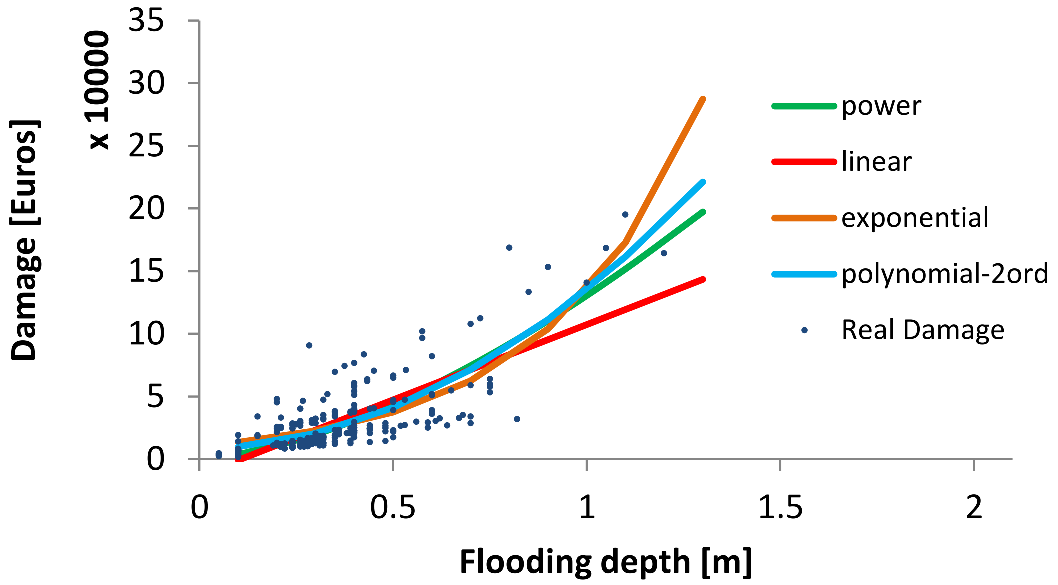
Urban flood damage appraisal

- The expected flood damage can be evaluated by an ex-ante analysis.
- Real damage data are interpolated by means of damage functions describing the relationship occurring between the hydraulic characteristics of flood and the related damage



Uncertainty inherent to flood damage function

- Several regression laws with different level of simplification can be used as depth-damage functions thus influencing the damage appraisal obtained.



- Flooding data are often piecemeal, affected by measuring errors and spatially aggregated. In consequence, the flood damage assessment is usually affected by a degree of intrinsic uncertainty that cannot be realistically eliminated

Distributed Input Uncertainty

Step 1 - Rainfall temporal resampling

12 historical rainfall time series recorded between 1993-1997 with high temporal resolution (1-3 min) by a network of 8 raingauges in the analyzed area was adopted

For each historical rainfall event, the duration was divided into n equal time intervals ranging from 5 to 15 min with step of 1 min and for each of these temporal windows the average intensity was evaluated.

For each historical rainfall event 11 hyetographs with a coarser temporal resolution were obtained

According to the mass balance principle, for each time step, the real event and the re-sampled one are characterised by the same rainfall volume.

Distributed Input Uncertainty

STEP 2 - Rainfall spatial resampling

Rainfall data from some of the raingauges present in the watershed were neglected to analyse the influence of the spatial rainfall aggregation on the efficiency of rainfall-runoff model predictions

- The number of sampled raingauges was selected between 1 and 8 and all possible combinations of raingauges were evaluated.
- For each combination, urban sub-catchments were linked to the closer raingauge according to the minimum Euclidean distance criterion.

The total number of raingauge combinations was **255**

The analysed spatial and temporal resolution scenarios were in total **2805** (255x11).

Distributed Input Uncertainty

STEP 3 – Evaluation of rainfall performance indicators

To quantify the goodness of rainfall estimates with respect to the reference rainfalls adopting the whole raingauge network (8 RG)

To evaluate the dependencies between model performance and accuracy in the description of a rainfall event in time and space

BALANCE and GORE indices *(Andréiassian et al., 2001)* were computed:

$$BALANCE = \frac{\sum_{i=1}^{n} P_i^E}{\sum_{i=1}^{n} P_i} \qquad GORE = 1 - \frac{\sum_{i=1}^{n}\left(\sqrt{P_i^E}-\sqrt{P_i}\right)^2}{\sum_{i=1}^{n}\left(\sqrt{P_i}-\overline{\sqrt{P_i}}\right)^2}$$

BALANCE quantifies the overestimation (when >1) or underestimation (when <1) of the rainfall volume in each analysed scenario.

GORE compares the sum of squared errors in the rainfall estimate in each scenario to the temporal variance of the reference precipitation (obtained with the maximum temporal resolution and the whole raingauge network).

Distributed Input Uncertainty

STEP 4 – Modelling efficiency evaluation

- Modelling results obtained with historical rainfall events were compared with those achieved using coarser estimate rainfall data

The uncertainty inherent in flood modelling results was assessed in terms of the Nash and Sutcliff criterion (N-S) adopted as likelihood measure within the GLUE analysis

$$1 - \frac{{\sigma^2}_e}{{\sigma^2}_o}$$

Distributed Input Uncertainty

Methodology application to a case study

At first,

the mathematical model was calibrated for all 2805 considered scenarios by defining 500 random Monte Carlo sets of parameters (rainfall – runoff parameters and drainage system roughness) and then selecting the one providing the highest N-S criterion computed on flooding depths.

Then,

the rainfall performance indicators were evaluated and compared with model efficiency in the flood depths appraisal;

the model uncertainty was correlated to the availability of rainfall data considering the impact of different possible combinations of a fixed number of raingauges

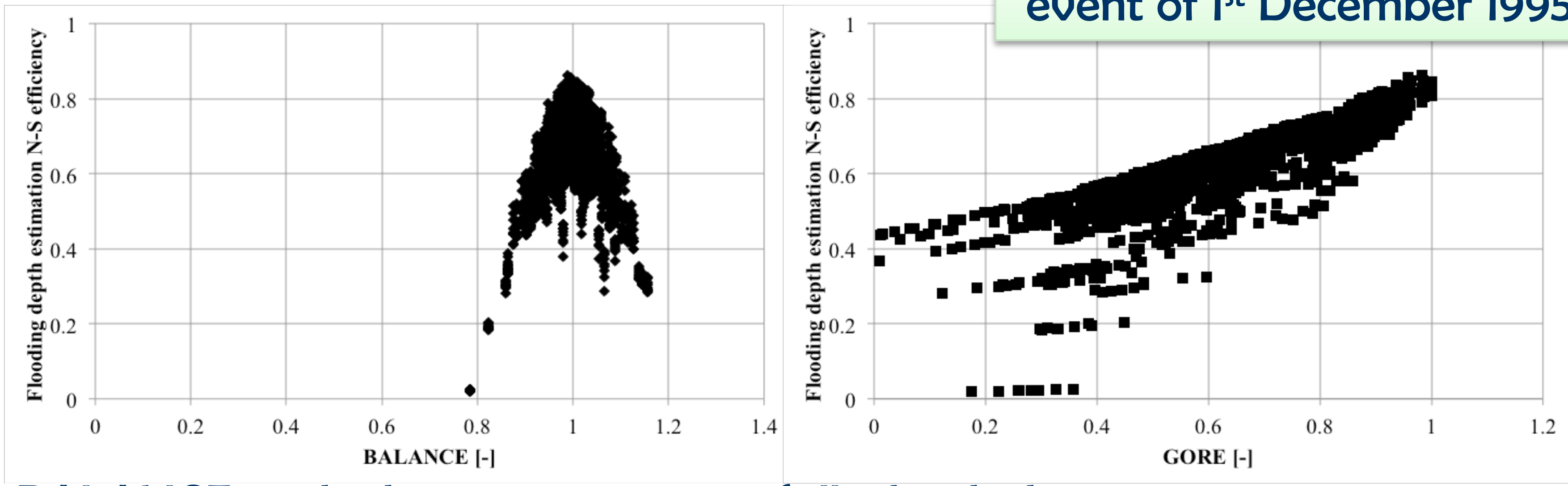
Università di Enna «Kore»
Centro di ricerca La.R.A.

