From Ensemble to Tree

An information based approach

L. Raso, D. Schwanenberg, P.J. van Overloop, N. van de Giesen

HEPEX International workshop on Post-processing and Verification of hydrological ensemble predictions

7-9 June 2011, UNESCO-IHE, Delft, the Netherlands

Brief Personal Presentation

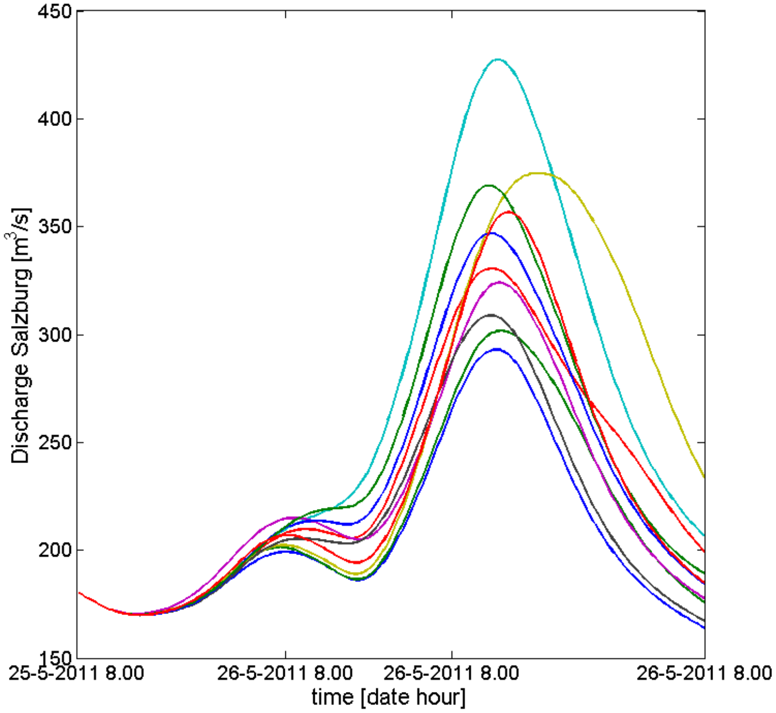
Current Research area:

- Optimal anticipatory control of water systems under uncertainty
  - Forecast uncertainty

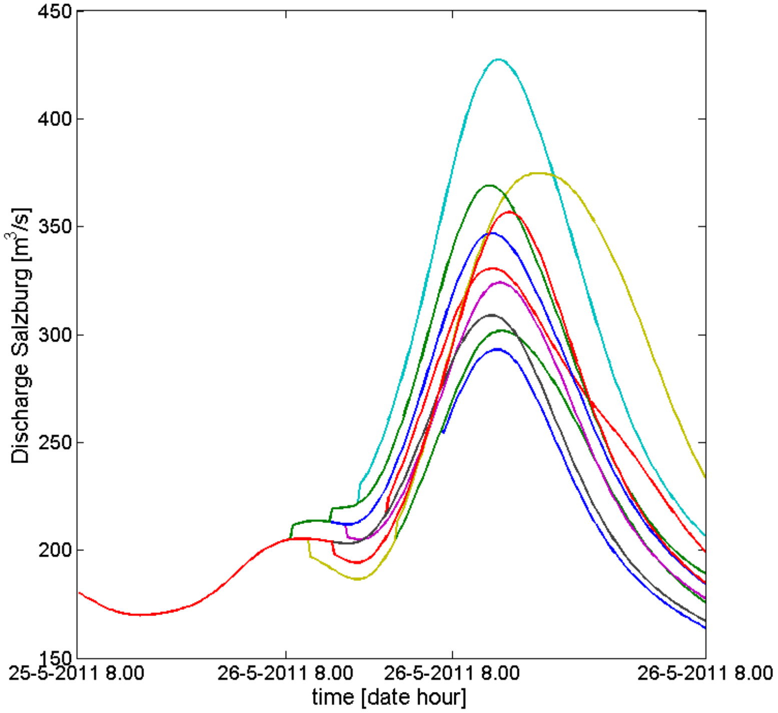
What is a tree?

