SimHydro 2019

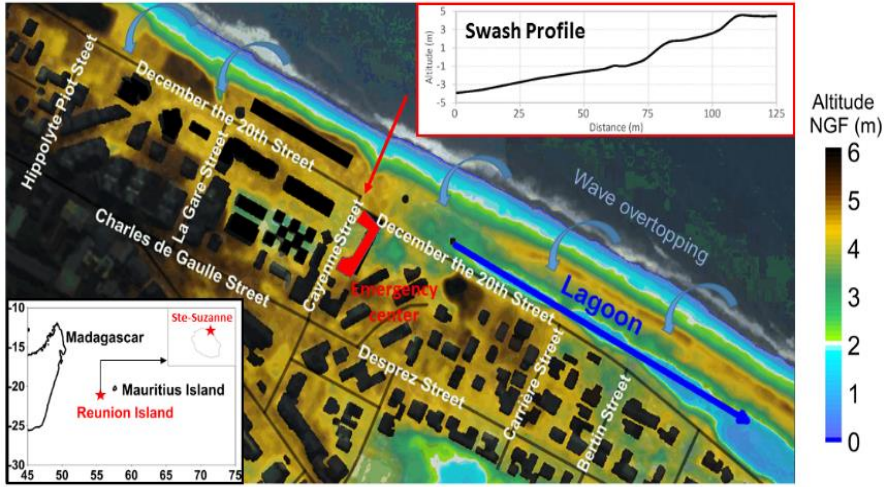
5TH INTERNATIONAL CONFERENCE
12 – 14 JUNE 2019
SOPHIA ANTIPOLIS - FRANCE

Early-warning system (EWS) for cyclone-induced wave overtopping aided by a suite of random forest approaches

J. Rohmer[1*], S. Lecacheux[1], R. Pedreros[1], D. Idier[1], F. Bonnardot[2]
j.rohmer@brgm.fr
1: BRGM
2: Direction Régionale de Météo-France pour l'Océan Indien

EWS at Ste Suzanne

- Objective:
  - **Probability** of marine flooding (wave overtopping) as input of the **Early Warning System**
  - **If there is overtopping**,
    - **Magnitude** of the flooding?
    - **Time evolution** of the flooding?



Problem definition

- Objective:
  - **Probability** of marine flooding (wave overtopping) as input of the **Early Warning System**
  - **If there is overtopping**,
    - **Magnitude** of the flooding?
    - **Time evolution** of the flooding?

- **Problem:**
  - **Computation time cost of numerical simulations** too high to be directly integrated in EWS

Problem definition

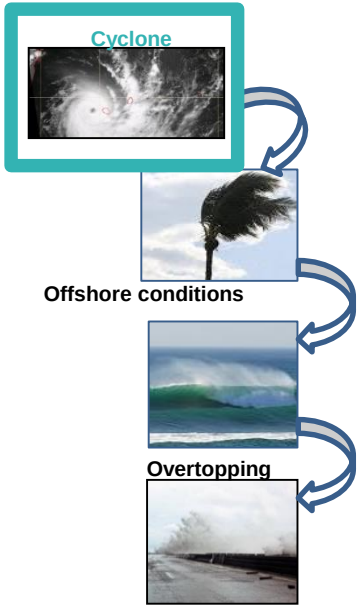
- Objective:
  - **Probability** of marine flooding (wave overtopping) as input of the **Early Warning System**
  - **If there is overtopping**,
    - **Magnitude** of the flooding?
    - **Time evolution** of the flooding?

- **Problem:**
  - Computation time cost of numerical simulations too high to be directly integrated in EWS

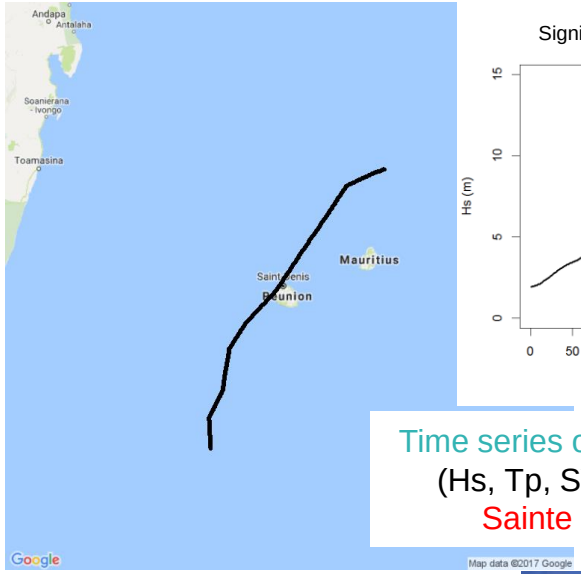
- **What we have:**
  - A database of pre-calculated simulations
  - ~500 synthetic cyclones
  - + their impact in terms of flooding

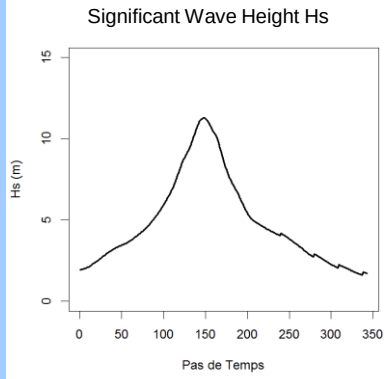
Database of pre-calculated numerical results

Example of synthetic cyclone
(Quetelard and Bonnardot, Meteo France)

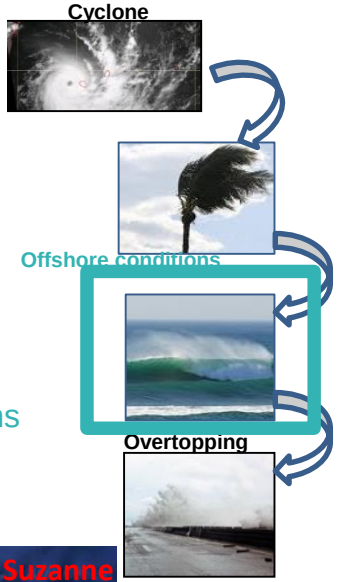


Database of pre-calculated numerical results





Time series of offshore conditions (Hs, Tp, SWL) offshore from Sainte Suzanne coast

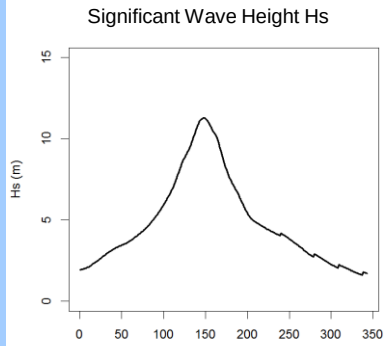


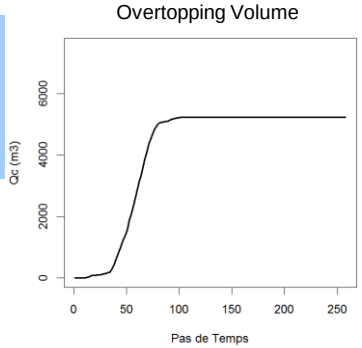
Regional wave Modelling using WWIII

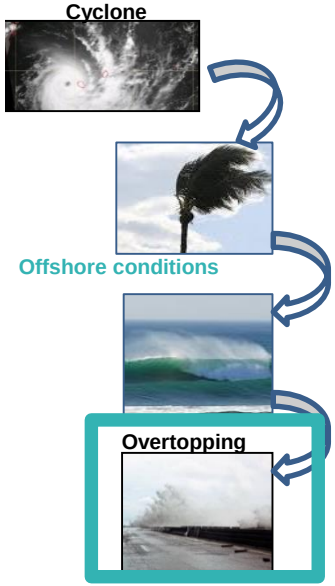
Compatible computation time cost!

Database of pre-calculated numerical results

Time series of cumulative water volume induced by **wave overtopping**

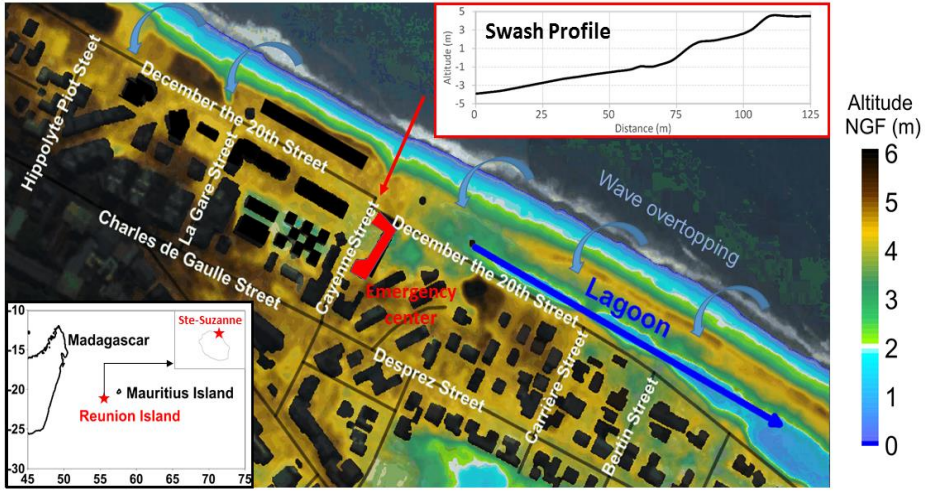






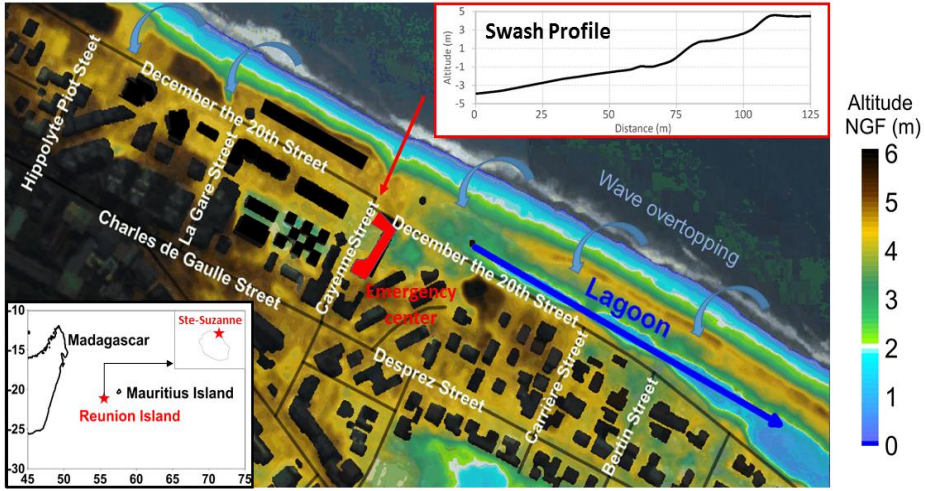
Very high computation time cost (>several hours)!

