Identification of dam behaviour by means of machine learning classification models

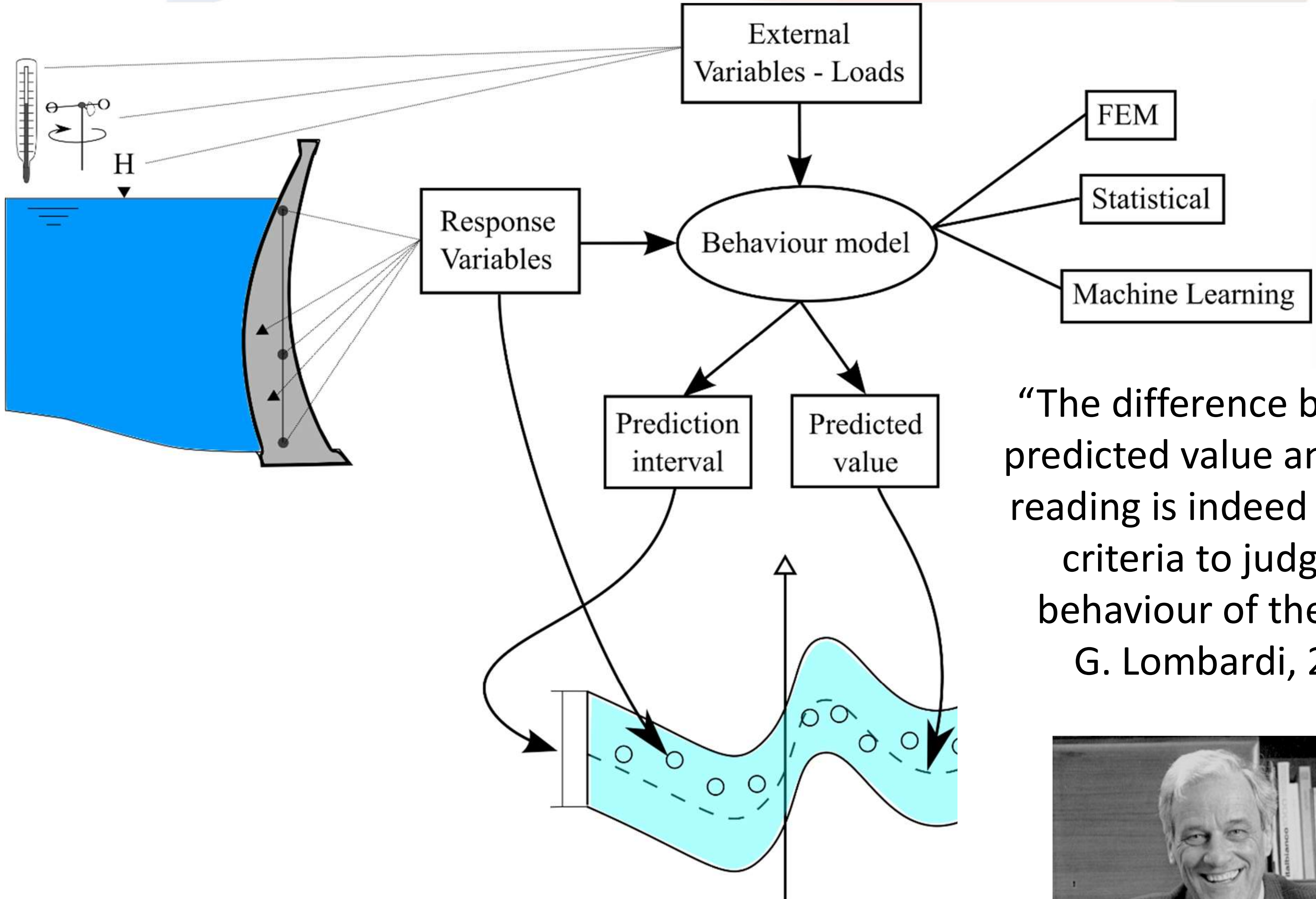
F. Salazar, A. Conde, D.J. Vicente
INTERNATIONAL CENTER FOR NUMERICAL METHODS IN ENGINEERING
CIMNE