Difference between a tree and an ensemble

- Tree and Ensemble are two different models of uncertainty



ensemble

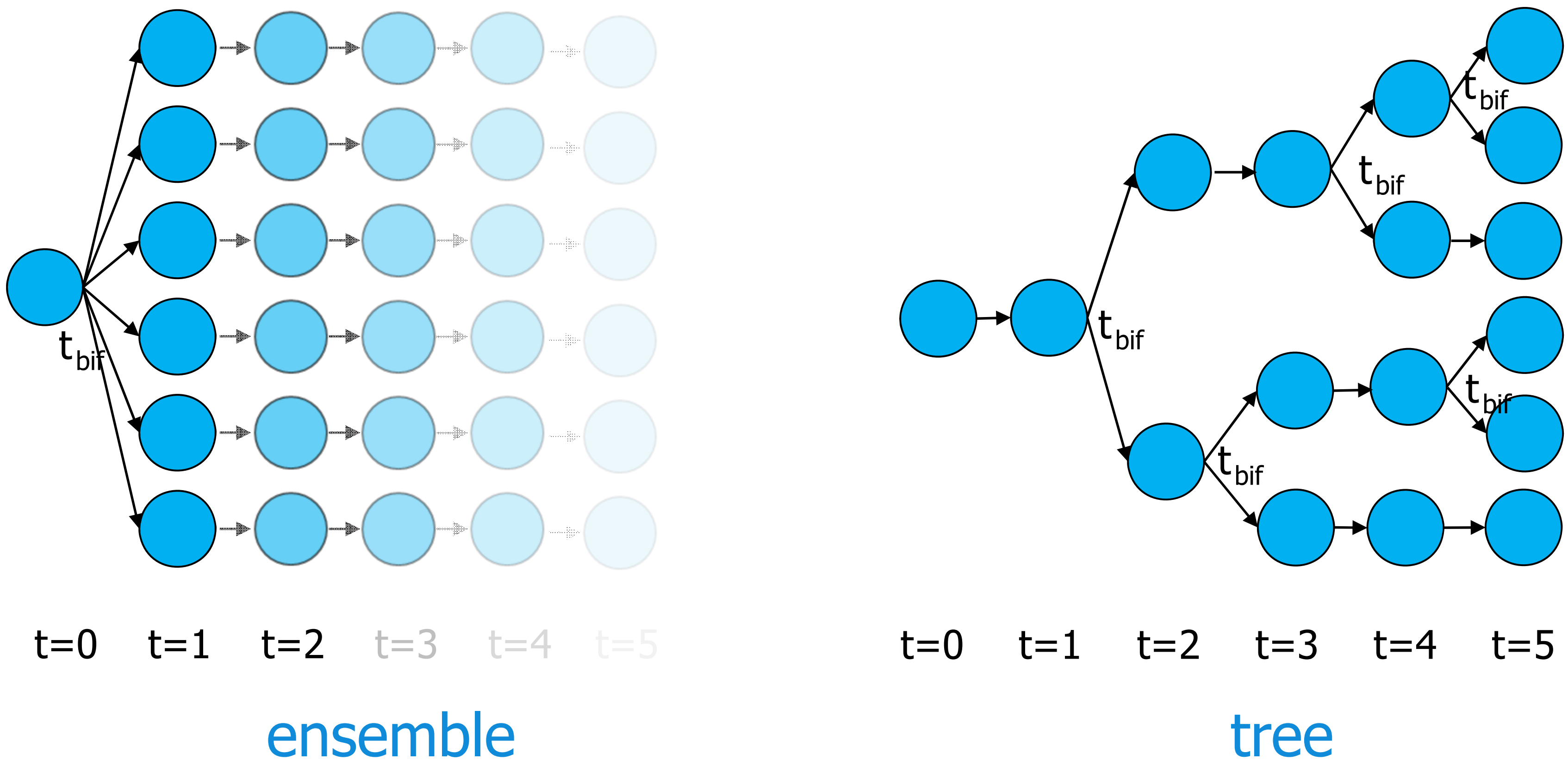


tree

What is a tree?

Difference between a tree and an ensemble

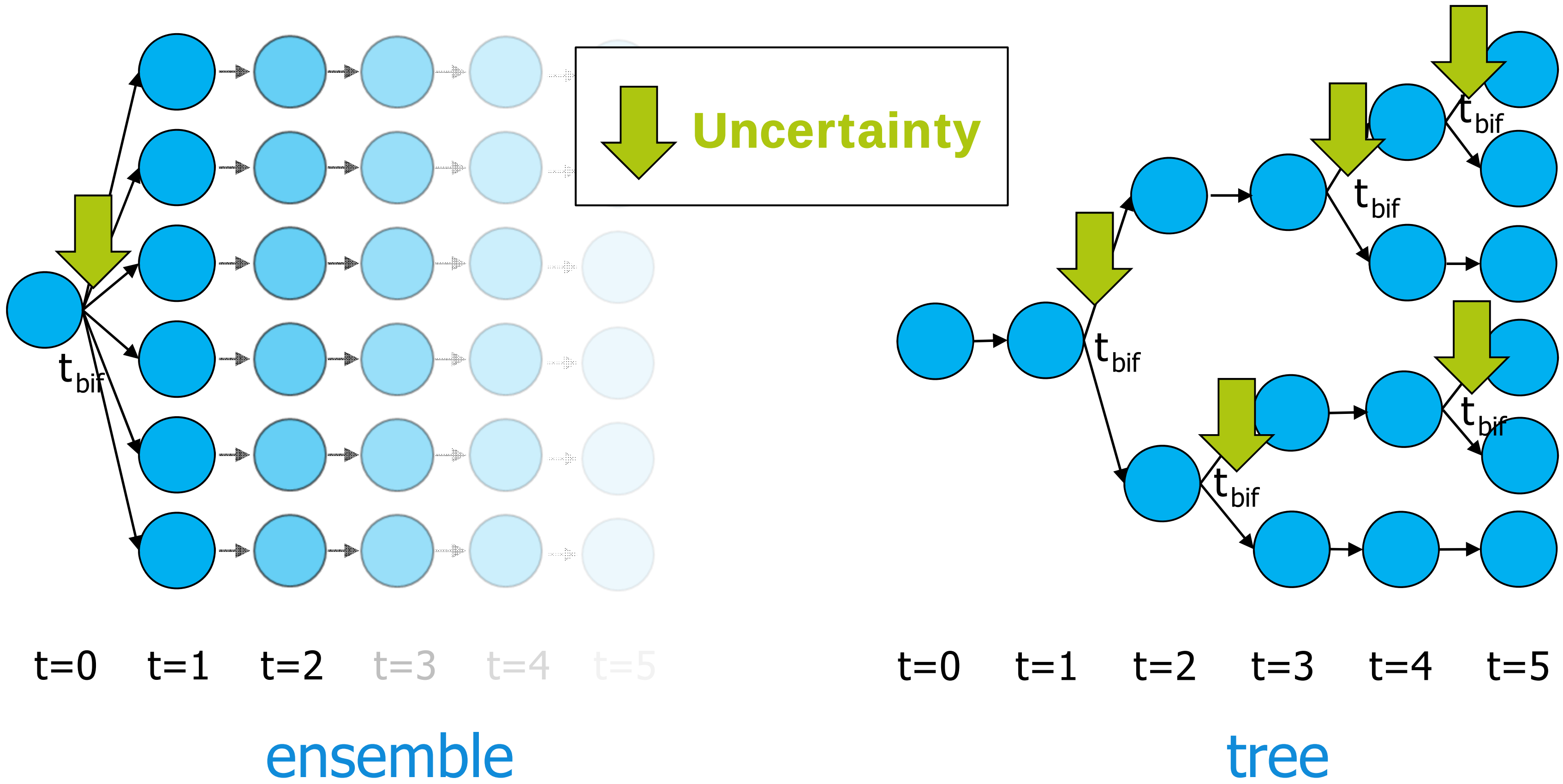
- Tree and Ensemble are two different models of uncertainty