Distributed Input Uncertainty

BALANCE and GORE indices *vs* model efficiency N-S

Comparison of rainfall indices and N-S criterion computed on the flooding depths obtained in each of the 15 flooding locations of the watershed by the simulations carried out for each of the 2805 considered scenario

event of 1st December 1995

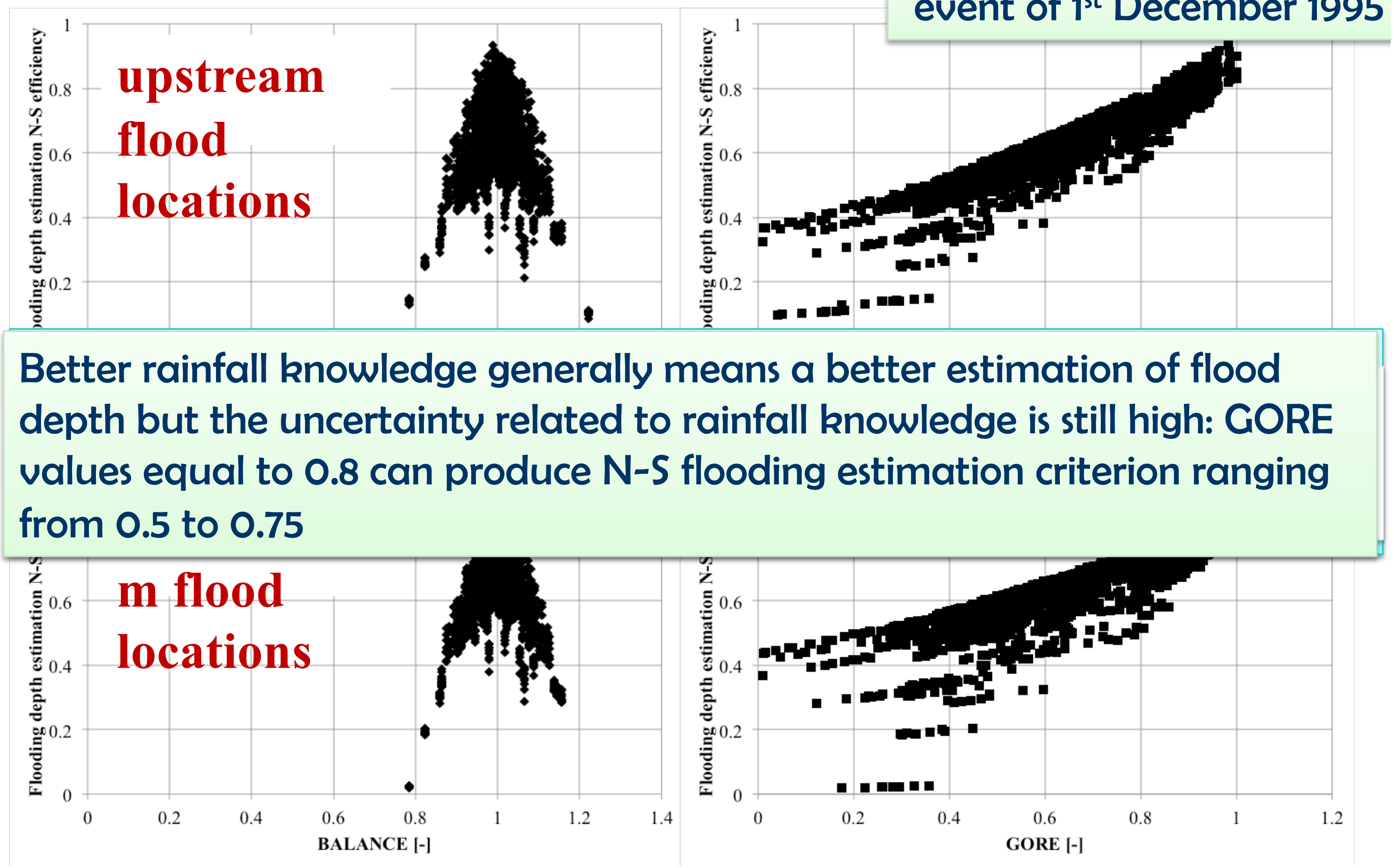


BALANCE is higher sentitive: rainfall depth has a major impact on flooding; a wrong estimation of rainfall can impact negatively on model performance

The ratio between GORE and N-S is nearly linear: the model, during calibration phase, is able to slightly compensate the wrong estimation of rainfall input obtaining a reliable estimation of the flooding depth.

Distributed Input Uncertainty

BALANCE and GORE indices *vs* model efficiency N-S separating the upstream and downstream flooding areas



Better rainfall knowledge generally means a better estimation of flood depth but the uncertainty related to rainfall knowledge is still high: GORE values equal to 0.8 can produce N-S flooding estimation criterion ranging from 0.5 to 0.75

Distributed Input Uncertainty

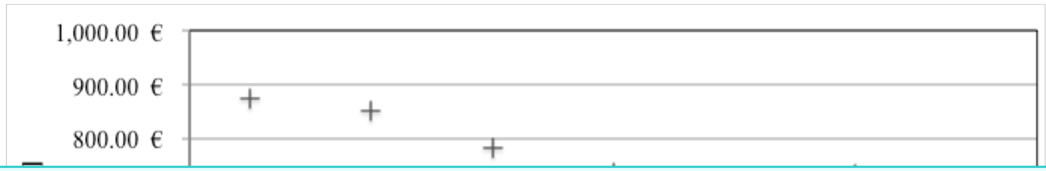
Uncertainty affecting flood damage estimates

A GLUE analysis was carried out by running 500 Monte Carlo Simulations changing model parameters for each of the 2805 considered scenarios

To estimate total monetary flood damage depth-damage curves for **buildings** and **vehicles** obtained by available data in the area were used

$$D = 867.85 \cdot h^{0.8409}$$

$$D = 1035.70 \cdot h^{0.5110}$$



Uncertainty bands (5th and

- The model is generally able for providing a good estimate of the measured damage.
- Total monetary damage appraisal is in the range of ±10% around the measured value.
- The calibration efficiency is progressively better if more rainfall data are available but the added value of the single raingauge after the forth is not relevant.

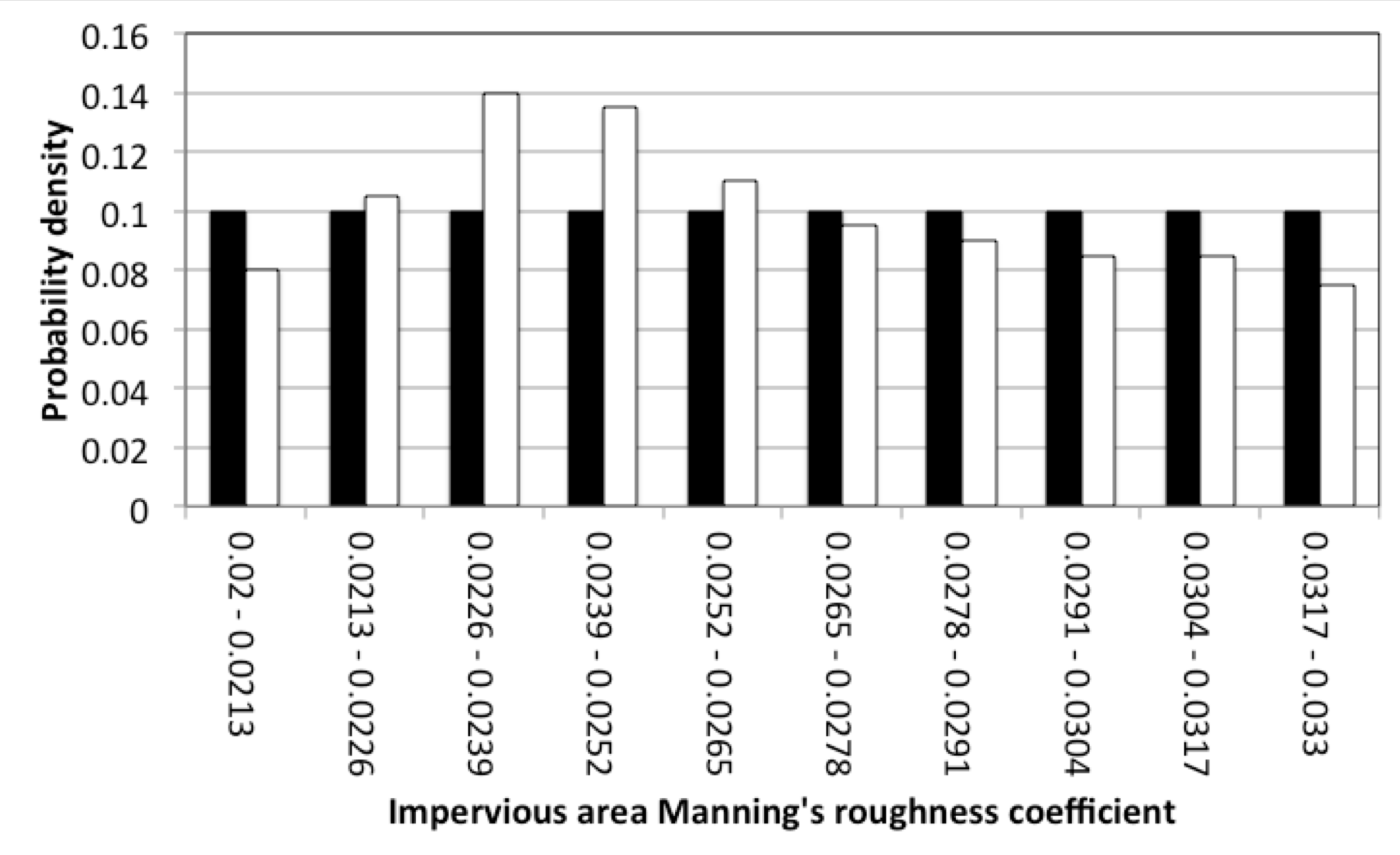
The application of the Bayesian analysis