CIMNE

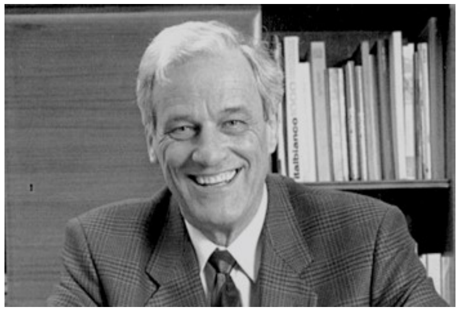
Numerical analysis of dams

ICOLD-BW 2019
9th-11th September
MILANO

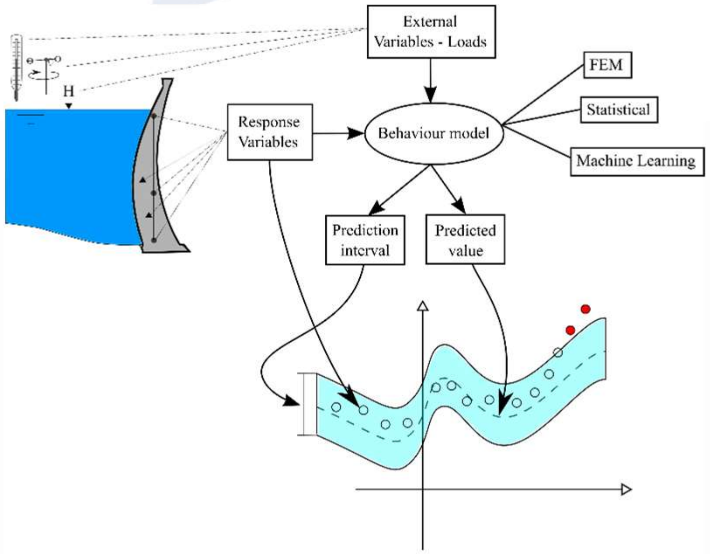
1. INTRODUCTION



“The difference between predicted value and actual reading is indeed the true criteria to judge the behaviour of the dam”
G. Lombardi, 2004



1. INTRODUCTION



Each response variable is analyzed separatedly

As many model evaluations/analysis as relevant outputs

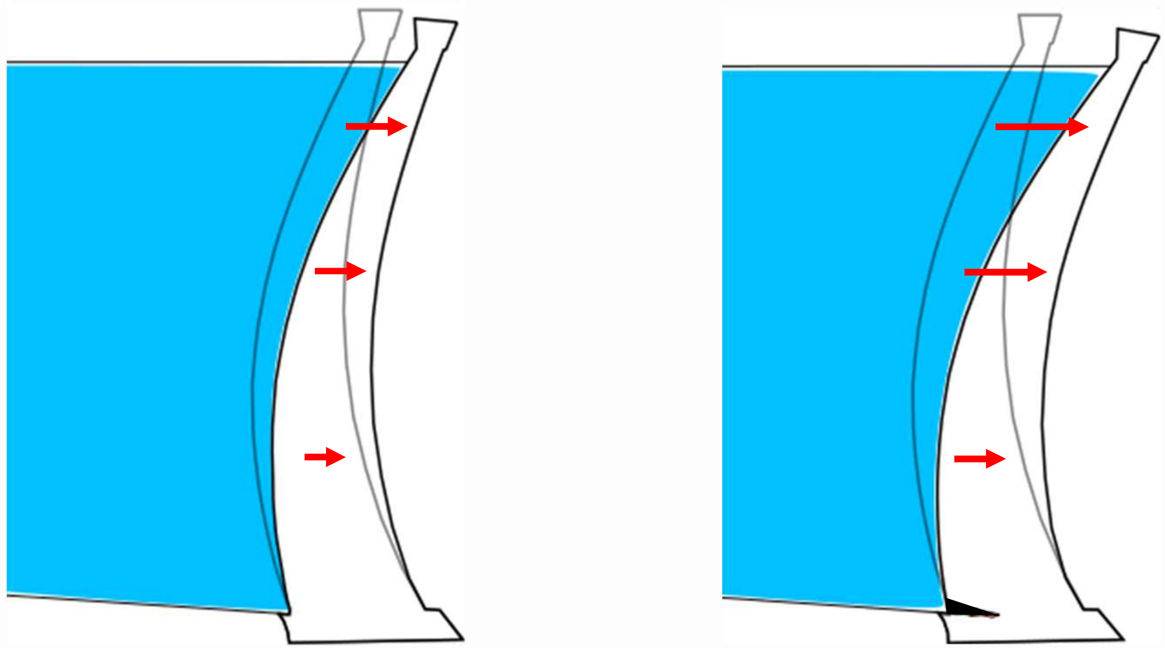
Relation between anomalies and potential failure modes is required