Wave overtopping modeling at local scale

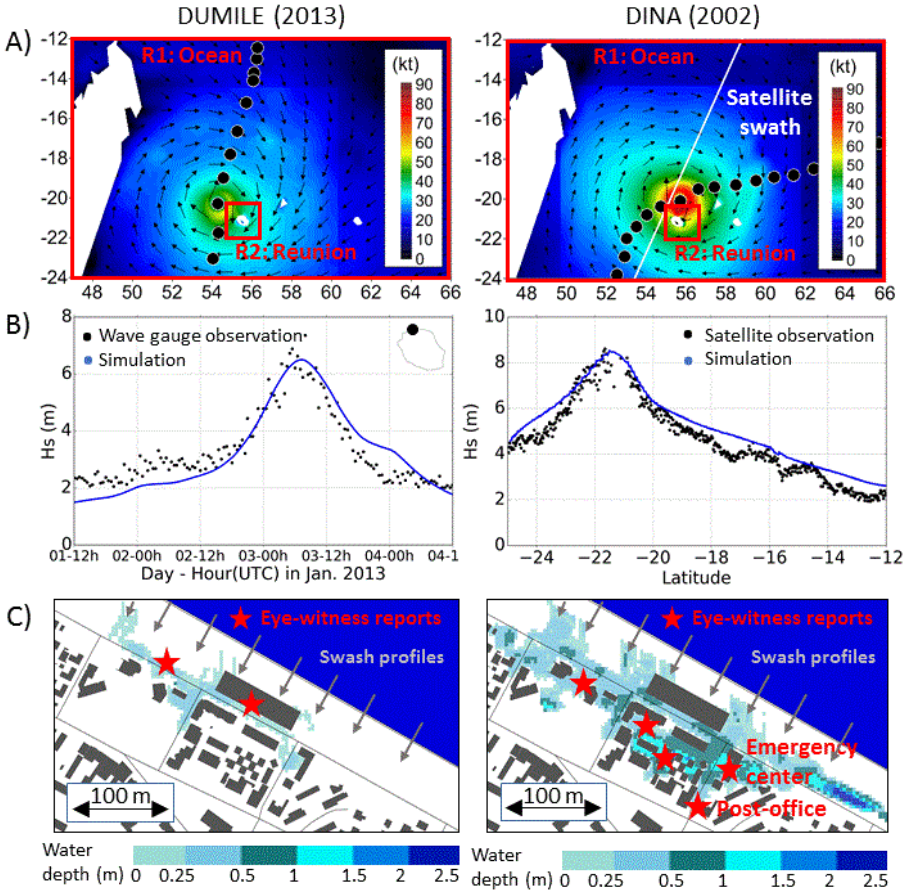


Very high computation time cost (>several hours)!

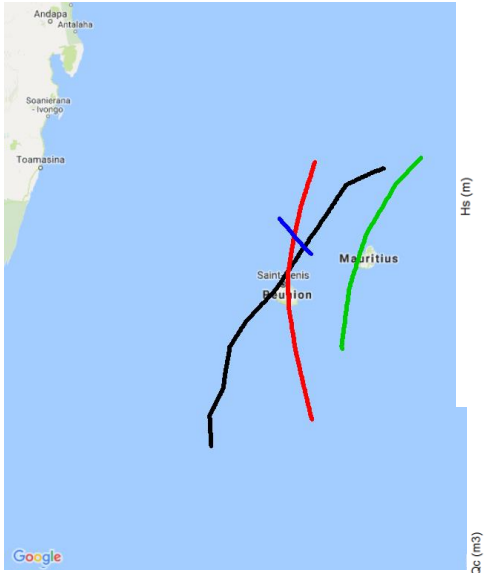
Wave overtopping modeling at local scale



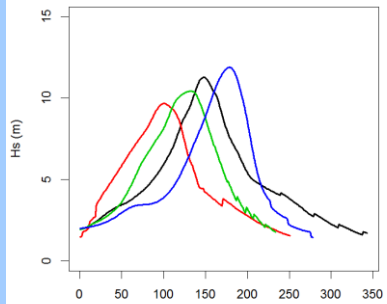
Very high computation time cost (>several hours)!

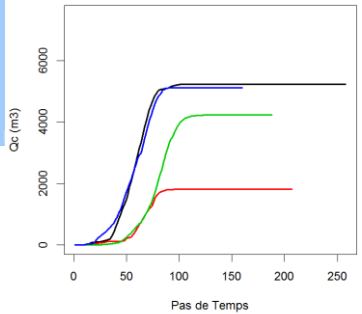


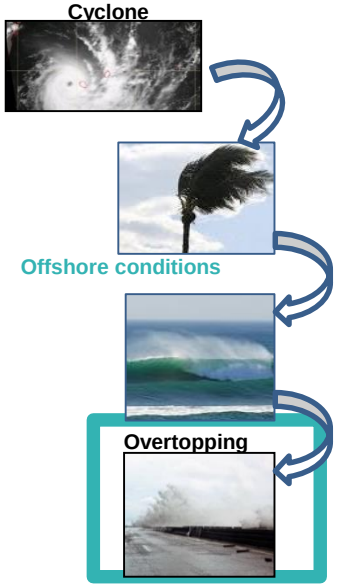
Database of pre-calculated numerical results



Database of pre-calculated simulations for ~500 synthetic cyclones







Very high computation time cost!

Methods

- **Take advantage of the** database (~500 synthetic cyclones)

- Construct the **link** (statistical) betw.:
  - The **offshore conditions** (maximum values for wave, sea level)
  - The **consequences induced by the cyclone:**
    - Overtopping occurence (0/1)
    - Max of cumulative volume

Methods

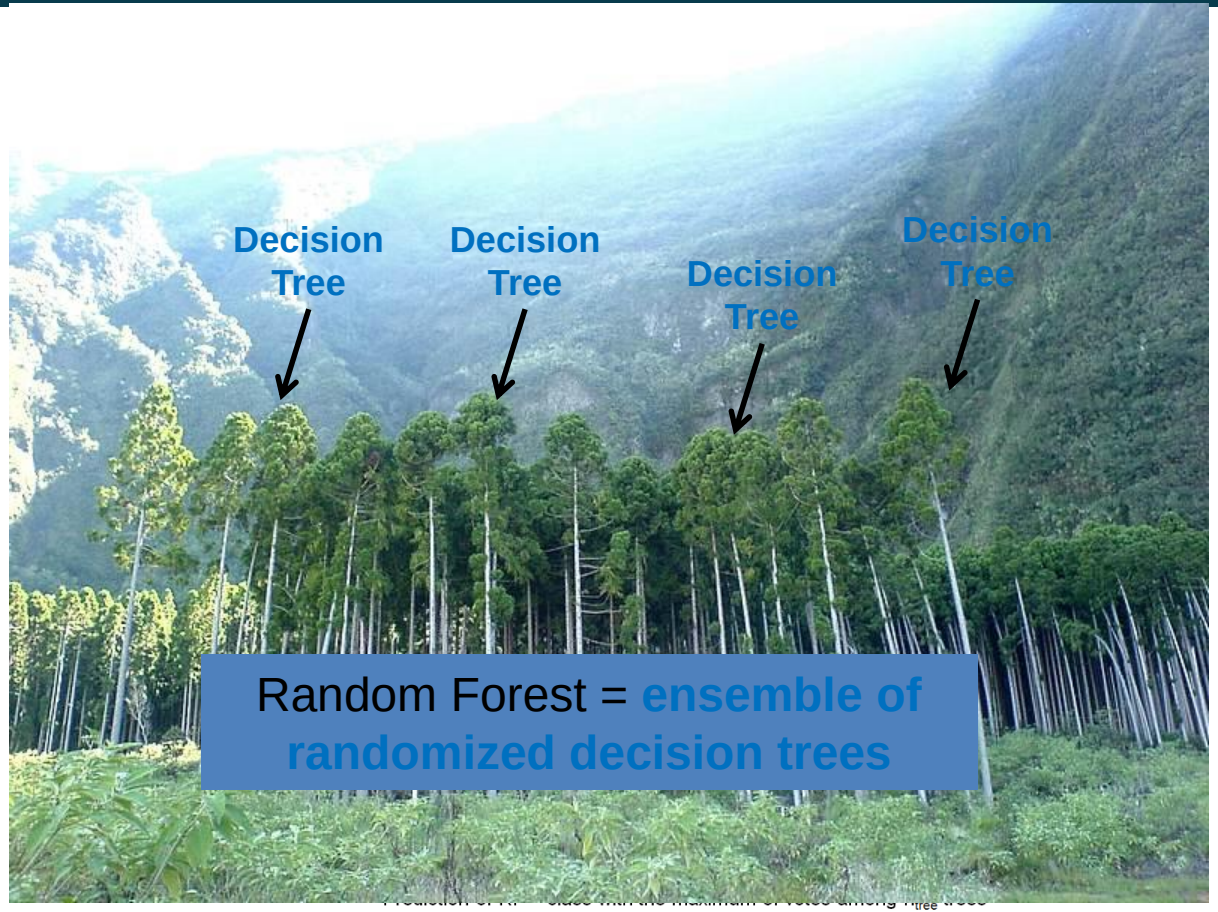
- **Take advantage of the** database (~500 synthetic cyclones)

- Construct the **link** (statistical) betw.:
  - The **offshore conditions** (maximum values for wave, sea level)
  - The **consequences induced by the cyclone:**
    - Overtopping occurence (0/1)
    - Max of cumulative volume

- Use this link to predict with a **negligible coputation time cost** (<min)

- Tools
  - A series of **random forest** approaches
  - **Flexible** to any problem (classification, regression, etc.)

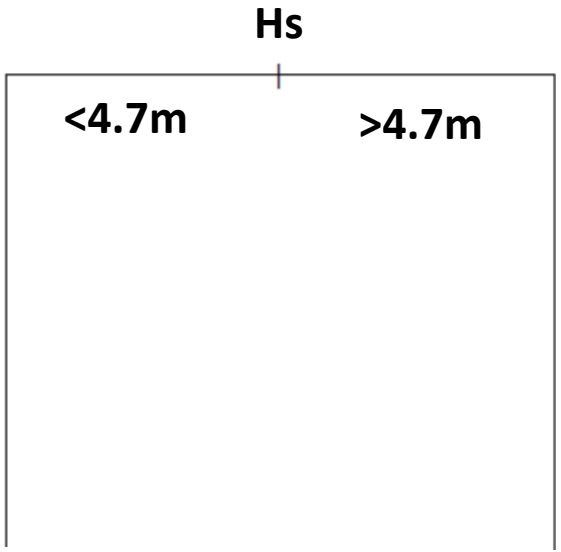
Random Forest



Random Forest = **ensemble of randomized decision trees**

Decision tree

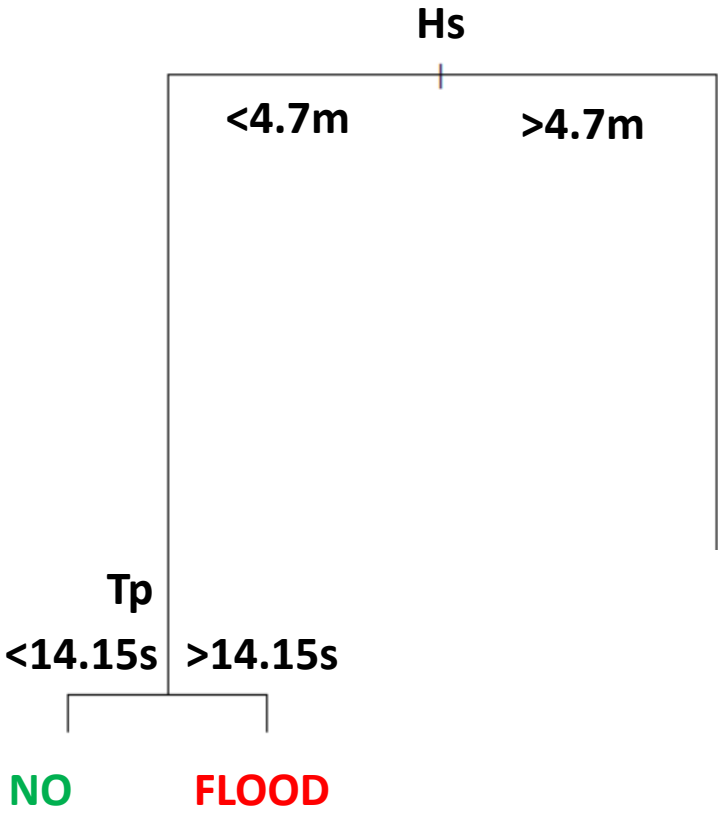
Offshore conditions leading to flooding?