- The dataset was divided in three parts in order to test the ability of Bayesian uncertainty analysis to learn from available data

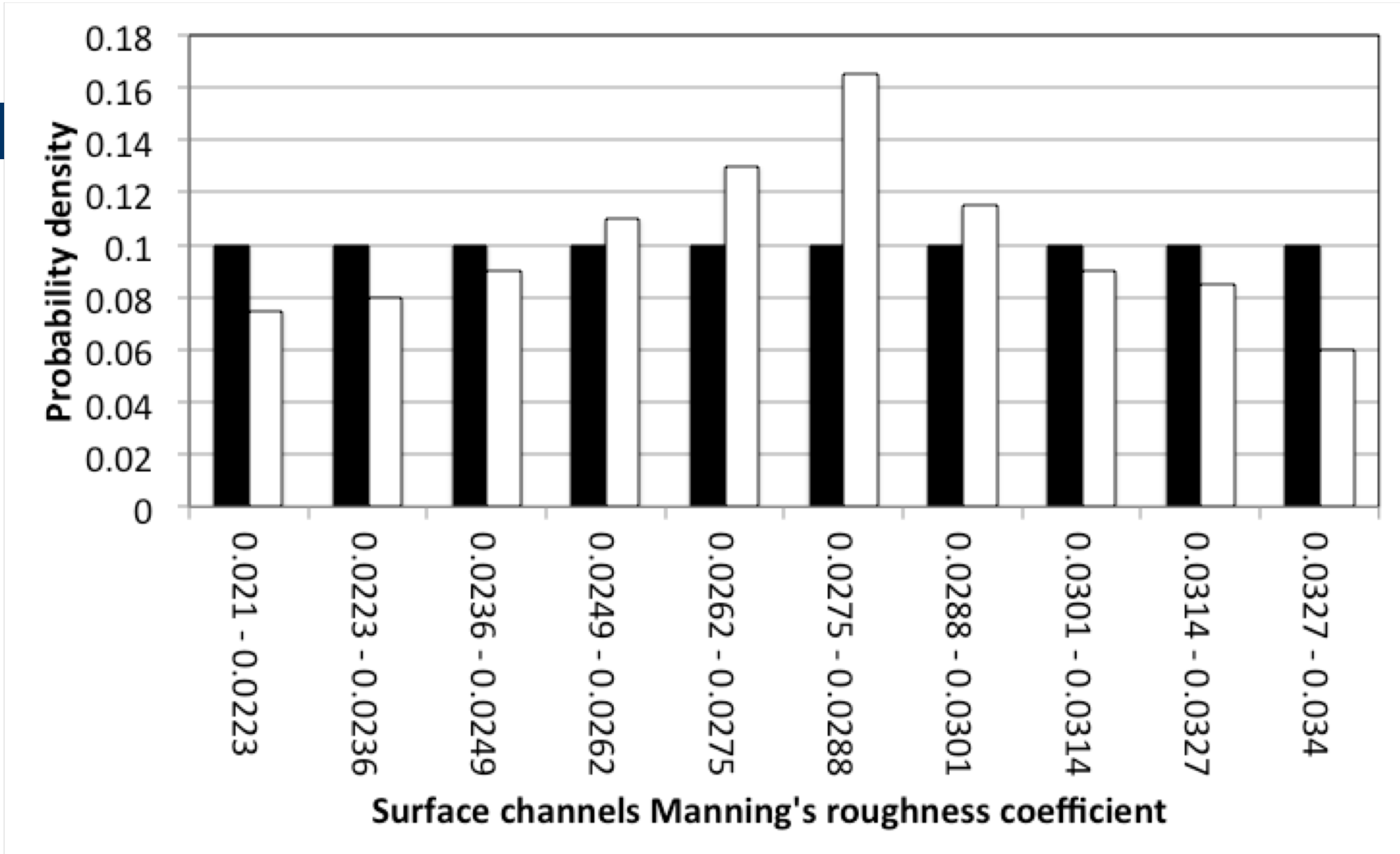
- The first Bayesian update starts from uniform parameter distributions

Parameters	M.U.	Min	Max
Impervious area surface storage	mm	0.5	2.0
Pervious area surface storage	mm	3.5	8.5
Impervious area Manning's roughness	-	0.020	0.033
Pervious area Manning's roughness	-	0.025	0.050
Max infiltration rate (Horton)	mm/h	62.0	117.2
Saturated soil infiltration rate (Horton)	mm/h	12.2	22.7
Underground drainage system Manning's roughness	-	0.014	0.025
Surface channels Manning's roughness	-	0.021	0.034

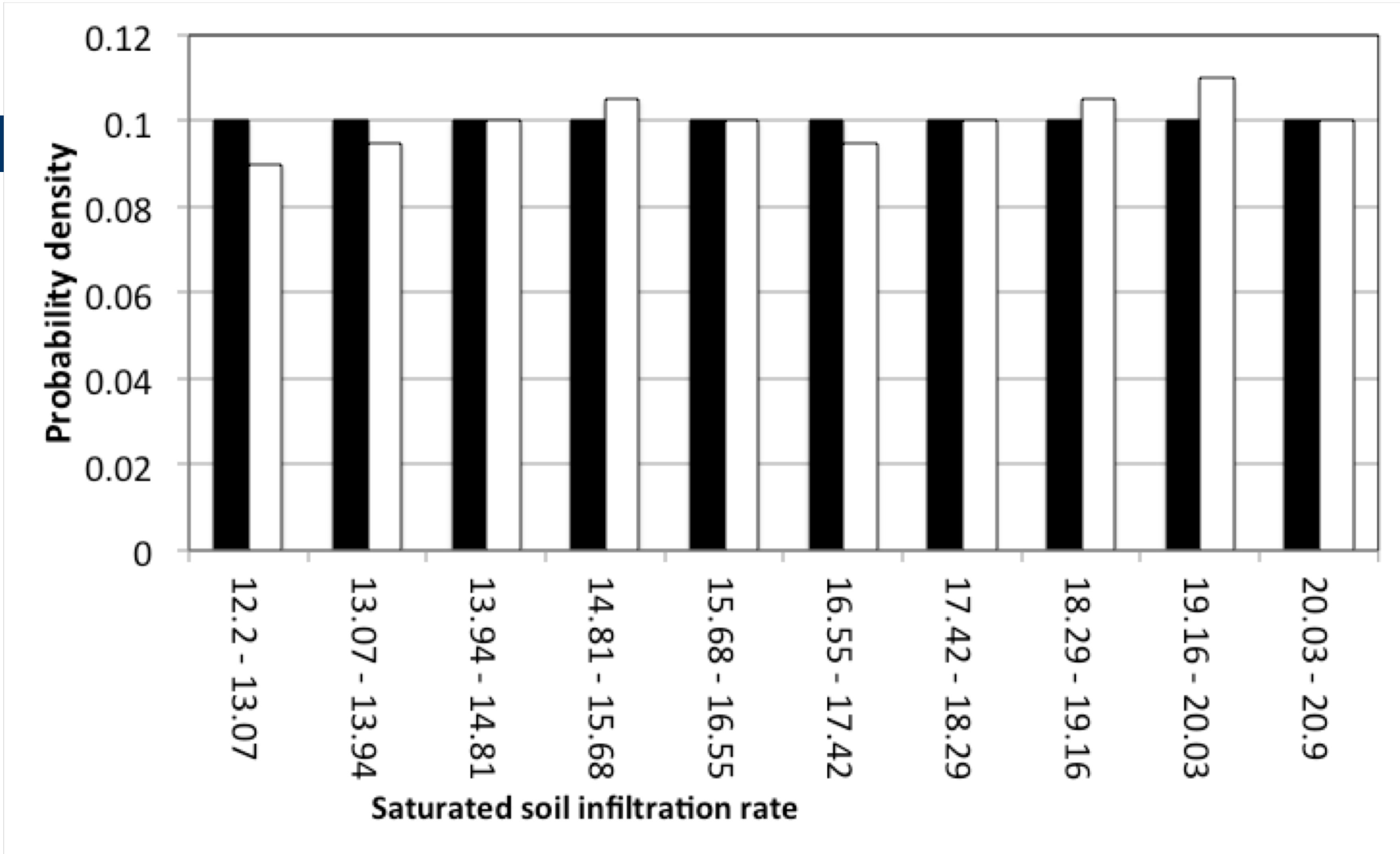
The results of the first Bayesian update