What is a tree?

Difference between a tree and an ensemble

- Tree and Ensemble are two different models of uncertainty



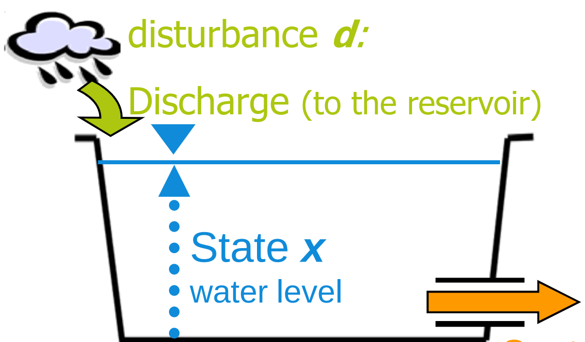
Why using a tree

The “two-stage” optimization problem

Two interdependent decisions, under uncertainty, in this sequence:

Decision | Uncertainty | Decision

Example: control of a reservoir



Release at t_1	Precipitation event	Release at t_2
u_1	D:[0,1]; P($d = 1$) $= p$	u_2

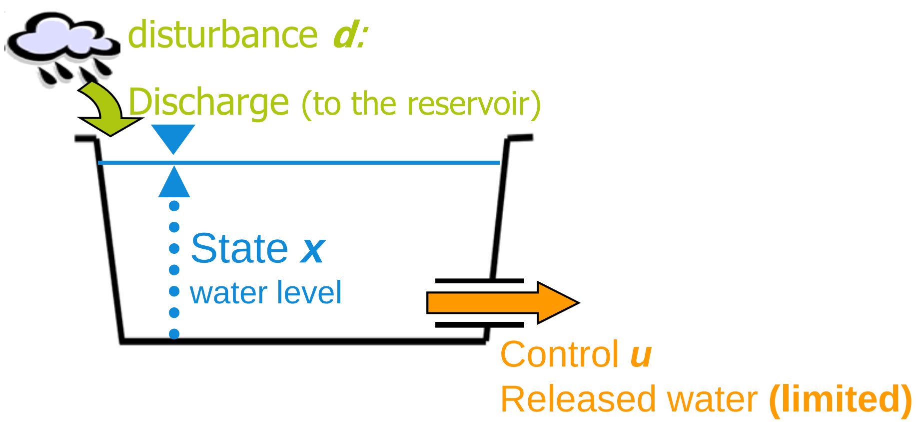
Why using a tree

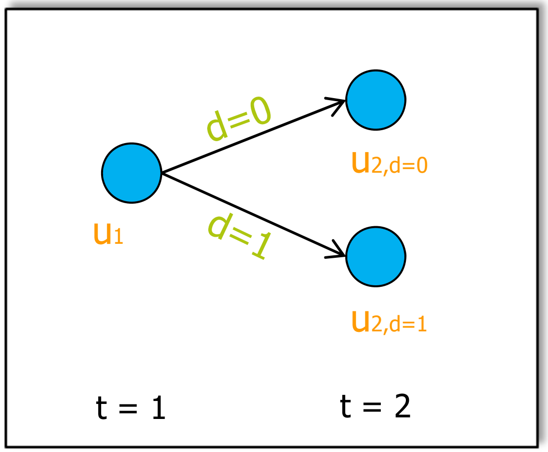
The "two-stage" optimization problem

Two interdependent decisions, under uncertainty, in this sequence:

Decision | Uncertainty | Decision

Example: control of a reservoir





Release at t_1
u_1

Precipitation event
D:[0,1]; $P(d=1) = p$

Release at t_2
u_2

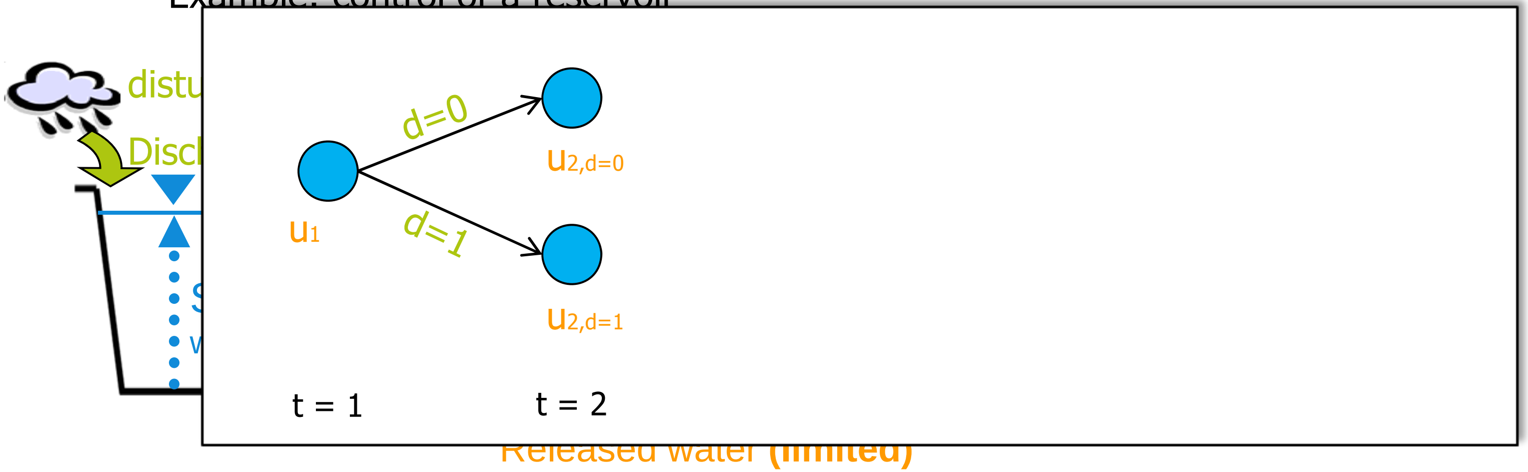
Why using a tree

The “two-stage” optimization problem

Two interdependent decisions, under uncertainty, in this sequence:

Decision | Uncertainty | Decision

Example: control of a reservoir



Release at t_1 | Precipitation event | Release at t_2

u_1 | D:[0,1]; P($d = 1$) $= p$ | u_2

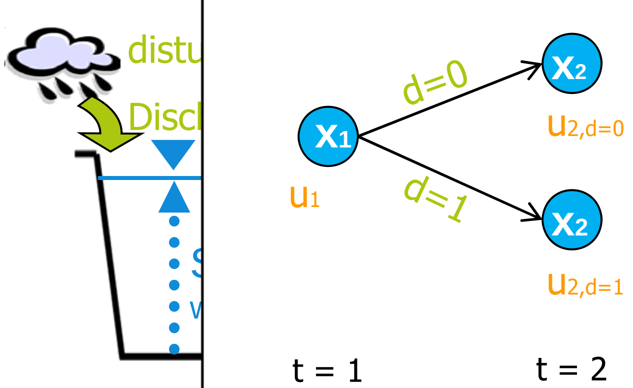
Why using a tree

The “two-stage” optimization problem

Two interdependent decisions, under uncertainty, in this sequence:

Decision	Uncertainty	Decision
Release at t_1	Precipitation event	Release at t_2
u_1	D:[0,1]; P($d = 1$) $= p$	u_2

Example: control of a reservoir



Decision variables:

$$\left\lfloor u_1, u_{2|d=0}, u_{2|d=1} \right\rfloor$$

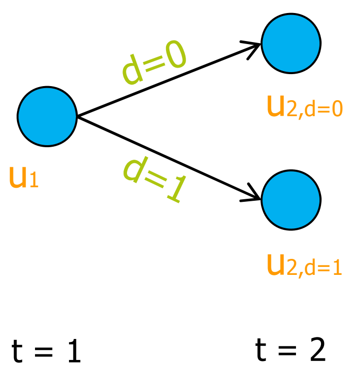
Why using a tree

The “two-stage” optimization problem

Two interdependent decisions, under uncertainty, in this sequence:

Decision — Uncertainty — Decision

Example: control of a reservoir



Released water **(limited)**

Release at t_1	Precipitation event	Release at t_2
u_1	D:[0,1]; $P(d=1) = p$	u_2

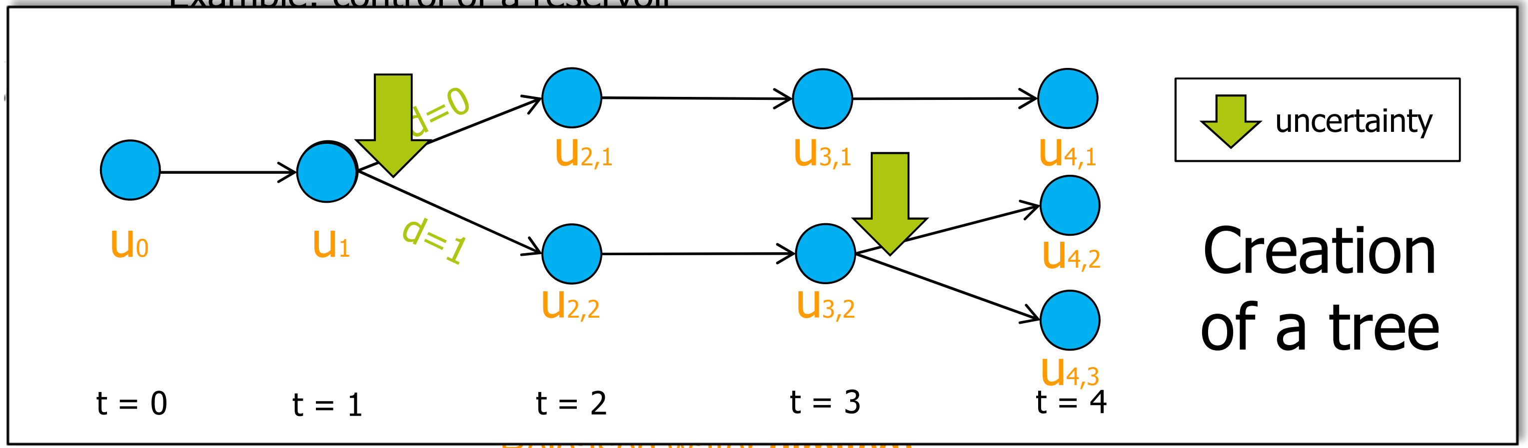
Why using a tree

The “two-stage” optimization problem

Two interdependent decisions, under uncertainty, in this sequence:

Decision | Uncertainty | Decision

Example: control of a reservoir



Released water **(limited)**

Release at t_1
u_1

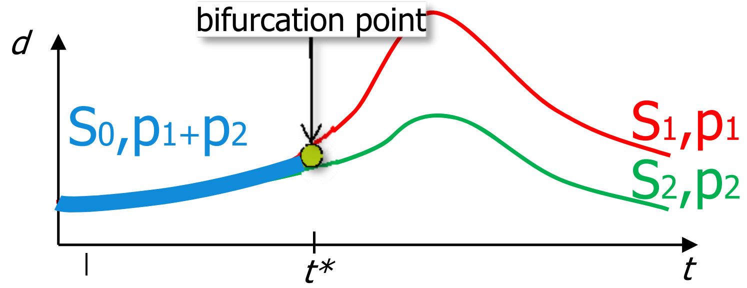
Precipitation event
D:[0,1]; P($d = 1$) = p

Release at t_2
u_2

How to create a tree from an ensemble

Existing procedures for scenario aggregation

- Existing procedure: "*optimal scenario reduction*" technique
  - Aggregate scenarios when "sufficiently close", i.e. average distance on [0,t*] smaller than a fixed threshold



This method requires the definition of

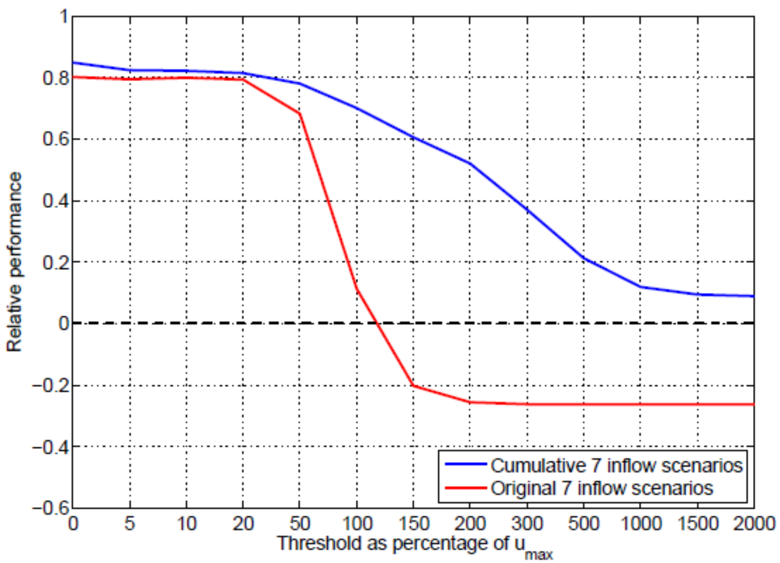
- Distance
- Threshold

How to create a tree from an ensemble

Existing procedures for scenario aggregation

- Existing procedure: "*scenario reduction technique*"
  - Aggregate scenarios when "sufficiently close", i.e. average distance on $[0,t^*]$ smaller than a fixed threshold

Threshold value sensitivity analysis



Optimal control of the reservoir using a tree

Relative performance index:
1: perfect forecast, no uncertainty
0: deterministic forecast

Importance of placing the bifurcation points a the right moment

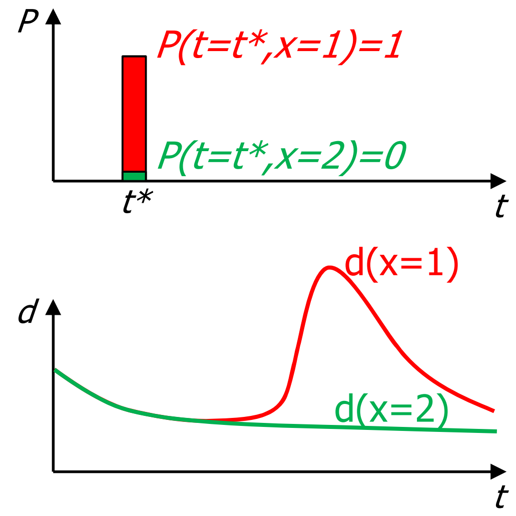
- Threshold

How to create a tree from an ensemble

An Information Based approach

- Bifurcation points are to be placed as soon as the uncertainty is solved, i.e. where probability of occurrence is sufficiently close to one

- Example: stream response with delay
    - Two scenarios:
        - Scenario x=1: rain at *t**
        - Scenario x=2: no rain
    - Observable dimensions:
        - Precipitation *P*
        - Discharge *d*



Rule: find $\min_t p(x) > p^*$

How does information enter the system?

Update of scenarios probability

- Observable variables $y(t, x, j)$
  - t = time *[1,…,h]*
  - x = scenario *x ϵ X:{1,2,…,N}*
  - j = observable dimension *discharge, rain, etc,…*

$y_t \mid x$ is a random variable f_y distributed

$$y_t = \hat{y}_t + \varepsilon \qquad \varepsilon \sim N\left(0, \sigma^2\right)$$

- Use of the Bayesian theorem:

$$p(x|\hat{y}_t) = \frac{f_y(\hat{y}_t|x) \cdot p(x)}{\sum_i^N f_y(\hat{y}_t|x_i) \cdot p(x_i)}$$