[Breiman (1984)]

Decision tree

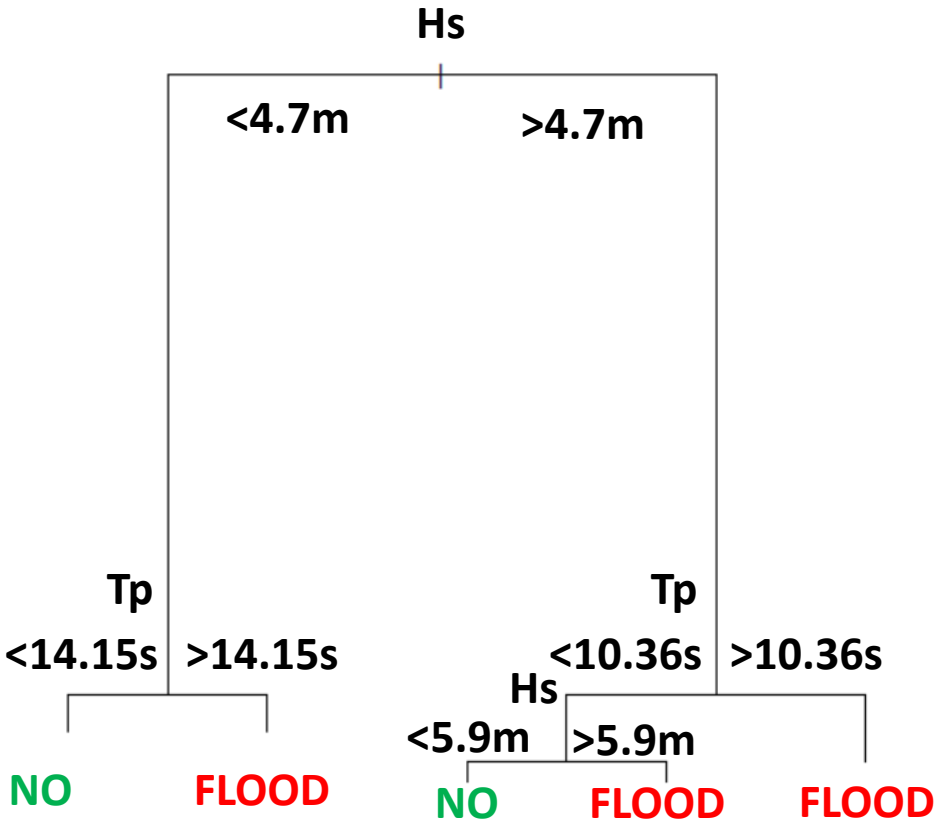
Offshore conditions leading to flooding?



[Breiman (1984)]

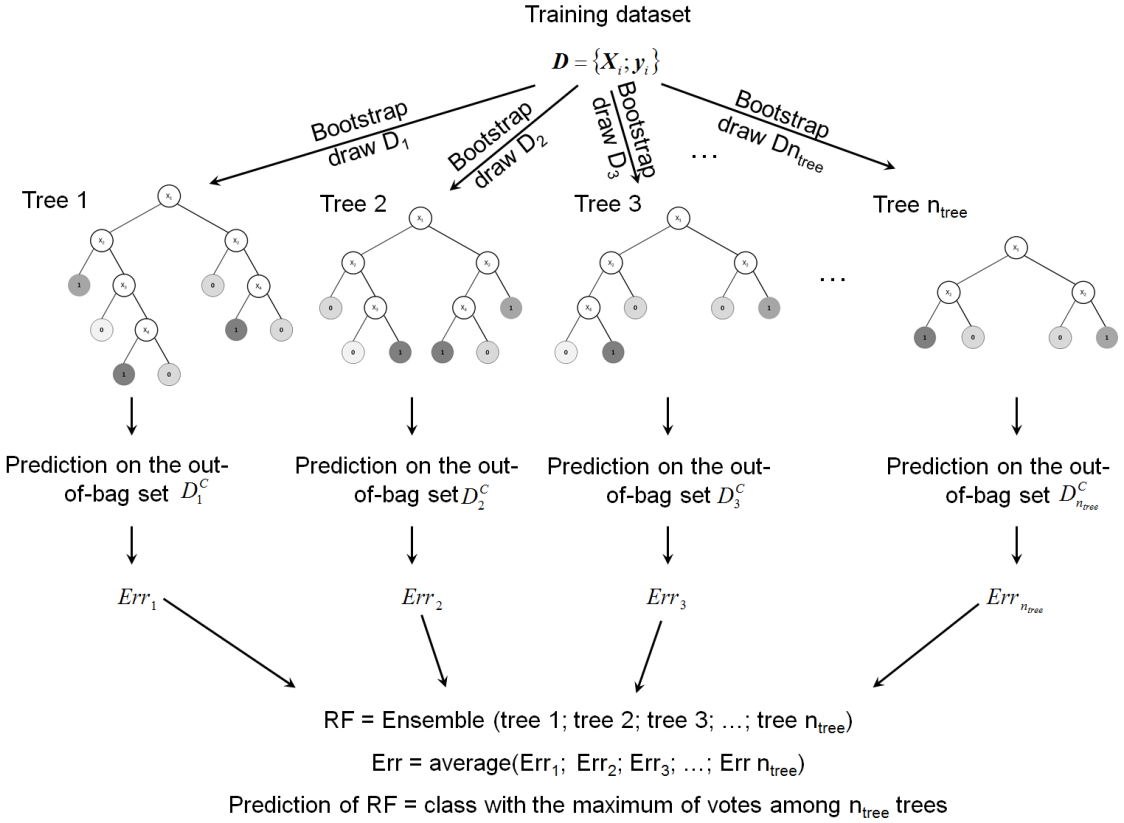
Decision tree

Offshore conditions leading to flooding?



[Breiman (1984)]

Random Forest



- Less sensitive to **threshold values**
- High accuracy
- Only a few parameters to tune
- **Versatile** to different problems (classification, regression, multi-output)

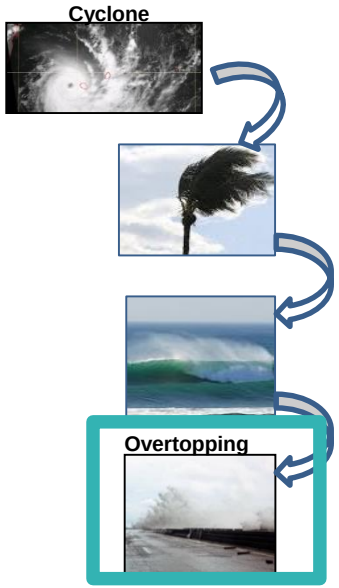
Validation?

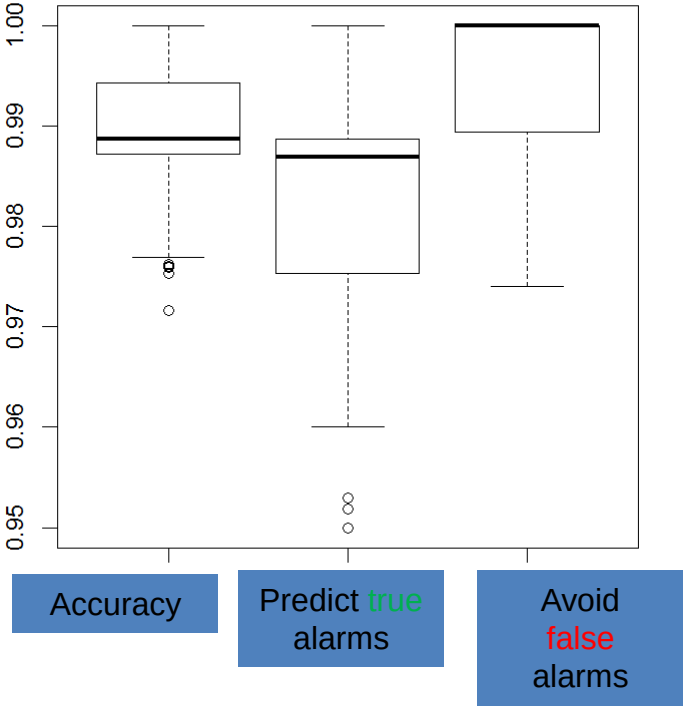
- **Historical cases**
  - Dina-2002 et Dumile-2012;

- Use of the database (**cross-validation**)

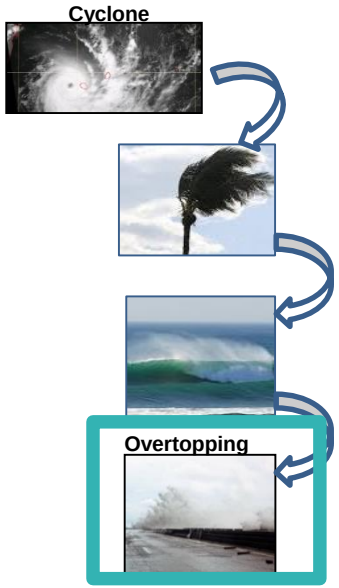
- Performance criteria
  - **Accuracy**
  - **True alarms**
  - **False alarms**

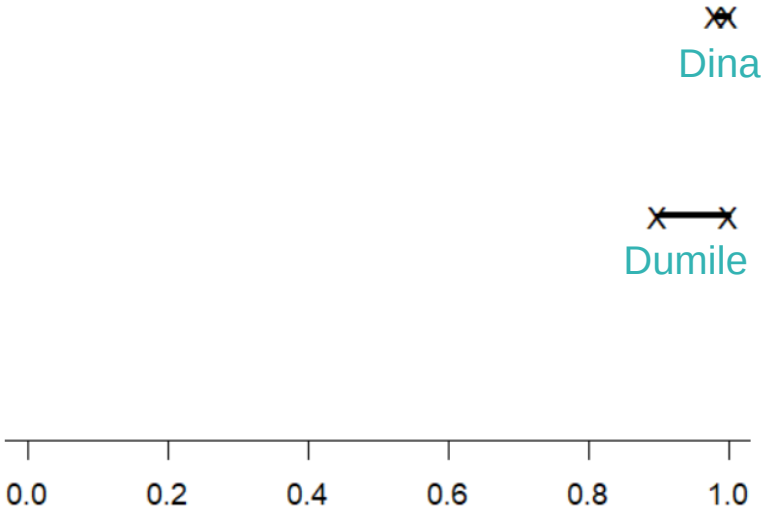
Occurrence of flooding - cross-validation





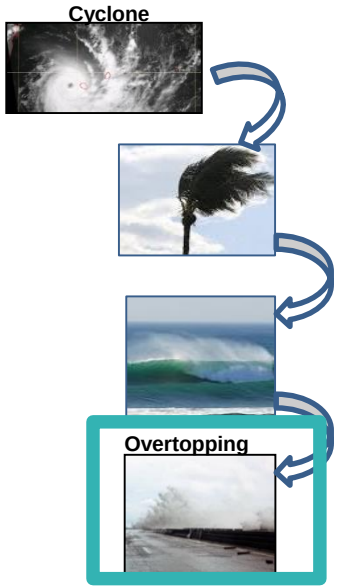
Validation with historical events

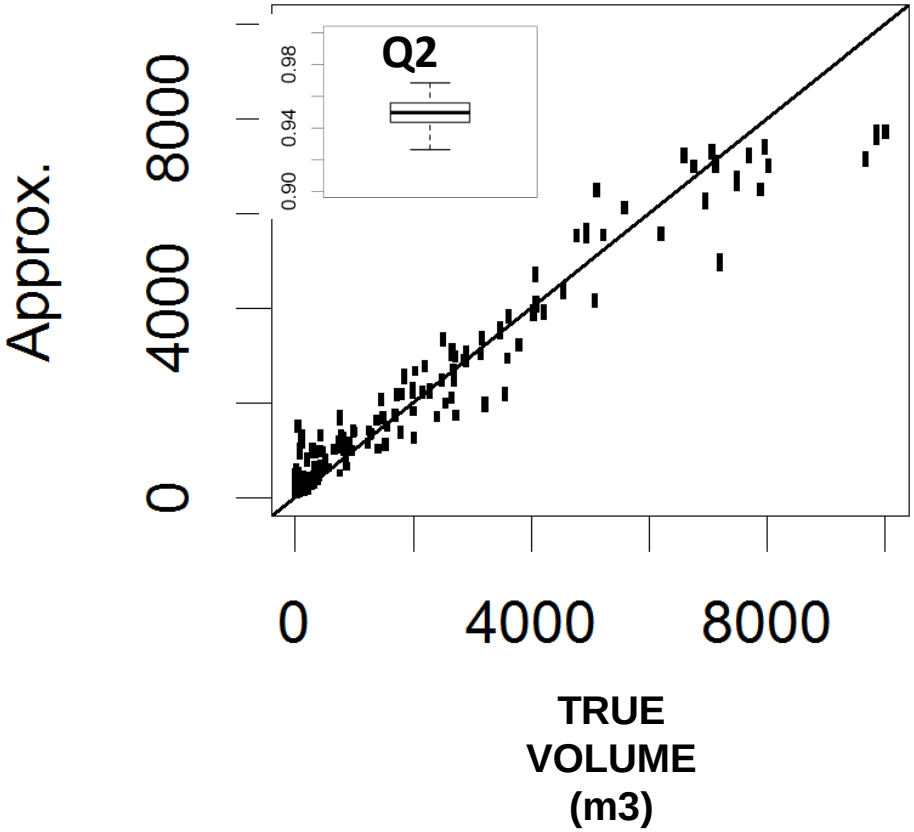




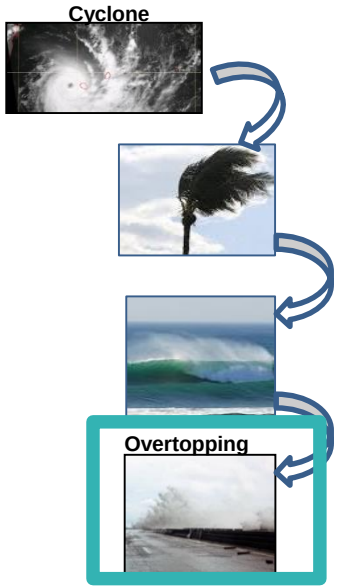
Probability of flooding
Dina: [98-100%]; Dumile [90-100%]