The results of the first Bayesian update

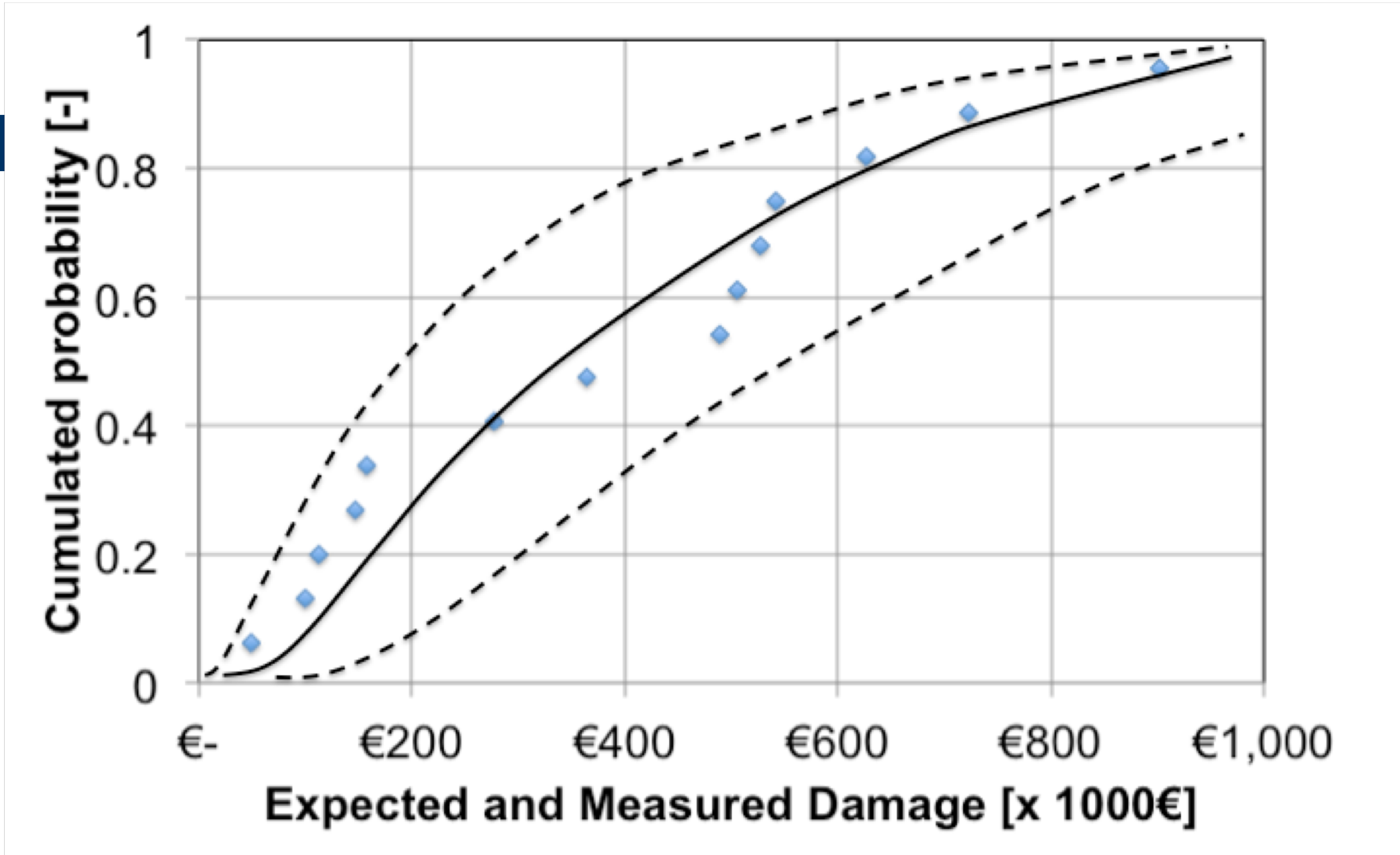


The results of the first Bayesian update

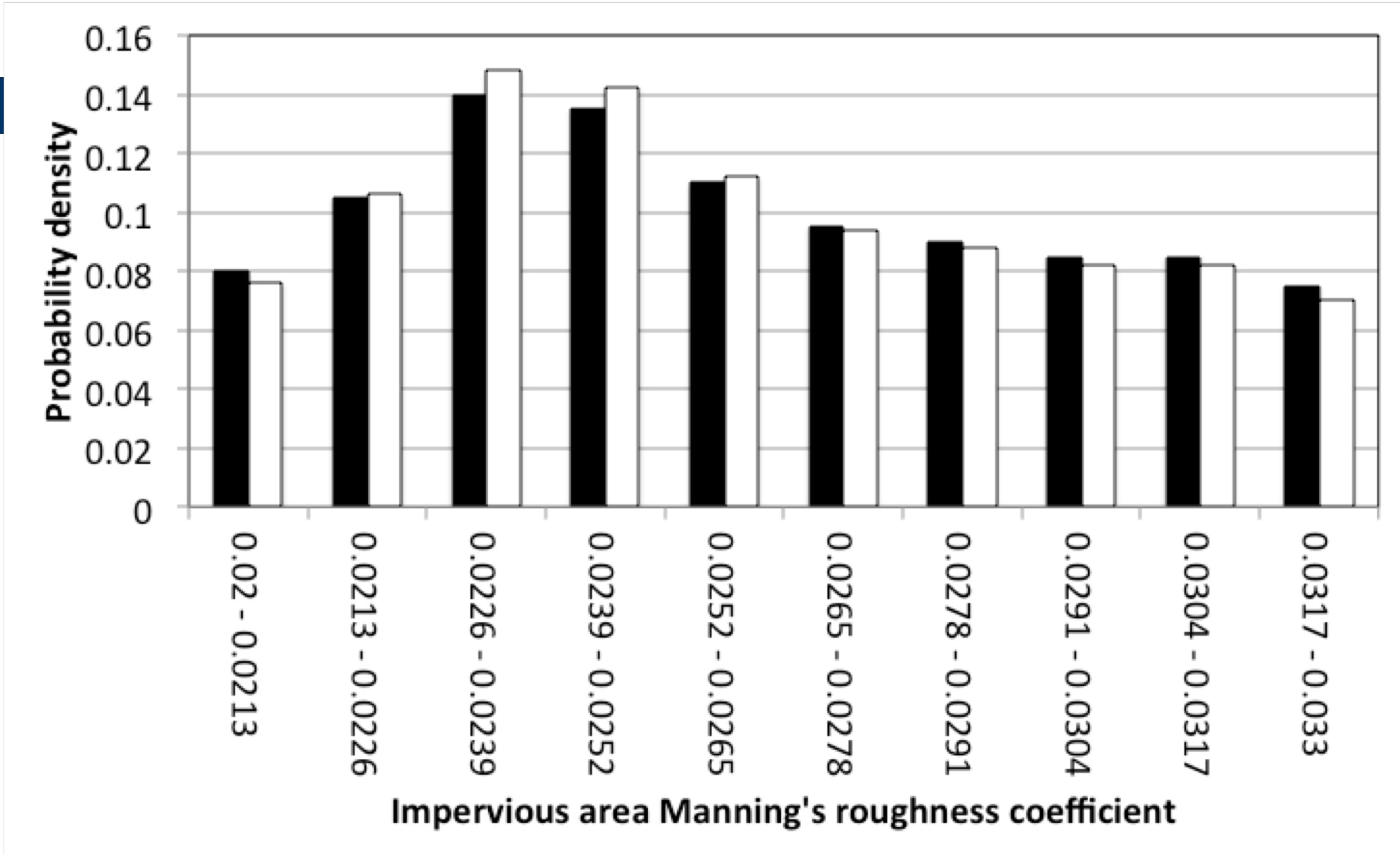


The results of the first Bayesian update