Joint analysis of a set of variables

Identification of behaviour patterns

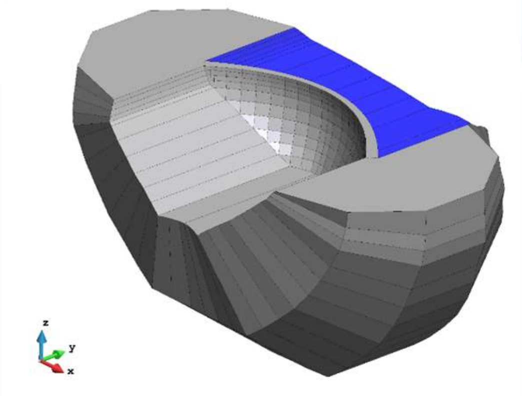
1. INTRODUCTION

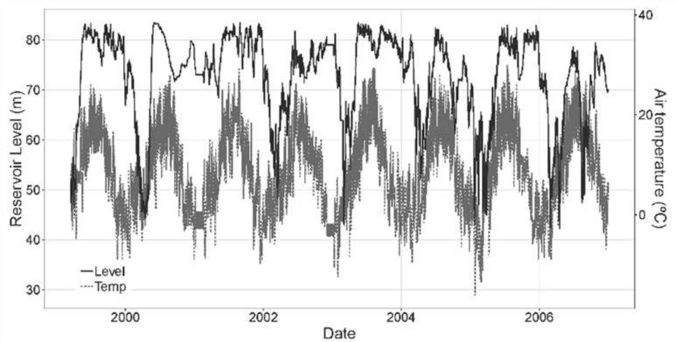
Joint analysis of a set of variables

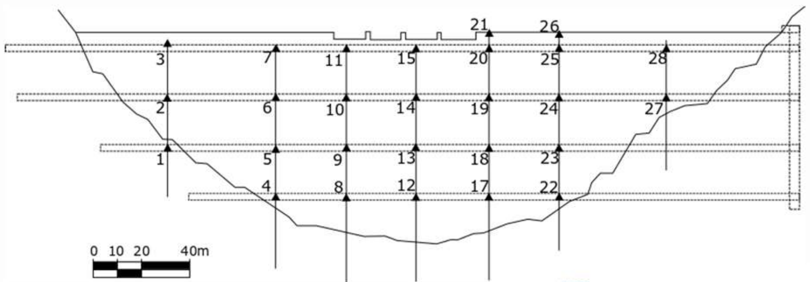


2. METHODOLOGY

1. FEM model
2. Calibration
3. Reference data series
4. Definition of potential anomalies
5. Simulation
6. Anomalous data series
7. Classification model fitting
8. Pattern recognition