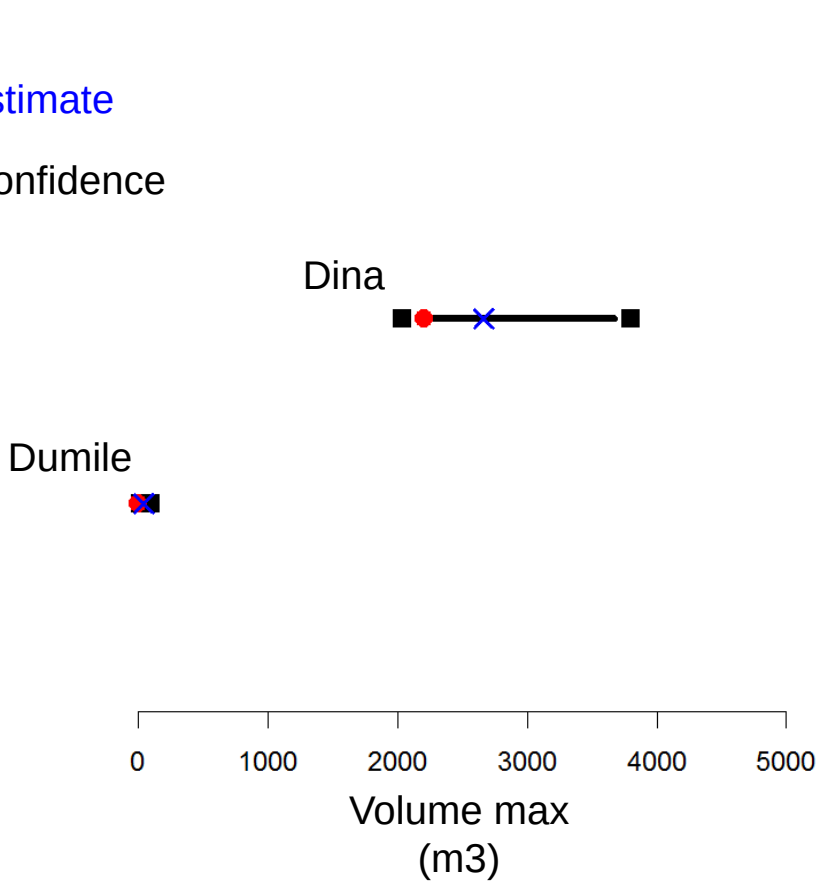
Magnitude of flooding - cross-validation



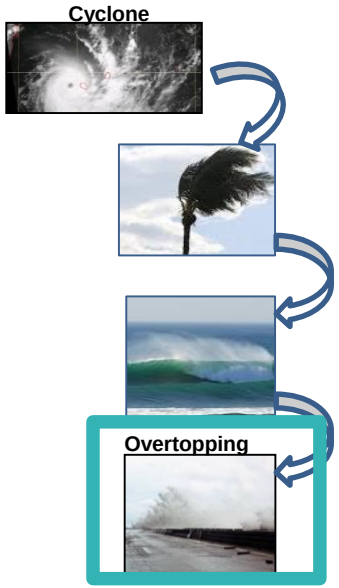


Validation with historical events





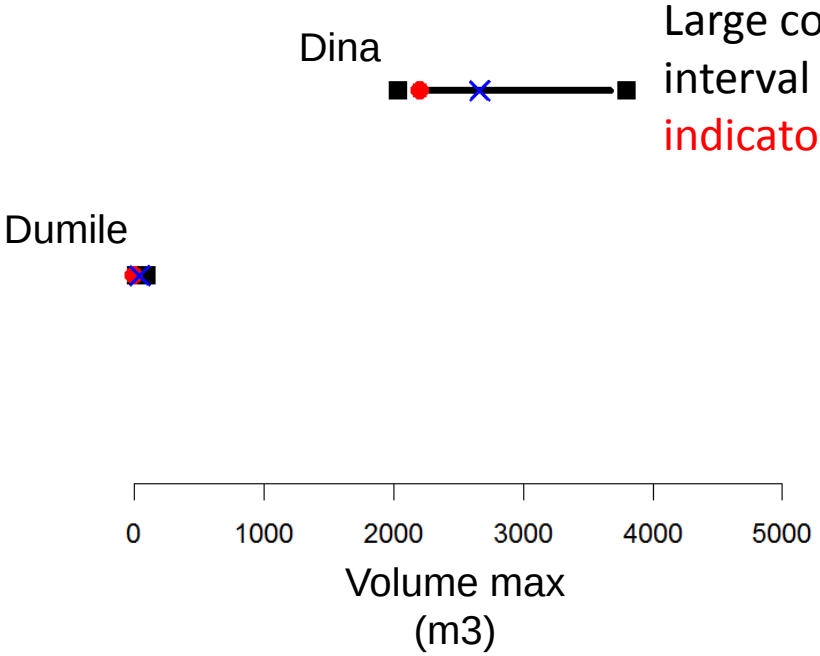
Validation with historical events



Truth

Best estimate

95% Confidence interval



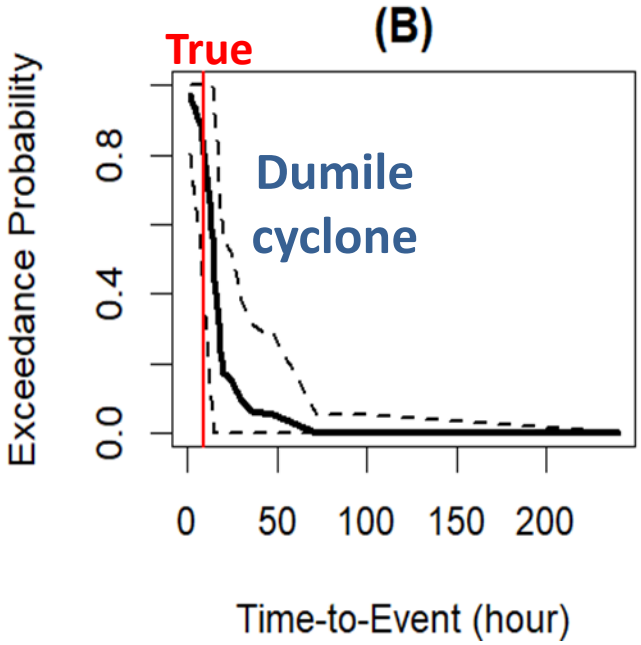
Large confidence interval width → indicator of RF error

Towards time evolution prediction

Time-to-Event=the difference between the time instant when Volume>0 and the one when the considered cyclone is at a distance <400km from Reunion Island

Towards time evolution prediction

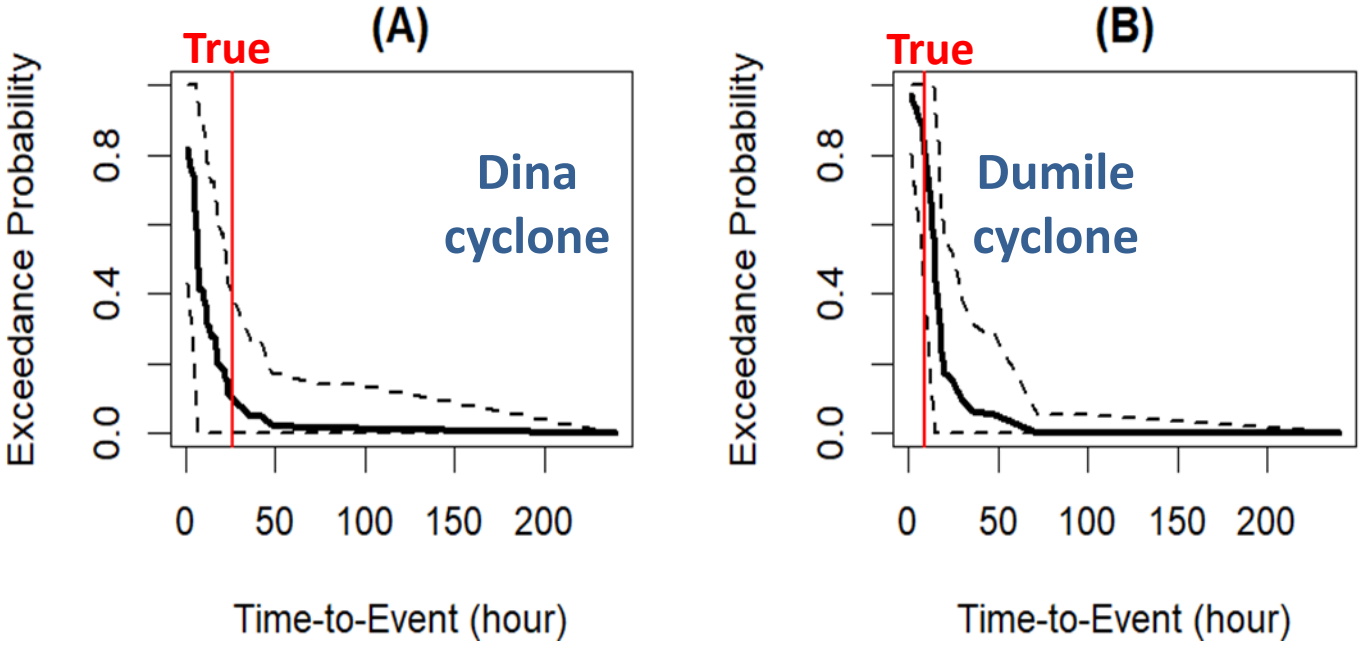
Time-to-Event=the difference between the time instant when Volume>0 and the one when the considered cyclone is at a distance <400km from Reunion Island



→ Adaptation of random-forest model to handle multivariate outputs (Ishwaran et al. 2017)

Towards time evolution prediction

Time-to-Event=the difference between the time instant when Volume>0 and the one when the considered cyclone is at a distance <400km from Reunion Island



→ Adaptation of random-forest model to handle multivariate outputs (Ishwaran et al. 2017)

Speeding up EWS?

- **Method**
  - Through the **statistical analysis** of precalculated numerical simulations
  - A suite of **Random forest procedures**
  - **Flexible** approach
- **To overcome**
  - Hydrodynamic simulator with **high computation time cost**
  - The input-output relationship is **non-smooth** (highly nonlinear)
- **Further work**
  - Inputs/outputs are **spatio-temporal quantities**
  - Integrate **approximation uncertainties**

Thank you for your attention!