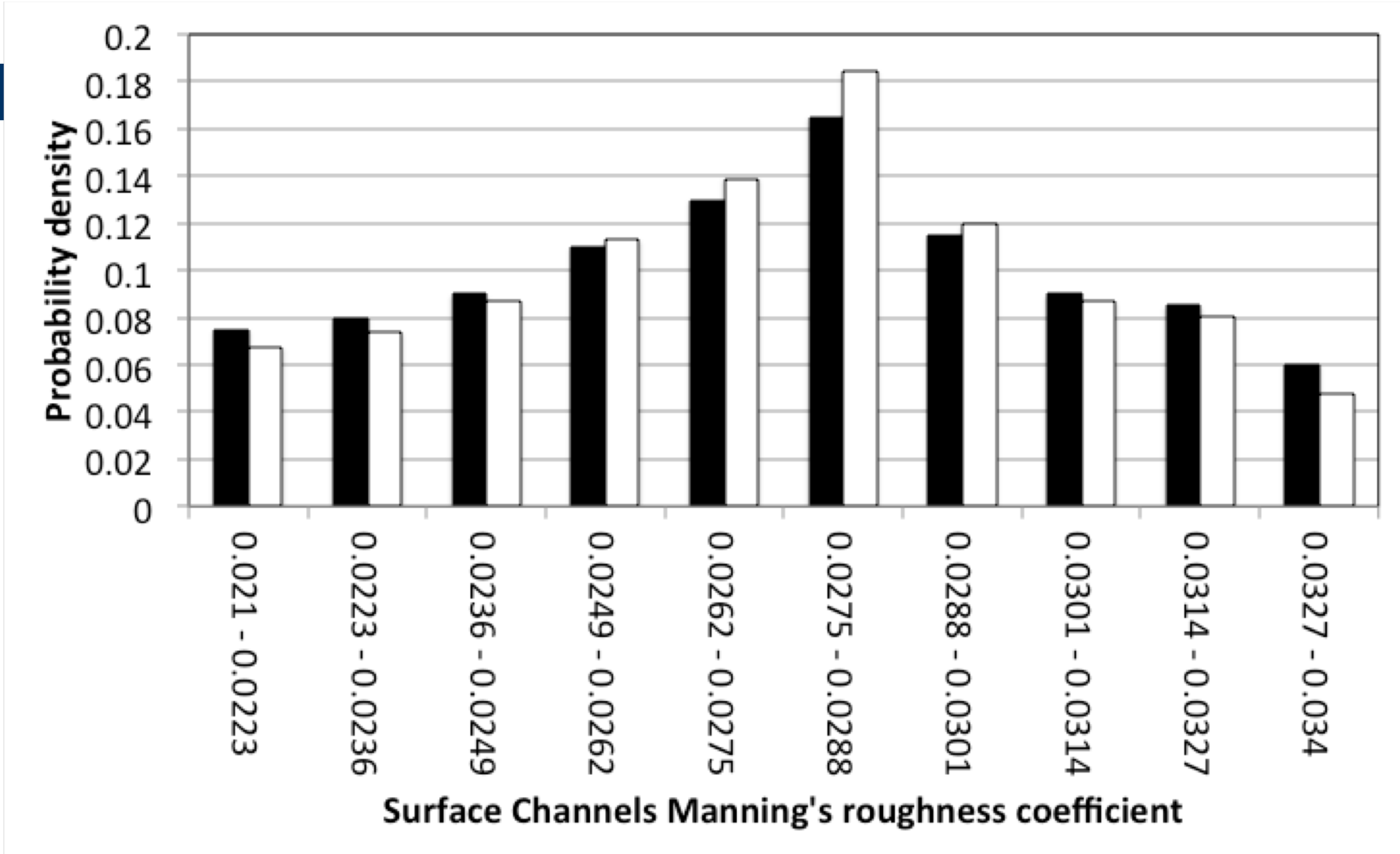
Total flood damage with power damage law



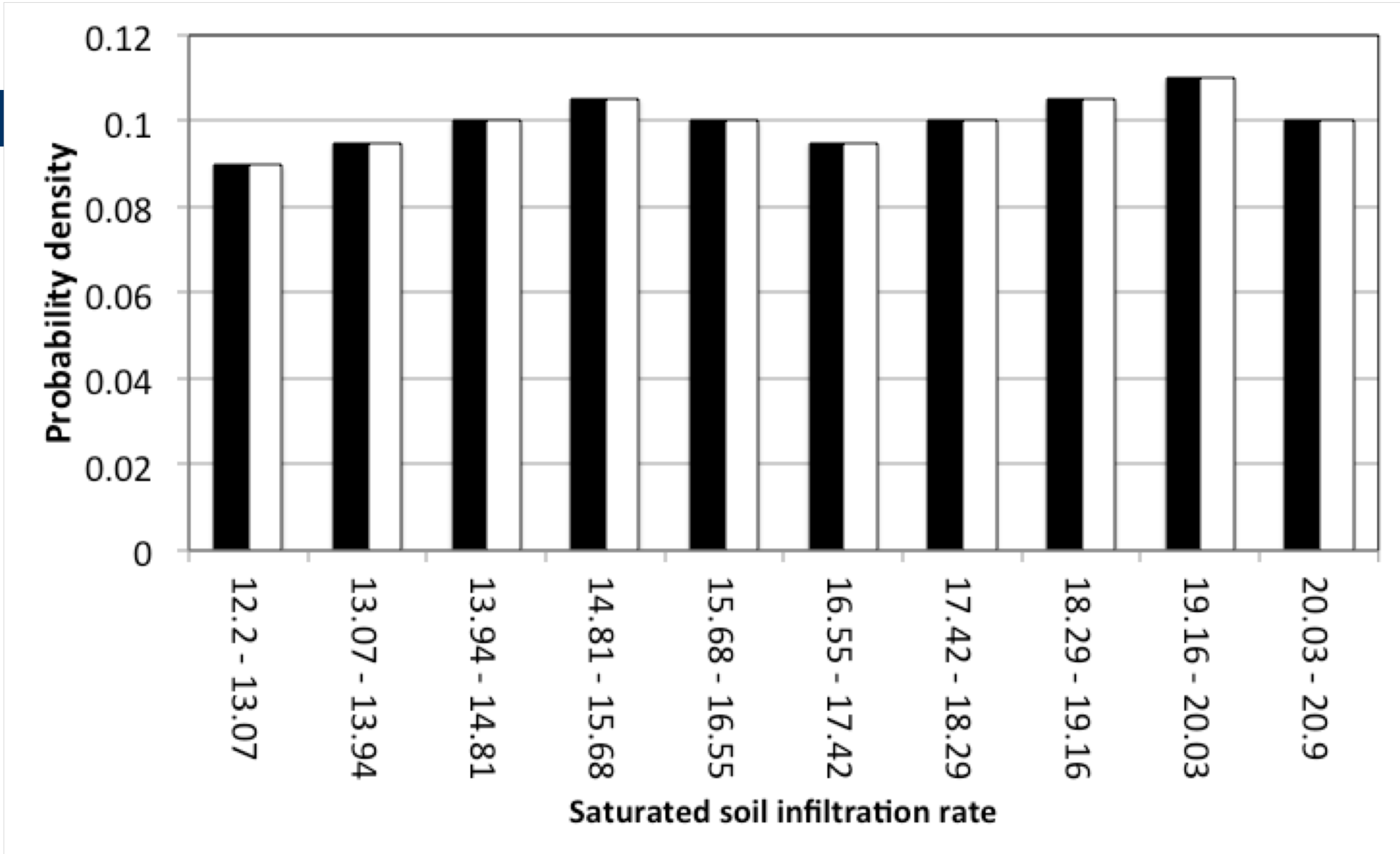
The results of the second Bayesian update