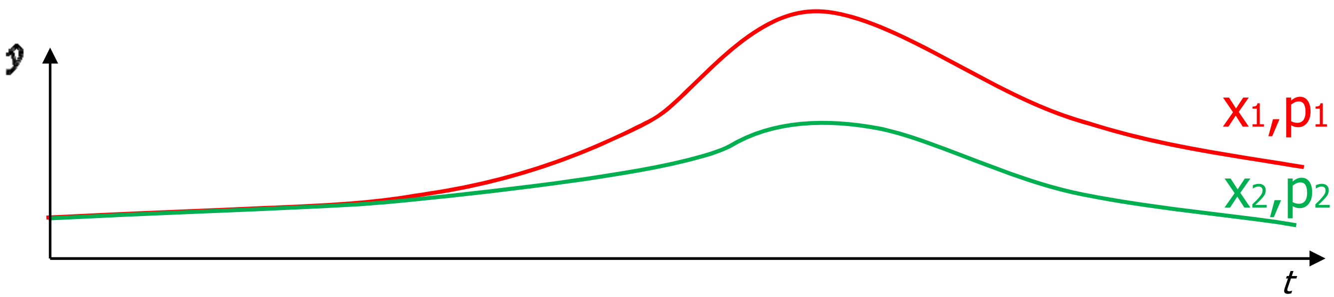
Update of the probability $p(x)$ given the observation $\hat{y}_t$

How does information enter the system?

Update of scenarios probability

- Observable variables $y(t, x, i)$

- Two scenarios $[x_1, x_2]$ with *a priori* probability $[p_1, p_2]$
- Observable dimension: $\hat{y}$ (present discharge)



- Use of the Bayesian theorem:

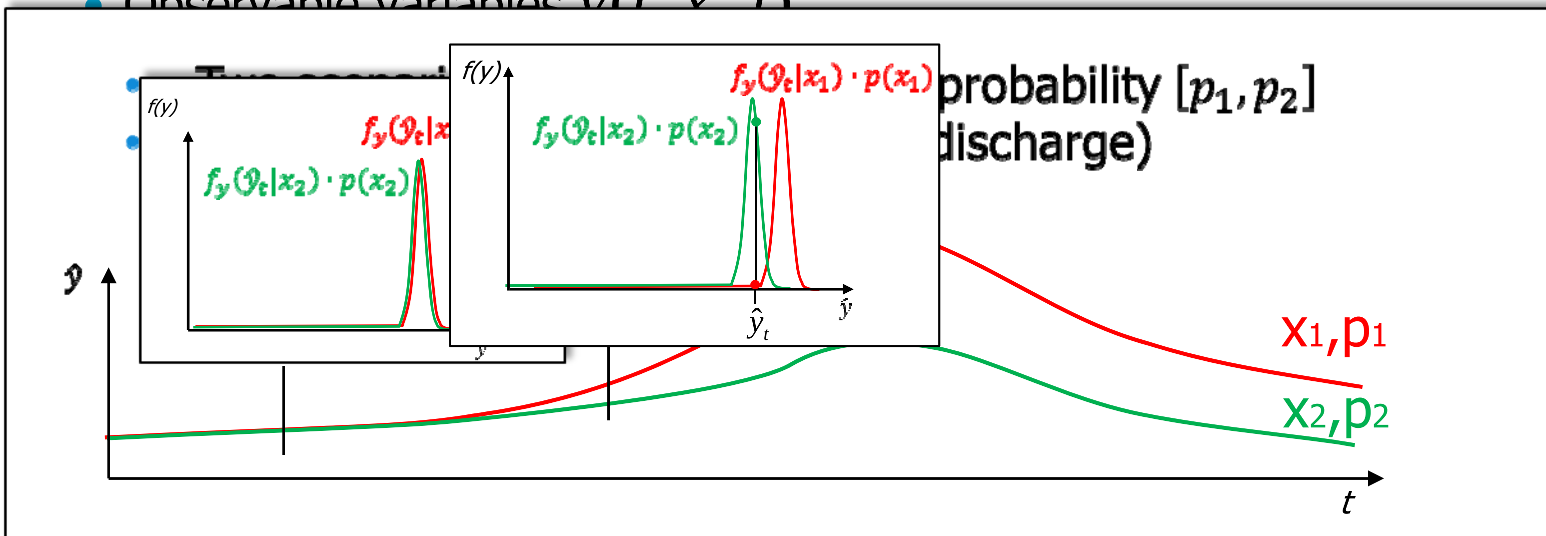
$$p(x|\hat{y}_t) = \frac{f_y(\hat{y}_t|x) \cdot p(x)}{\sum_i^N f_y(\hat{y}_t|x_i) \cdot p(x_i)}$$

Update of the probability $p(x)$ given the observation $\hat{y}_t$

How does information enter the system?

Update of scenarios probability

- Observable variables $y(t, x, i)$



- Use of the Bayesian theorem:

$$p(x|\hat{y}_t) = \frac{f_y(\hat{y}_t|x) \cdot p(x)}{\sum_i^N f_y(\hat{y}_t|x_i) \cdot p(x_i)}$$

Update of the probability $p(x)$ given the observation $\hat{y}_t$

How does information enter the system?

Update of scenarios probability

- Observable variables $y(t, x, j)$
  - t = time *[1,…,h]*
  - x = scenario $x \in X:\{1,2,\dots,N\}$
  - j = observable dimension *discharge, rain, etc,…*

$y_t \mid x$ is a random variable f_y distributed

$$y_t = \hat{y}_t + \varepsilon \qquad \varepsilon \sim N\left(0, \sigma^2\right)$$

- Use of the Bayesian theorem:

$$p(x|\hat{y}_t) = \frac{f_y(\hat{y}_t|x) \cdot p(x)}{\sum_i^N f_y(\hat{y}_t|x_i) \cdot p(x_i)}$$

Update of the probability $p(x)$ given the observation $\hat{y}_t$

How does information enter the system?