Early-warning system for cyclone-induced wave overtopping:

a suite of random forest approaches

J. Rohmer[1*], S. Lecacheux[1], R. Pedreros[1], D. Idier[1], F. Bonnardot[2]
j.rohmer@brgm.fr
1: BRGM
2: Direction Régionale de Météo-France pour l'Océan Indien

Géosciences pour une Terre durable
brgm

Only one part of the problem…

Weather Forecast

Wind model

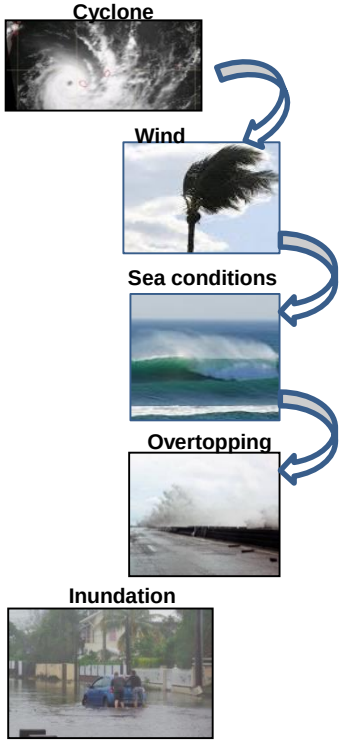
Wave model

Flooding model

Metamodel

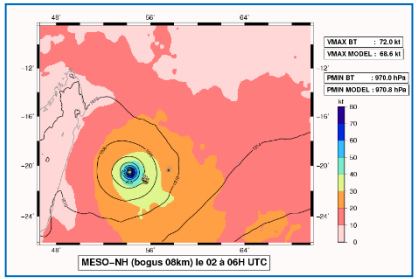
Cascade of uncertainties

This presentation

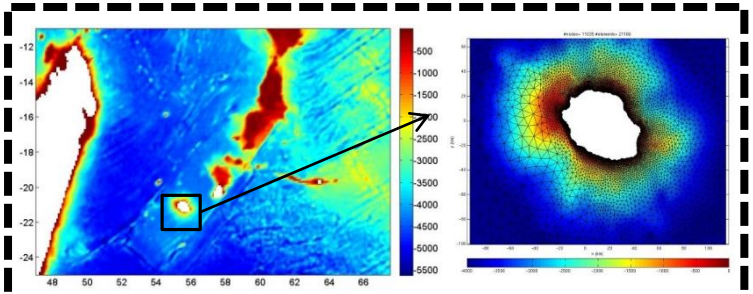


Appendices

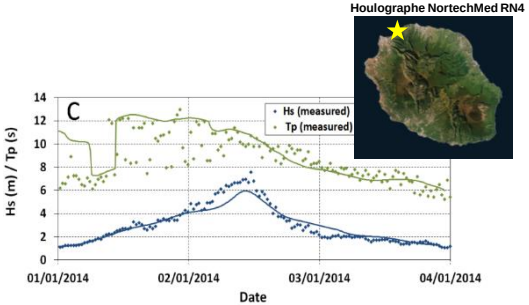
Wave modeling at regional scale

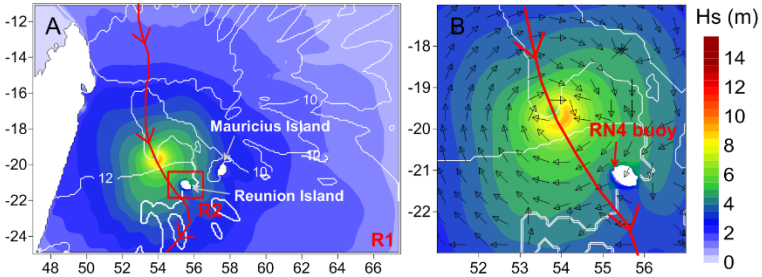


Wavewatch 3 two-way nested grids (10km => 300m)



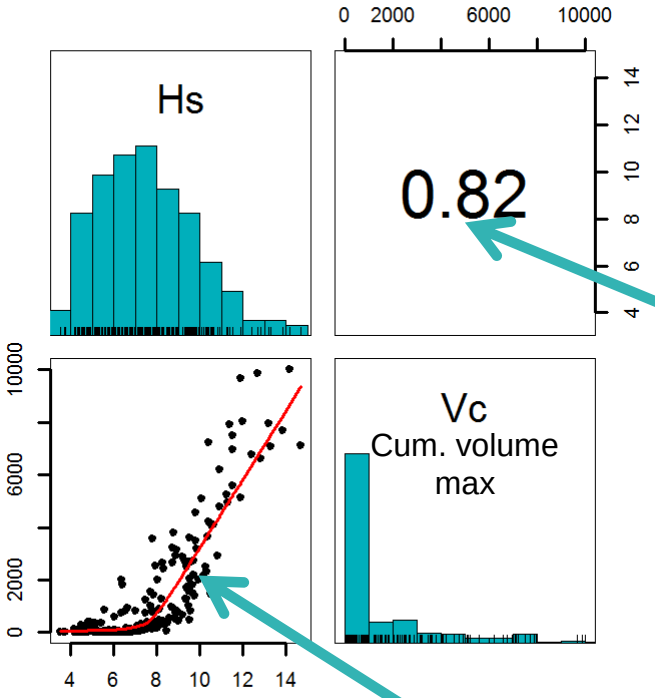
Computing time ~ 25min on 24 CPU for 24h simulated





BEJISA le 02/01/2014 à 12UTC

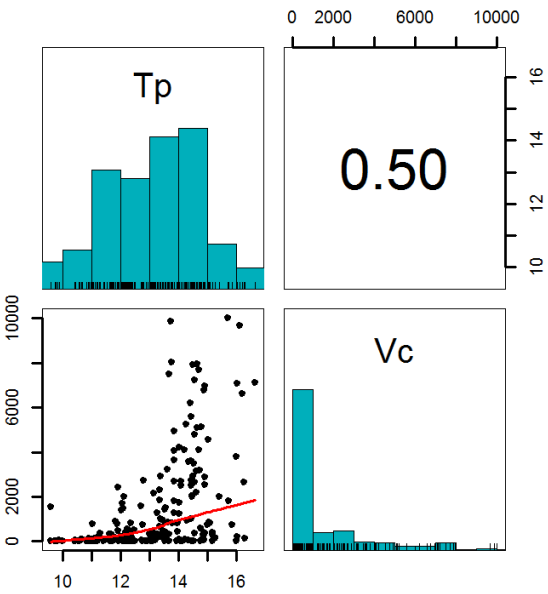
Empirical analysis of the database

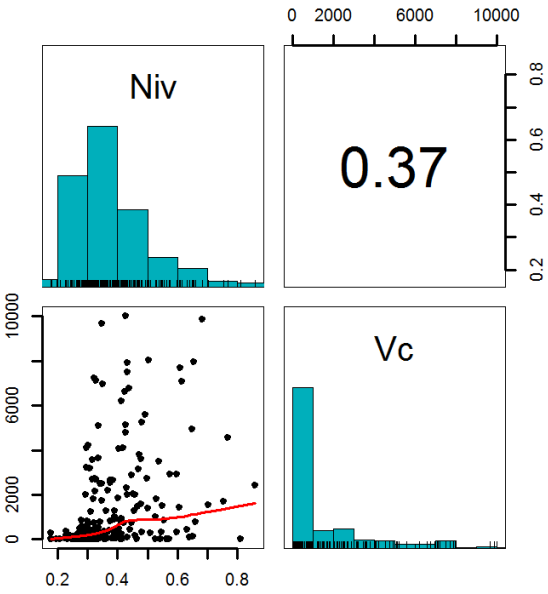


Lin. correlation

Highly nonlinear behaviour

Empirical analysis of the database

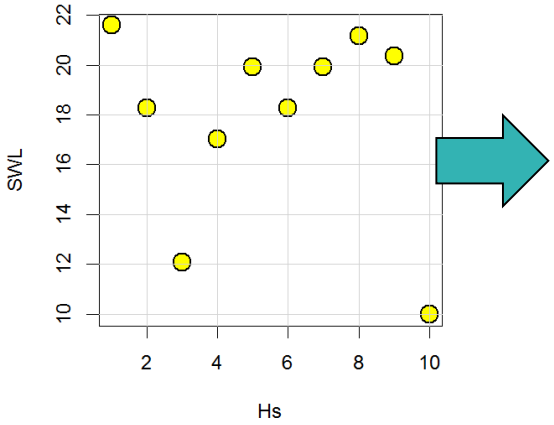




Use knowledge on Hs, Tp et TWL to predict the occurrence of overtopping via regression (random forest)

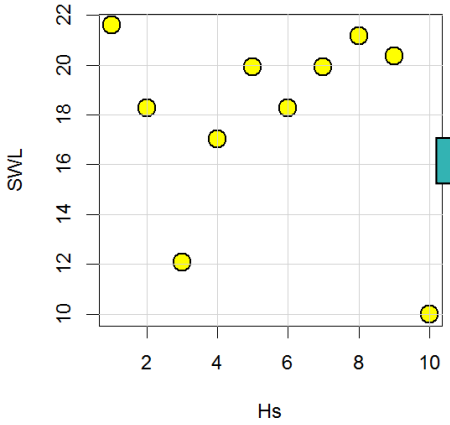
PRINCIPES

10 Scénarios de simulations

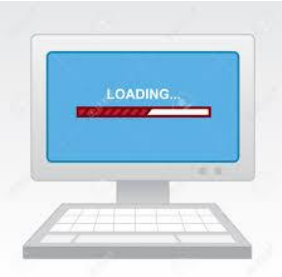


PRINCIPES

10 Scénarios de simulations



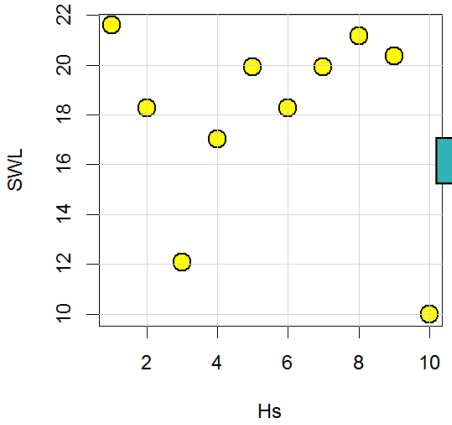
Simulation numérique



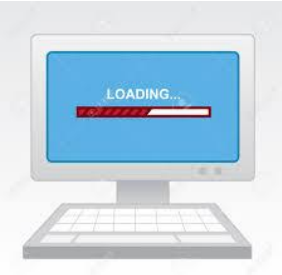
Etape couteuse en temps de calcul!