The results of the second Bayesian update

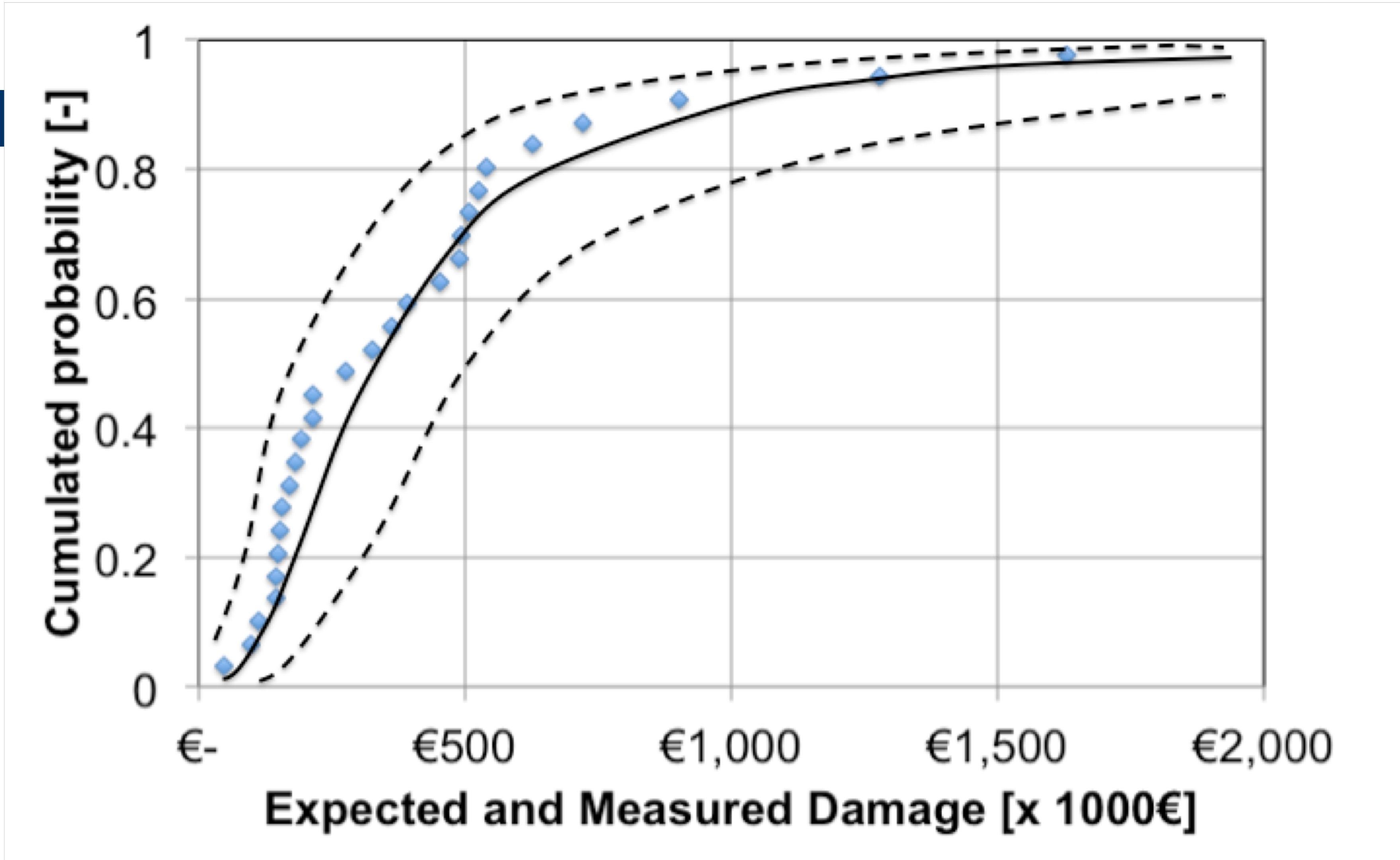


The results of the second Bayesian update



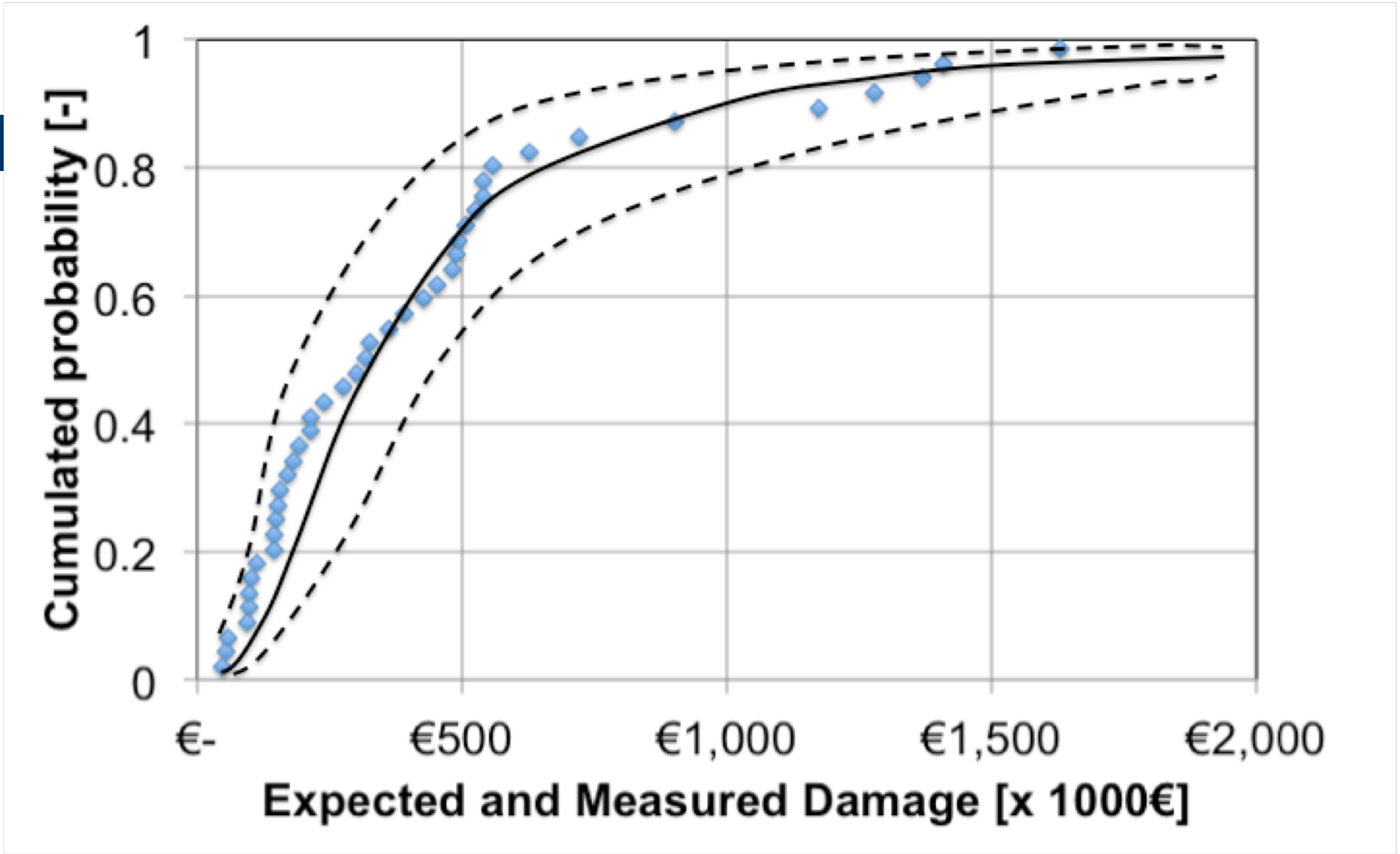
The results of the second Bayesian update

Total flood damage with power damage law



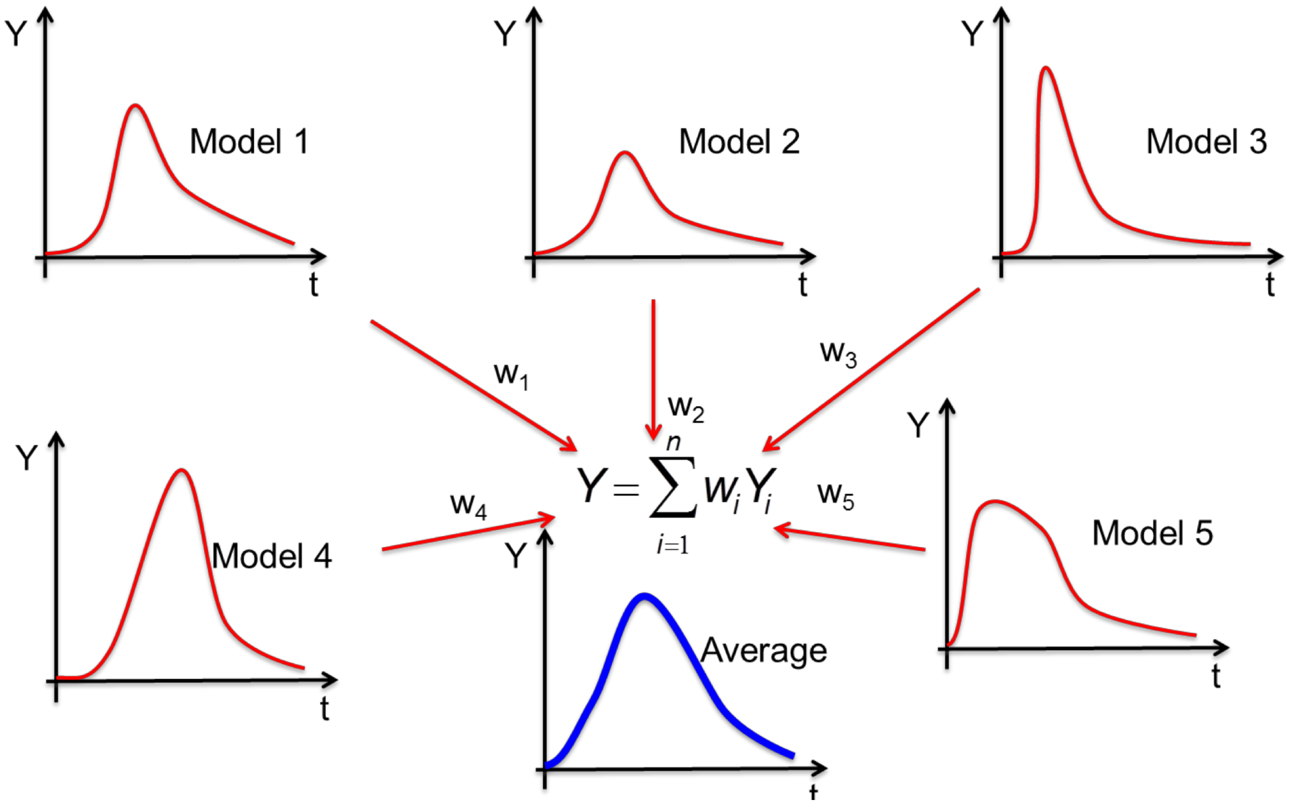
The results of the third Bayesian update

Total flood damage with power damage law



Bayesian Model Averaging (BMA)

Bayesian Model-Averaging technique looks to overcome the limitations of a single model by linearly combining a number of competing models into a single new model forecast



Model structure disinformation

BMA infers the posterior distribution of forecasting variables by weighing individual posterior distributions based on their probabilistic likelihood measures, with the better performing predictions receiving higher weights than the worse predictions.