Update of scenarios probability

- Observable variables $y(t, x, j)$
  - t = time *[1,…,h]*
  - x = scenario *$x \in X:\{1,2,\ldots,N\}$*
  - j = observable dimension *discharge, rain, etc,…*

$y_t \mid x$ is a random variable f_y distributed

$$y_t = \hat{y}_t + \varepsilon \qquad \varepsilon \sim N\left(0, \sigma^2\right)$$

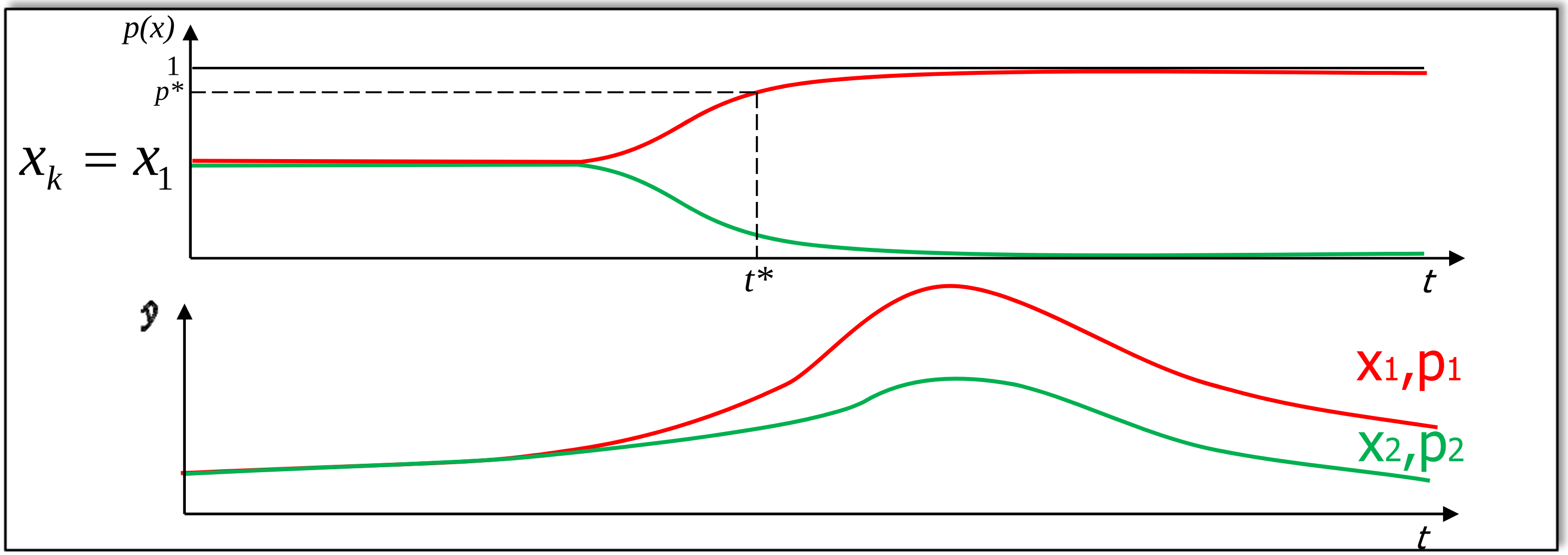
- Use of the Bayesian theorem on average:

$$p(x|y_t) = \int \frac{f_y(y_t|x) \cdot p(x)}{\sum_i^N f_y(y_t|x_i) \cdot p(x_i)} f(y_t|x_k)dx$$

Update of the probability $p(x)$ given the average observation $\hat{y}_t|x_k$

How does information enter the system?

Update of scenarios probability



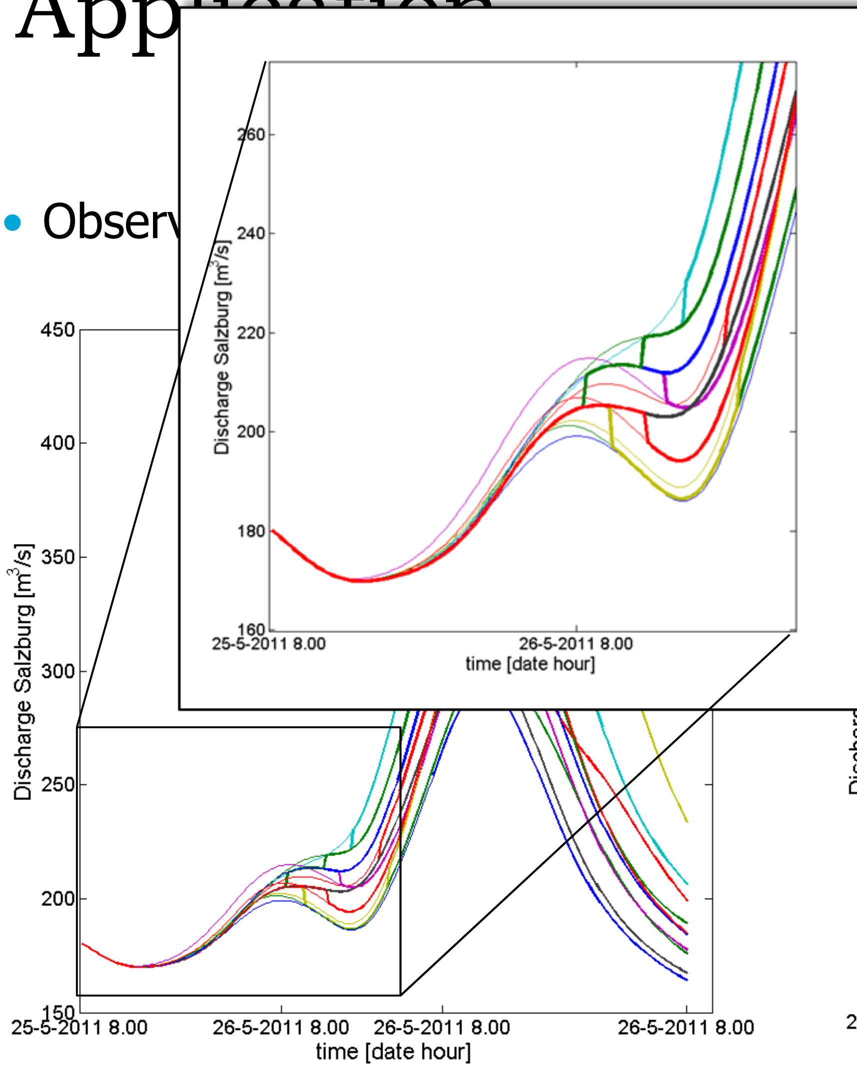
- Use of the Bayesian theorem on average:

$$p(x|y_t) = \int \frac{f_y(y_t|x) \cdot p(x)}{\sum_i^N f_y(y_t|x_i) \cdot p(x_i)} f(y_t|x_k)dx$$

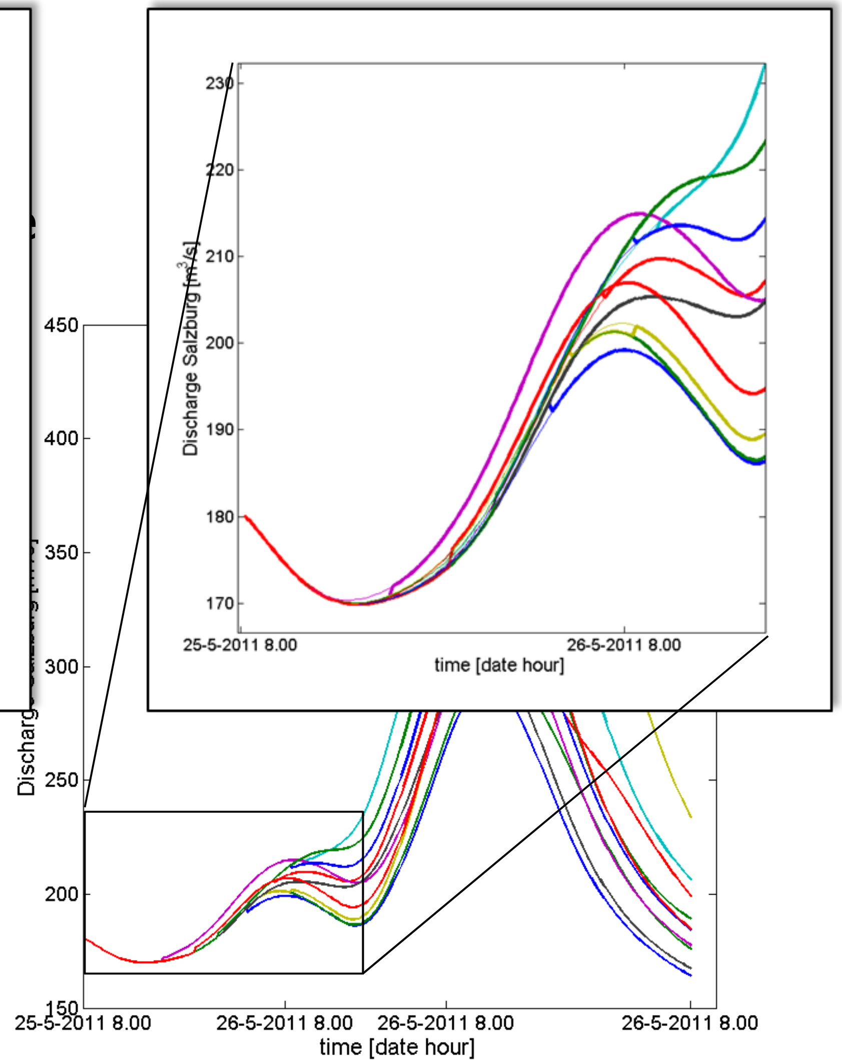
Update of the probability $p(x)$ given the average observation $\hat{y}_t|x_k$

Application

- Observ



$\sigma = 5\ m^3/s$



$\sigma = 1\ m^3/s$

Future application

- Observable variable: discharge forecast

$$Y_t \mid x \sim N(\mu, \Sigma)$$

$$Y_t = \left[y_{t|t}, y_{t+1|t}, \ldots, y_{t+h|t}\right]$$
$$\mu = \left[E\left(y_{t|t} \mid x\right), E\left(y_{t+1|t} \mid x\right), \ldots, E\left(y_{t+h|t} \mid x\right)\right]$$
$$\Sigma = COV\left(y_{t+i|t} \mid x, y_{t+j|t} \mid x\right)_{i=1,2,\ldots,h;\, j=1,2,\ldots,h}$$

Comments and conclusion

- In the cases when uncertainty is solved in time

  A tree is a <u>better model of uncertainty</u> than an ensemble

- It is possible to generate a tree from an ensemble
  - the “Information Based” method requires the definition how information enters the system

Question Time

From Ensemble to Tree

An information based approach

L. Raso, D. Schwanenberg, P.J. van Overloop, N. van de Giesen

HEPEX International workshop on Post-processing and Verification of hydrological ensemble predictions

7-9 June 2011, UNESCO-IHE, Delft, the Netherlands