PRINCIPES

10 Scénarios de simulations



Simulation numérique

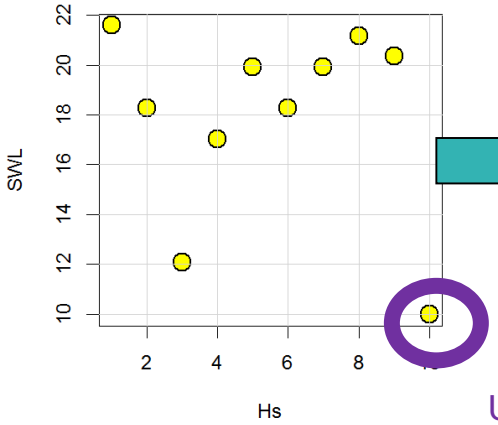


10 Résultats

Aire d'inondation

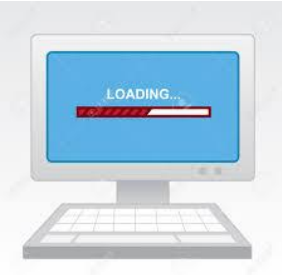
PRINCIPES

10 Scénarios de simulations



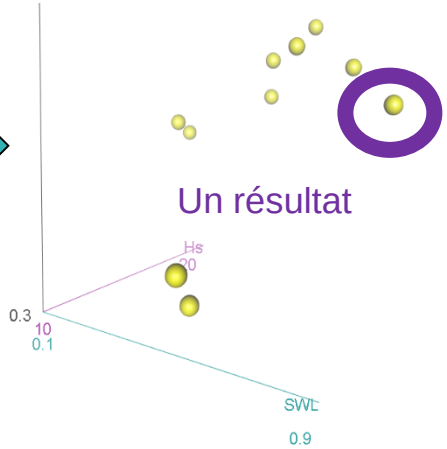
Un scénario

Simulation numérique



10 Résultats

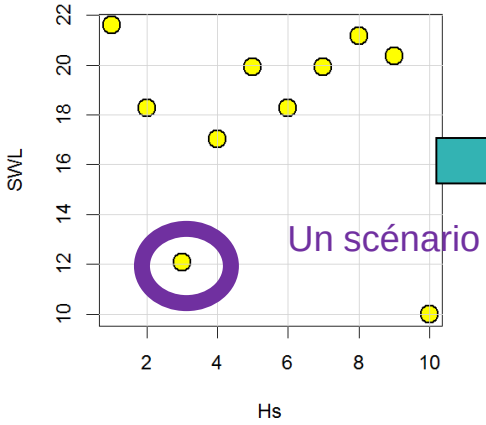
Aire d'inondation



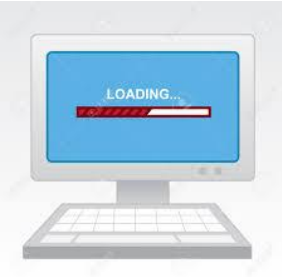
Un résultat

PRINCIPES

10 Scénarios de simulations

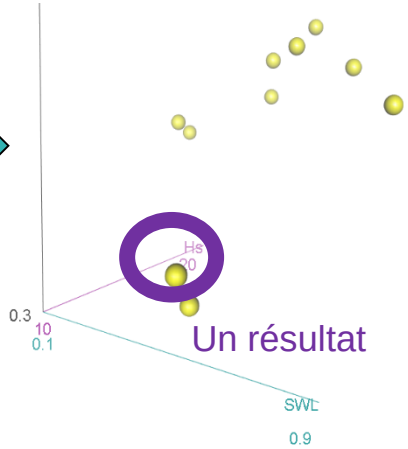


Simulation numérique



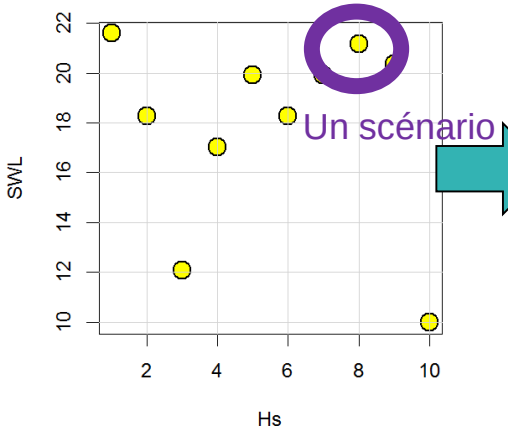
10 Résultats

Aire d'inondation

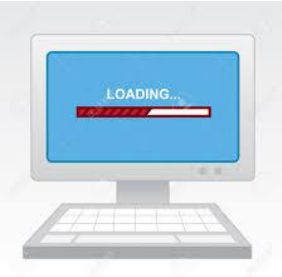


PRINCIPES

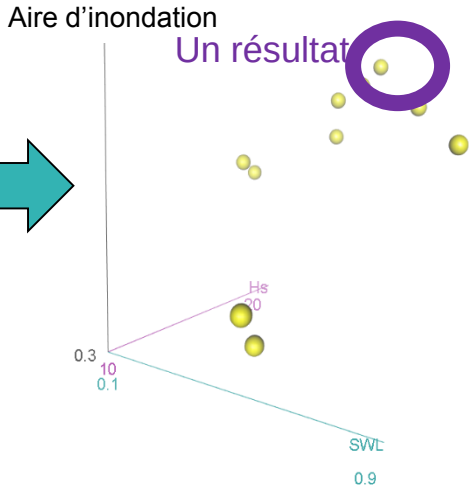
10 Scénarios de simulations



Simulation numérique

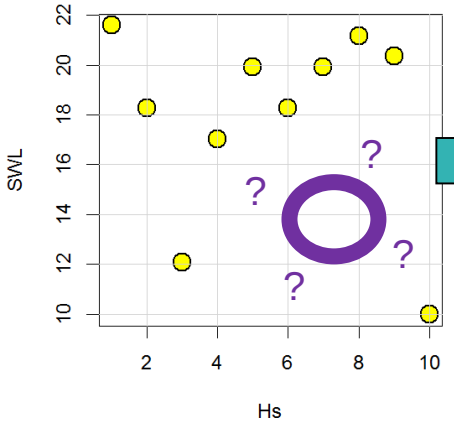


10 Résultats

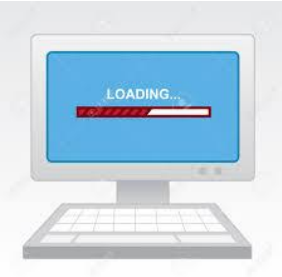


PRINCIPES

10 Scénarios de simulations

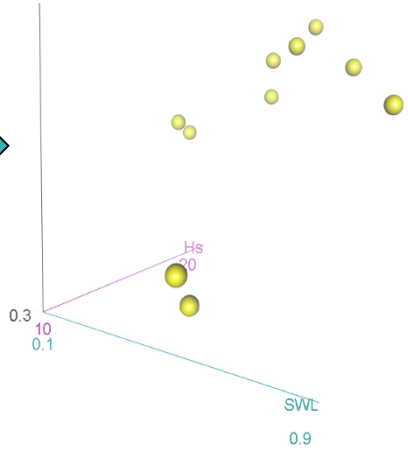


Simulation numérique



10 Résultats

Aire d'inondation

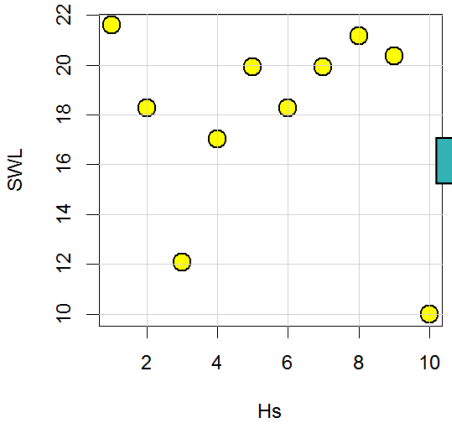


Que dire là où n'a pas fait de simulation ?

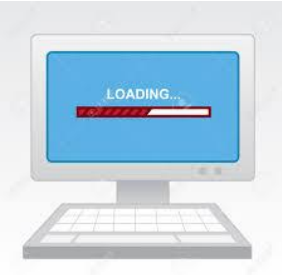
SANS utiliser le simulateur couteux en temps de calcul!!

PRINCIPES

10 Scénarios de simulations

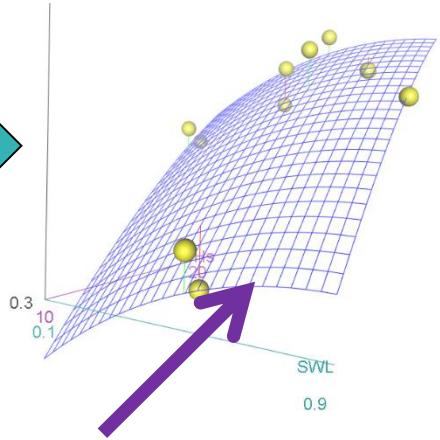


Simulation numérique



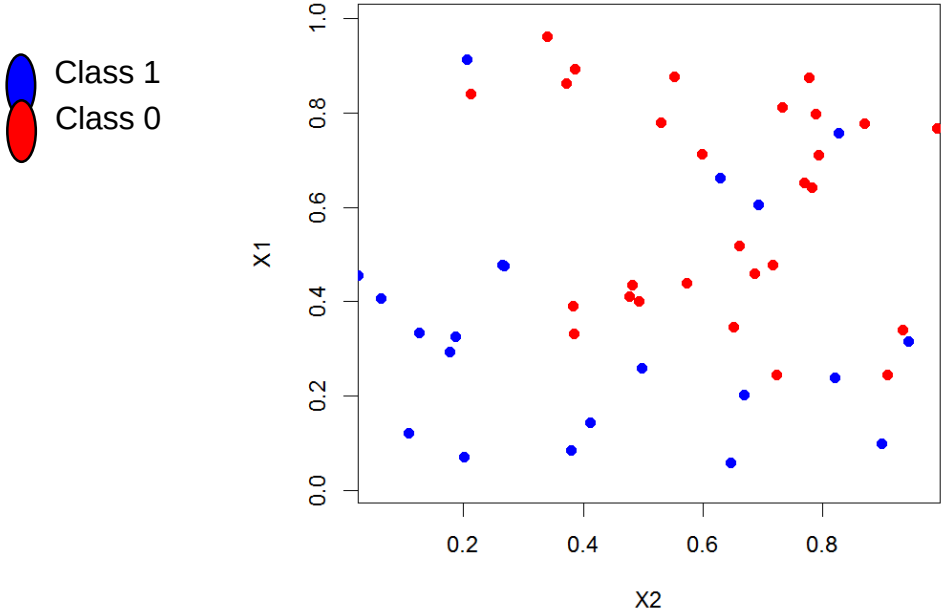
10 Résultats

Aire d'inondation



Combler les « trous »
avec un méta-modèle
(modèle statistique)

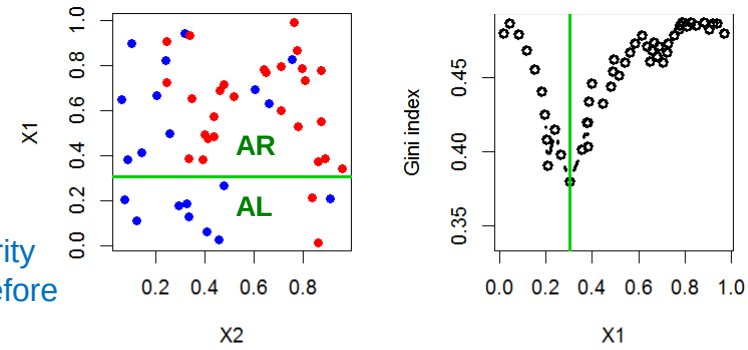
Classification using decision tree



Classification:

1. **Splitting rule: maximize the decrease in heterogeneity (measured by Gini impurity measure)**
2. **Stopping rule (Gini index does not evolve or max number)**

Classification using decision tree



Gini impurity measure before splitting

$$G = p_1(A).p_0(A) - \frac{N(A_L)}{N(A)} p_1(A_L).p_0(A_L) - \frac{N(A_R)}{N(A)} p_1(A_R).p_0(A_R)$$

A Inital domain

A_L Left domain after splitting

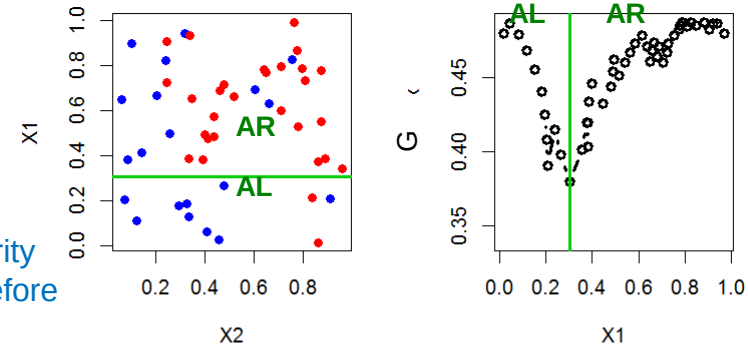
A_R Right domain after splitting

p_k Probability of class k = {0;1}

N Number of elements

[Breiman (1984)]

Classification using decision tree



Gini impurity measure before splitting

$$G = p_1(A).p_0(A) - \frac{N(A_L)}{N(A)} p_1(A_L).p_0(A_L) - \frac{N(A_R)}{N(A)} p_1(A_R).p_0(A_R)$$