$\mathcal{M} = (M_1,\ldots,M_K)$ set of models considered and

Δ= quantity to predict

$p(\Delta|\boldsymbol{D}) = \sum_{k=1}^{K} p(\Delta|M_k,\boldsymbol{D})\, p(M_k|\boldsymbol{D})$ **posterior** distribution of Δ given data ***D***

$p(\Delta|M_k,\boldsymbol{D})$ **posterior** distribution under M_k

$p(M_k|\boldsymbol{D})$ **posterior** model probability (**weight**)

All probabilities implicitly **conditional** on $\mathcal{M}$

Model structure disinformation

Because BMA provides a probabilistic form of synthesis results among several models, the mean value of distribution can be used as the result of multi-model forecasting.

BMA prediction (posterior mean) of Δ:

$$E\left[\Delta \middle| \boldsymbol{D}\right] = \sum_{k=1}^{K} E\left[\Delta \middle| \boldsymbol{D}, M_k\right] p\left(M_k \middle| \boldsymbol{D}\right)$$

The expected BMA prediction is the average of individual model predictions weighted by the likelihood w_k that the individual model M_k is the optimal model on the condition of the given data D

Model structure disinformation

BMA predictive uncertainty (posterior variance):

$$Var\left[\Delta|\boldsymbol{D}\right]=\underbrace{\sum_{k=1}^{K}Var\left[\Delta|\boldsymbol{D},M_k\right]p\left(M_k|\boldsymbol{D}\right)}_{\text{within-model variance}}$$

Reflecting data variability

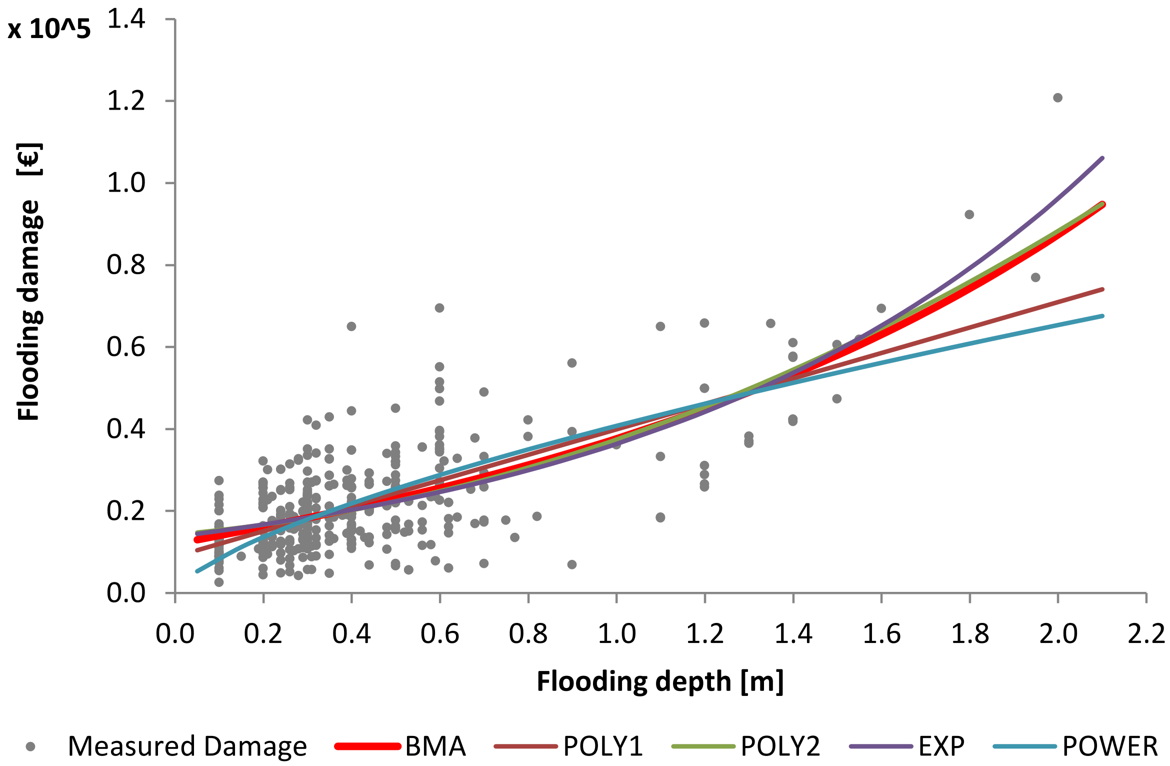
$$+\underbrace{\sum_{k=1}^{K}\left(E\left[\Delta|\boldsymbol{D},M_k\right]-E\left[\Delta|\boldsymbol{D}\right]\right)^2 p\left(M_k|\boldsymbol{D}\right)}_{\text{between-model variance}}$$

Reflecting model consistency

It represents an important uncertainty measure that better describes the predictive uncertainty than in a non-BMA scheme where uncertainty is estimated based only on the variance between-models and consequently results in under-dispersive predictions

Equivalent damage models

BMA prediction and the related uncertainty was compared to the predictions of 4 different formulations of damage curve functions:



- linear (POLY1)
- polynomial-2ord (POLY2)
- exponential (EXP)
- power with upper limit (POWER)

Università di Enna «Kore»
Centro di ricerca La.R.A.

Model structure disinformation

BMA methodology requires the estimation of BMA weights and variance

$$p(\Delta|\boldsymbol{D}) = \sum_{k=1}^{K} p(\Delta|M_k, \boldsymbol{D})\, p(M_k|\boldsymbol{D})$$

Each model was treated by Bayesian uncertainty analysis:

$$p(D|M_k) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma_{e,k}^2}} \exp^{\left(\frac{(D_i - y_i^k)^2}{-2\sigma_{e,k}^2}\right)} \Rightarrow w_k = p(M_k|D) = \frac{p(D|M_k)\cdot p(M_k)}{\sum_{n=1}^{K} p(D|M_n)\cdot p(M_n)}$$

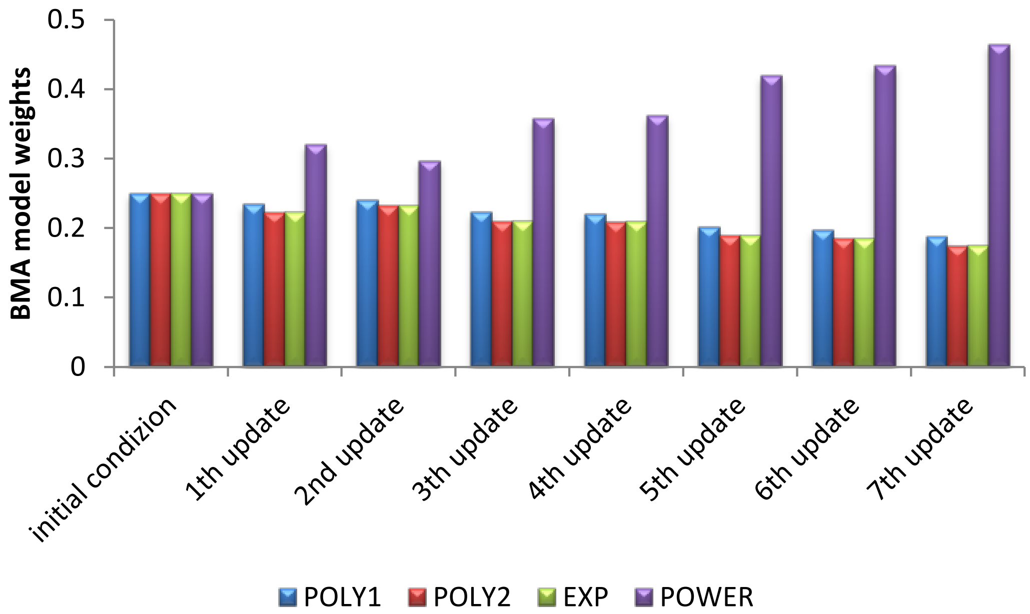
BMA weights and variance results by the application of the Expectation-Maximization (EM) algorithm to a log-likelihood function

$$l(\theta) = \log\left(\sum_{k=1}^{K} w_k \cdot p(\Delta|M_k, D)\right)$$

Model structure disinformation

During the analysis, the effect of the available data growth was taken also into account on the model uncertainty by seven updating of the analysed database (adding 4 events at time)

According to a Bayesian updating approach the model weights, w_k, resulted as a weighted average of its current forecast performance weighted by the conditional probabilities of the previous step.

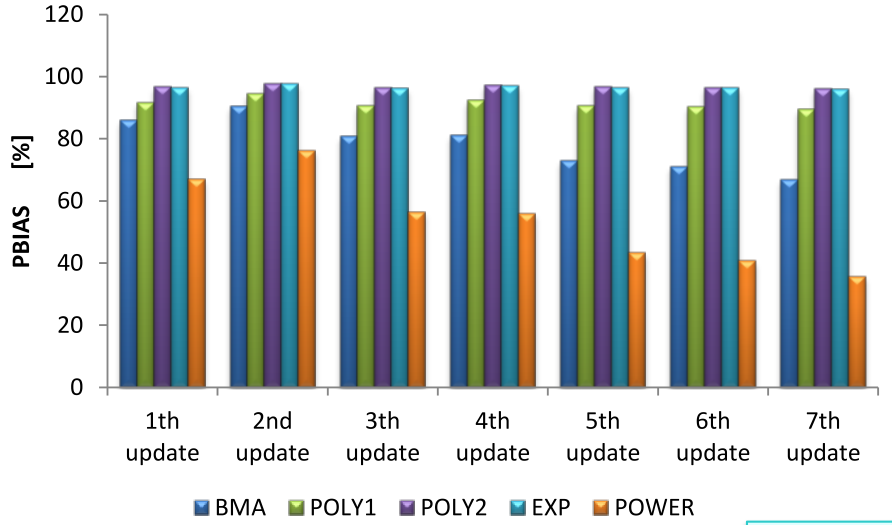


Starting from equal weights for all models, the weight increases for the better performing model and decreases for worse models

Model weight linked to Power function shows always the best performances

Università di Enna «Kore»
Centro di ricerca La.R.A.

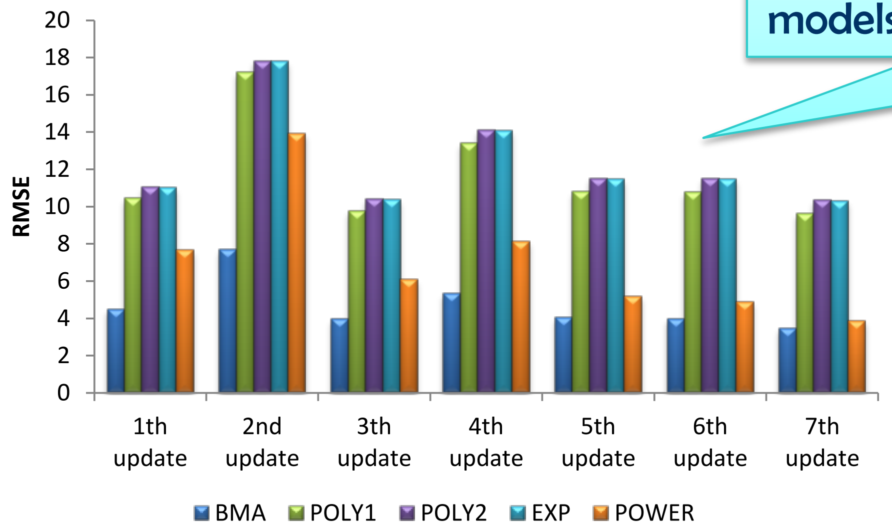
Model structure disinformation



$$PBIAS[y_{BMA} \mid D] = 100 * \frac{\left(D_i - \sum_{k=1}^{K} w_k \eta_{i,k}\right)}{\sum_{i=1}^{m} D_i}$$

$$PBIAS = 100 * \left(\frac{\sum_{i=1}^{m} \left(D_i - y_i^k\right)}{\sum_{i=1}^{m} D_i} \right)$$

RMSE$_{BMA}$ better describes the predictive uncertainty than in a non-BMA scheme where uncertainty is estimated based only on the variance between-models

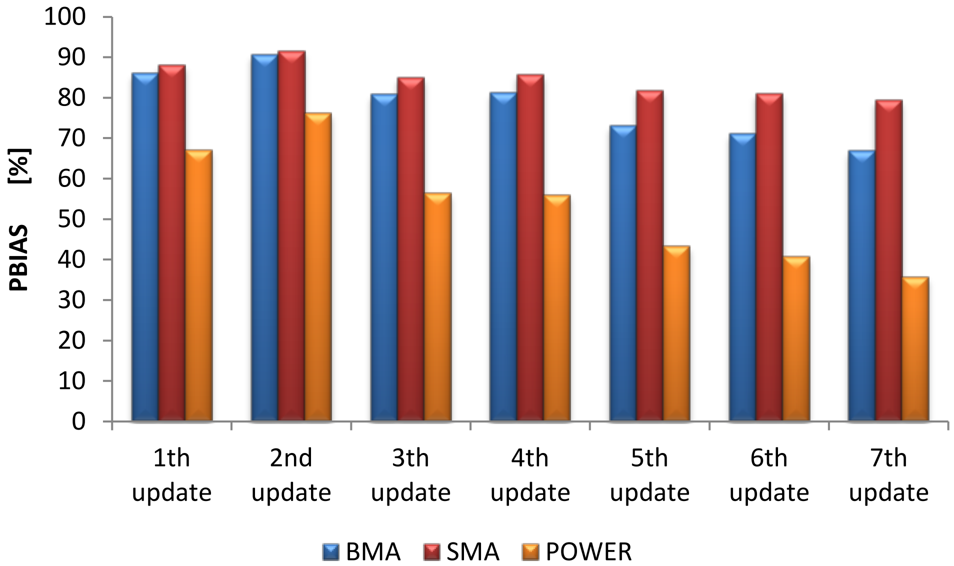


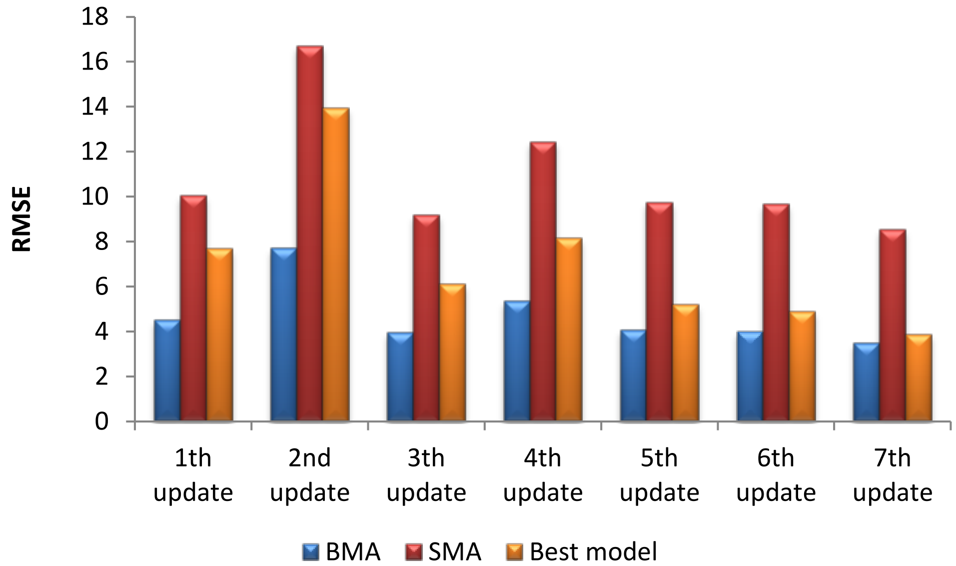
$$RMSE[y_{BMA} \mid D] = \sqrt{\sum_{k=1}^{K} w_k \cdot \left(\eta_k - \sum_{k=1}^{K} w_k \eta_k\right)^2 + \sum_{k=1}^{K} w_k \sigma_k^2}$$

$$RMSE_k = \sqrt{\sum_{i=1}^{m} \frac{\left(D_i - y_i^k\right)^2}{m-1}}$$

Model structure disinformation

Comparison between PBIAS and RSME statistics of the expected BMA predictions, along with that of the Simple Model Average (SMA)





BMA predictions are better than that of the SMA and of best individual predictions in terms of RSME

BMA PBIAS and RMSE are less than SMA ones

The improvement in terms of uncertainty reduction by BMA is around 10% with respect to the best individual model

A paradox to think about in your own work…

- Generally, the more physical understanding that is built into a model, the more parameter values must be specified to run the model
- The more parameter values that cannot be estimated precisely, the more degrees of freedom that will be available in fitting the observations (we cannot measure effective parameters everywhere).
- Therefore the more physical understanding that is built into a model, the greater the uncertainty is likely to be.
- A "perfect" model with unknown parameters is no protection against equifinality

Summary

- Maybe it is impossible to separate out sources of error from series of model residuals

but probably we do not need it in practice

- Epistemic sources of uncertainty result in non-stationarity in error characteristics;

UA hypotheses are a card tower so watch out

- Treating all uncertainties as aleatory can lead to dramatic over-conditioning

and over-estimation of uncertainty

- Non formal approaches allows for having a glance on the performance of our models without requiring restrictive assumptions

but produce equifinality and uncertainty estimation is only qualitative

- Limits of acceptability on model performance can help getting rid of bad models;

Every time your modeler heart is broken, a doorway cracks open to a world full of new beginnings, new opportunities

- Discussion and agreement regarding assumptions of analysis provide a basis for communication of concepts

If it happens early and often, all the better