A Inital domain

A_L Left domain after splitting

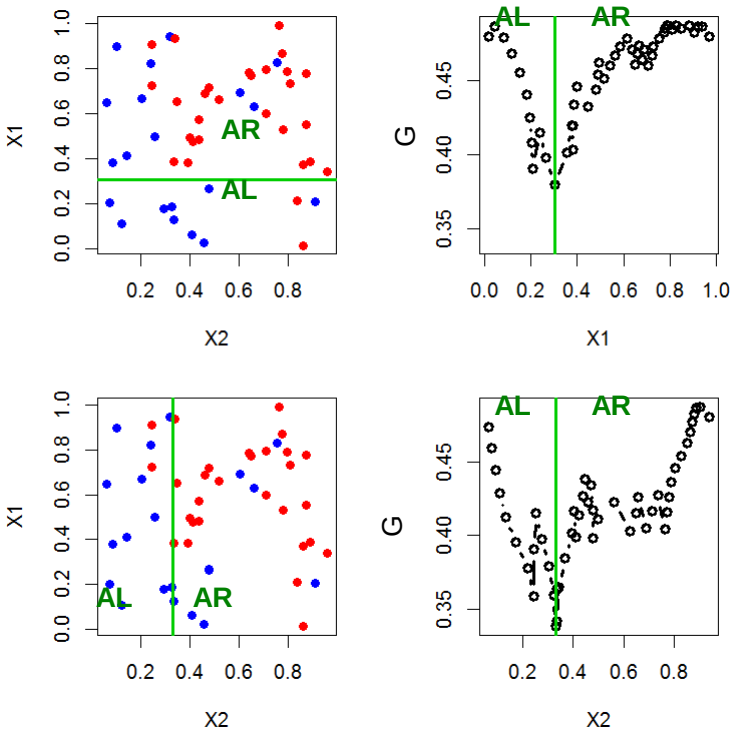
A_R Right domain after splitting

p_k Probability of class k = {0;1}

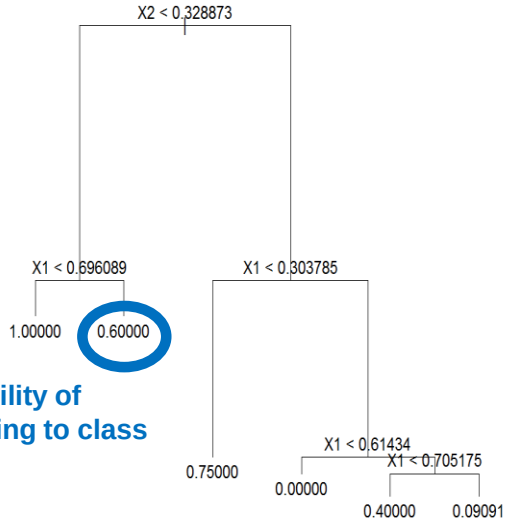
N Number of elements

[Breiman (1984)]

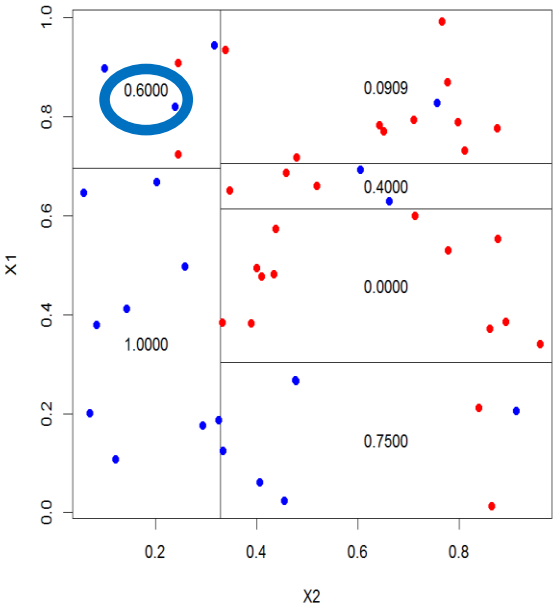
Classification using decision tree



Classification using decision tree

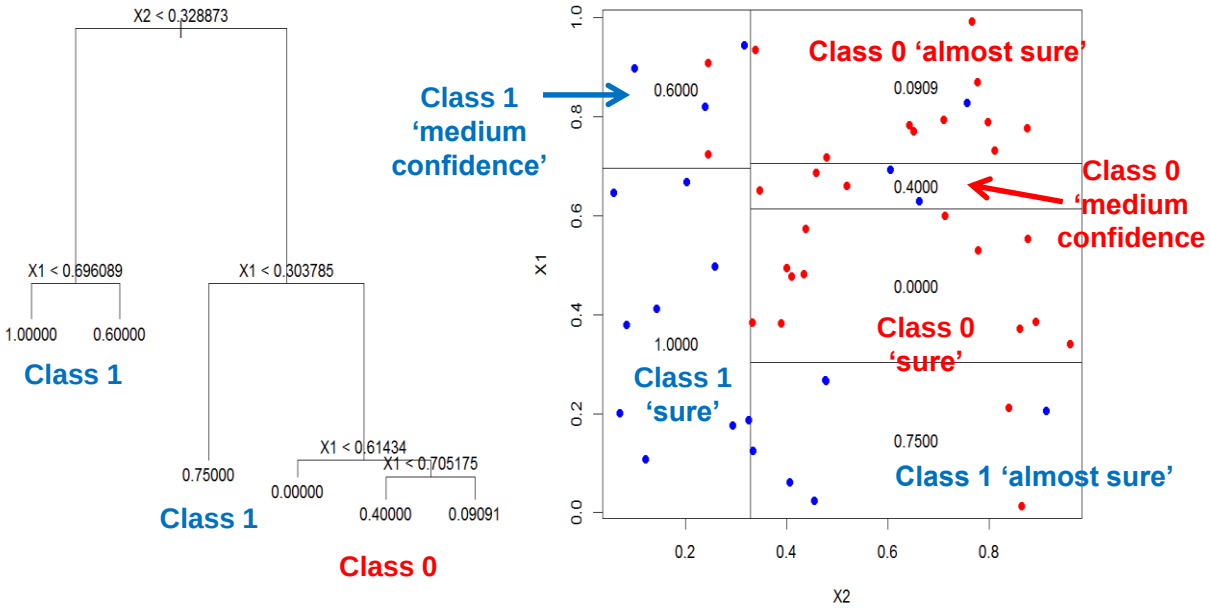


Probability of belonging to class 1

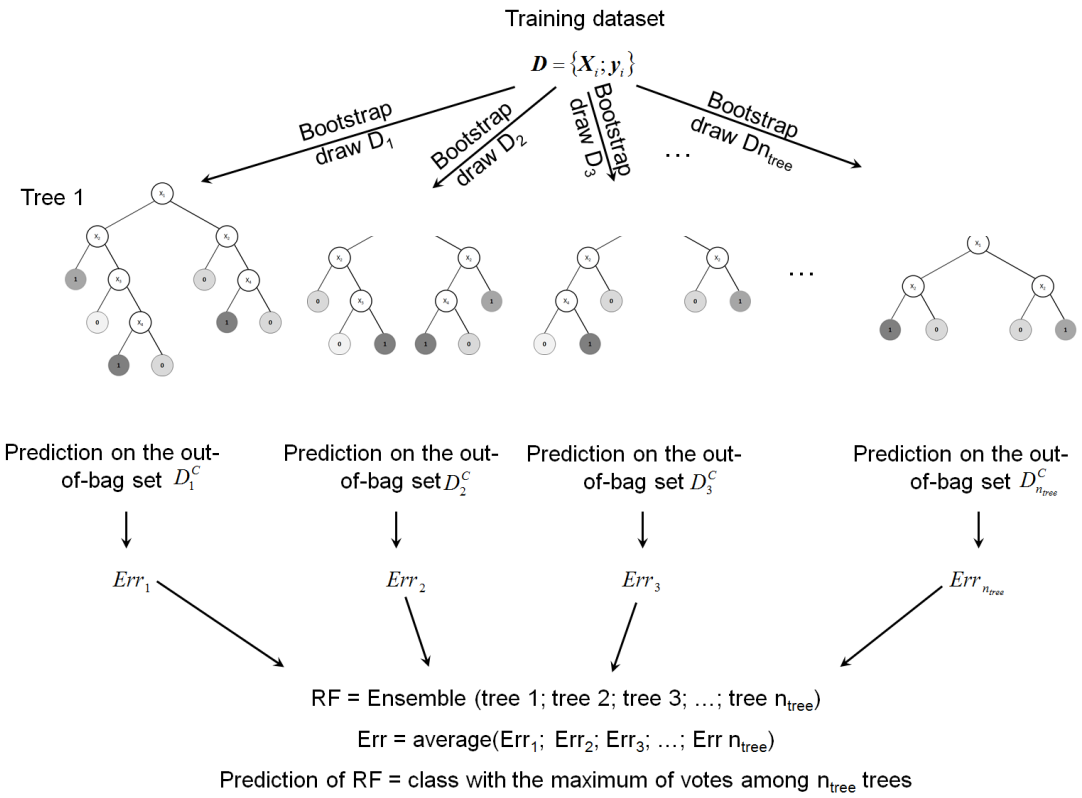


Package R: tree; Approch of Malley (2012) for estimating the class probability

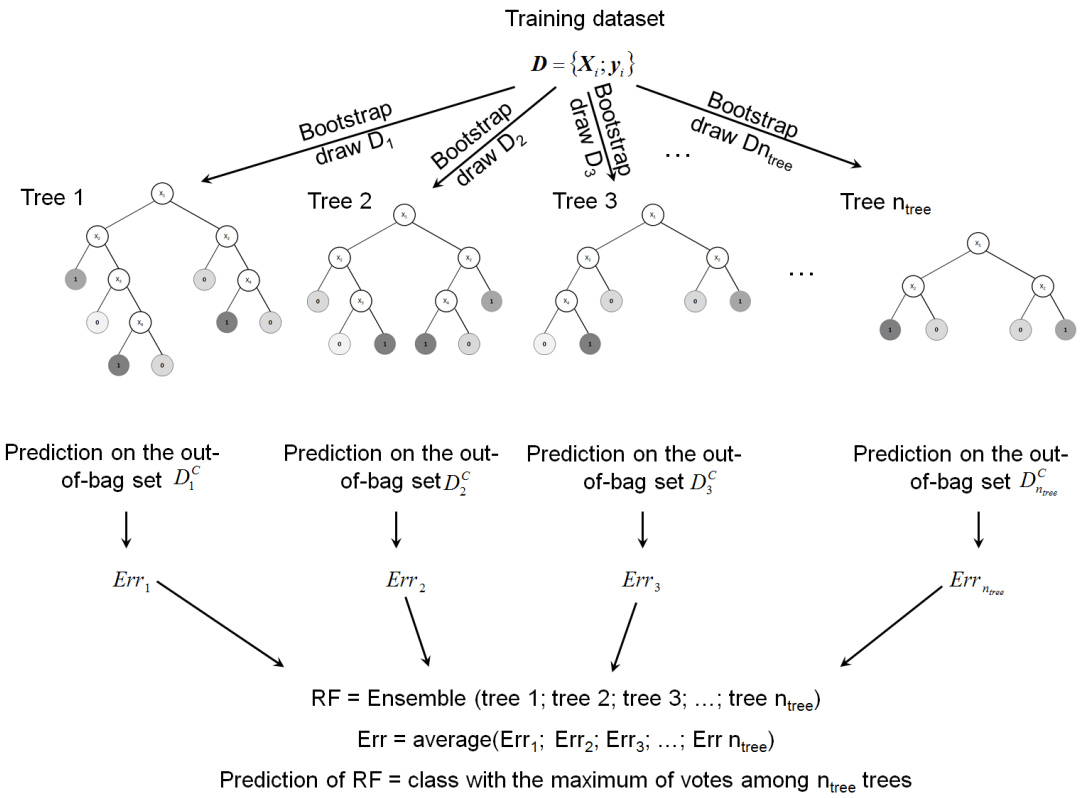
Classification using decision tree



Package R: tree; Approch of Malley (2012) for estimating the class probability

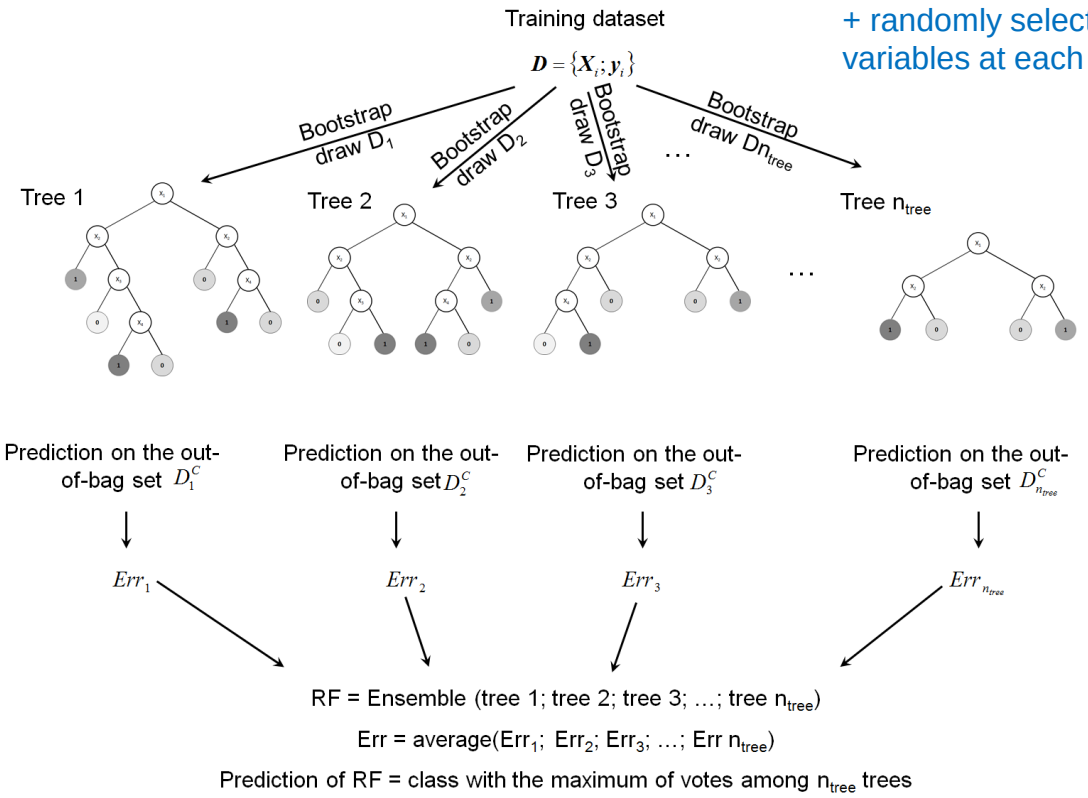


[Breiman (2001)]



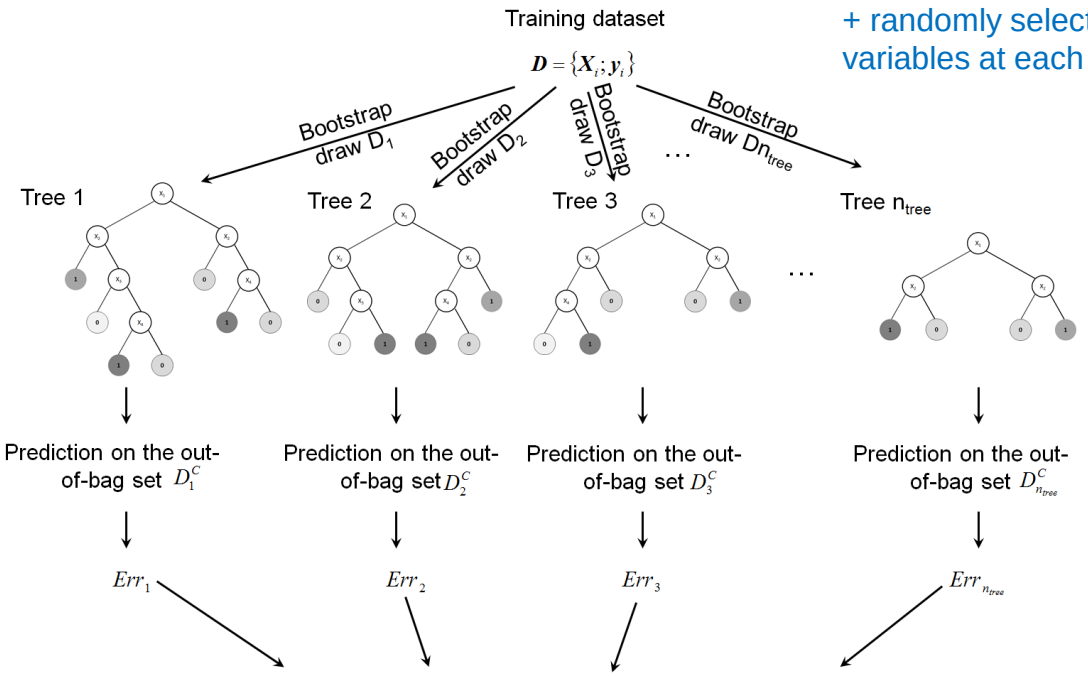
[Breiman (2001)]

+ randomly select mtry variables at each node



[Breiman (2001)]

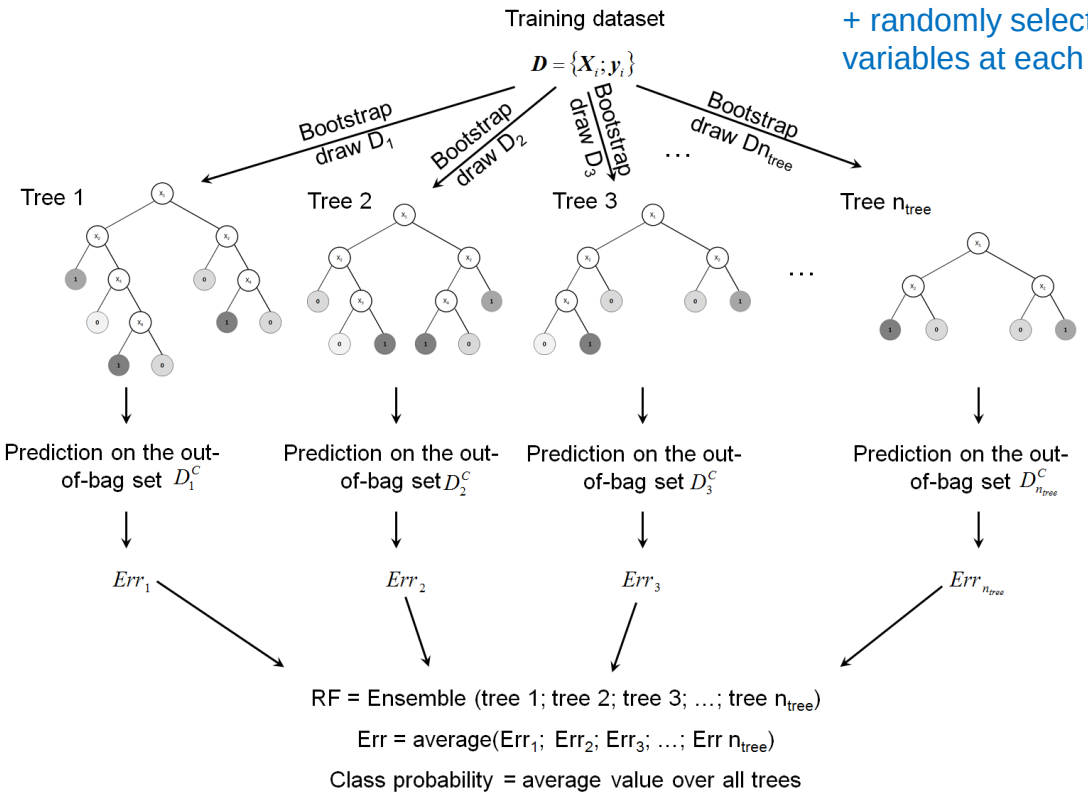
+ randomly select mtry variables at each node



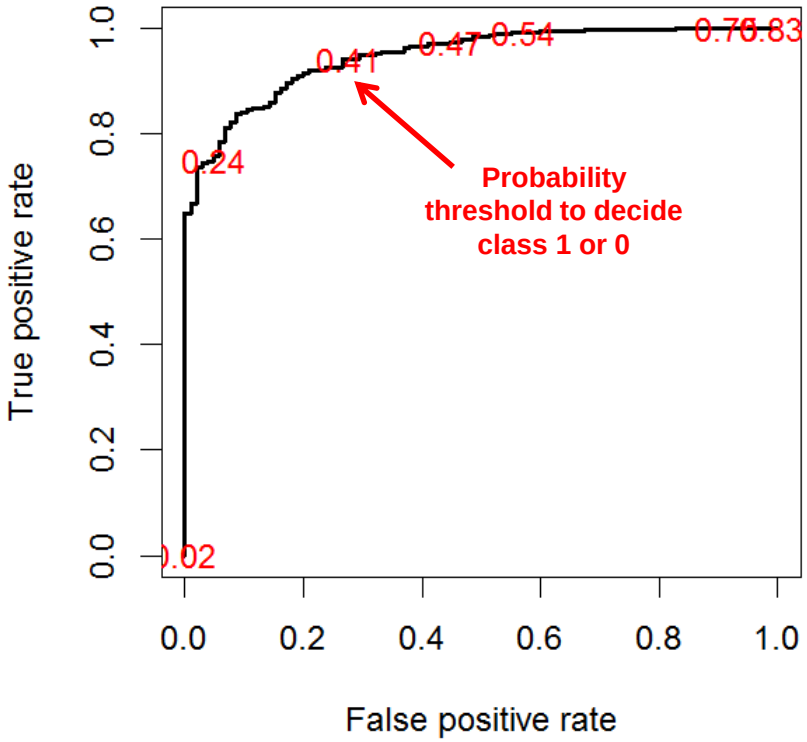
$$Err_j = \frac{1}{\left|\boldsymbol{D}_j^C\right|} \sum_{i \in \boldsymbol{D}_j^C} I(y_i \neq \hat{y}_i) \qquad j = 1, \ldots, n_{tree}$$

[Breiman (2001)]

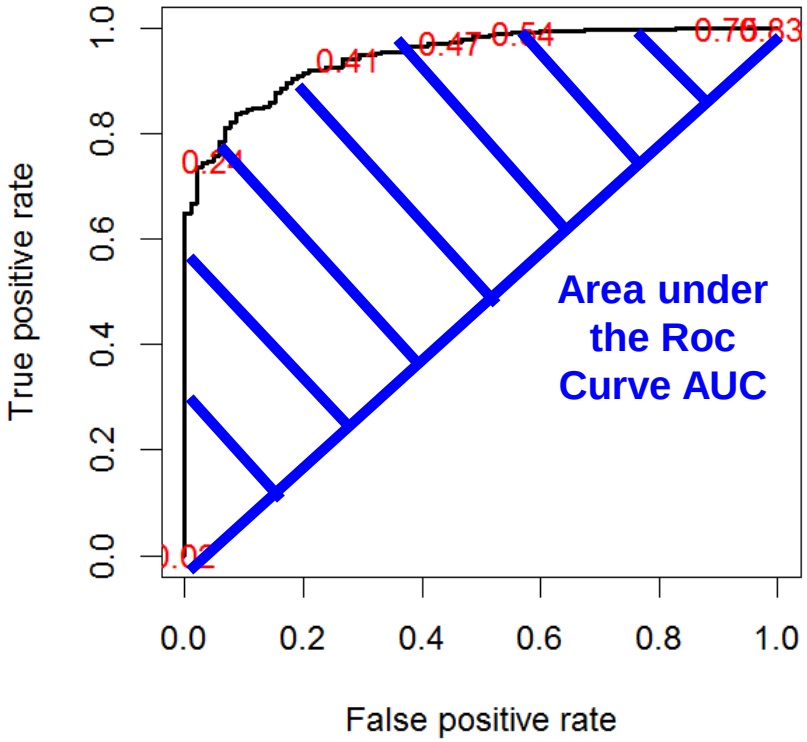
+ randomly select mtry variables at each node



[Breiman (2001)]; package R ranger



[Metz (1978)]; package R pROC



[Metz (1978)]; package R pROC