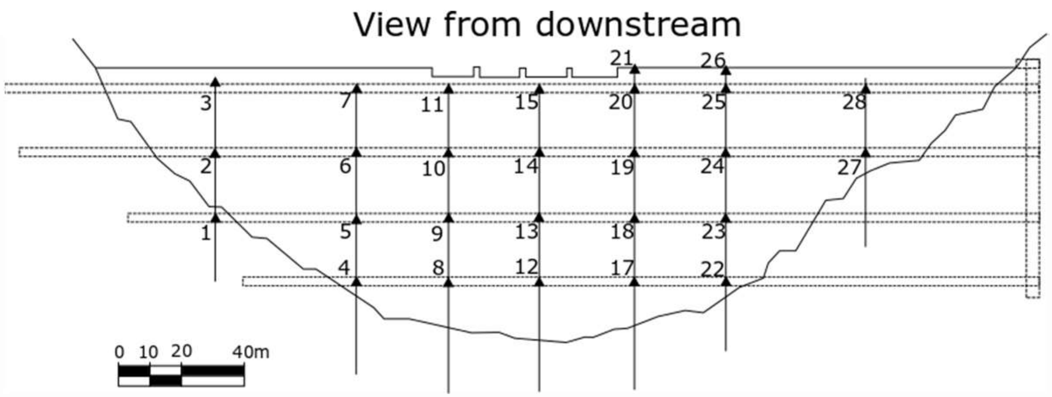


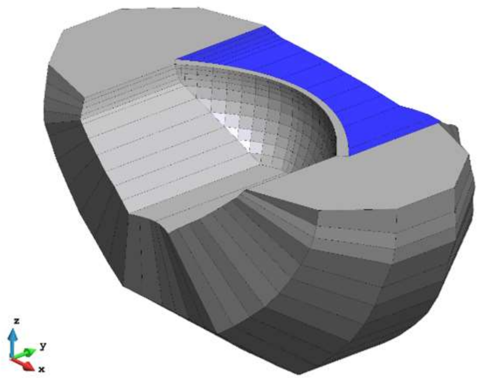


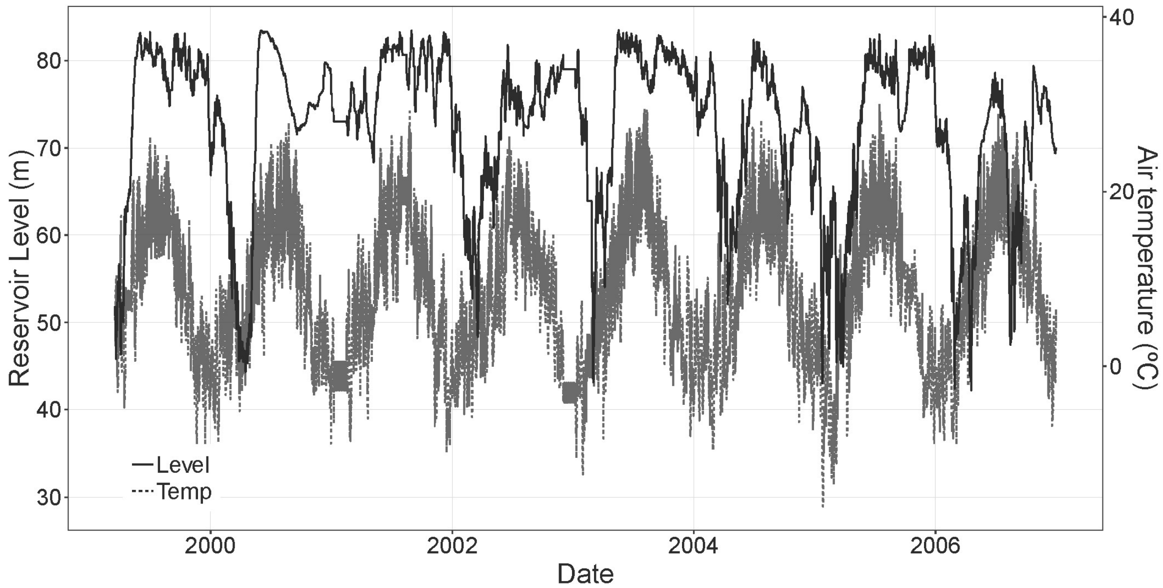
CIMNE

ICOLD-BW 2019
9th-11th September
MILANO
Numerical analysis of dams

2. METHODOLOGY

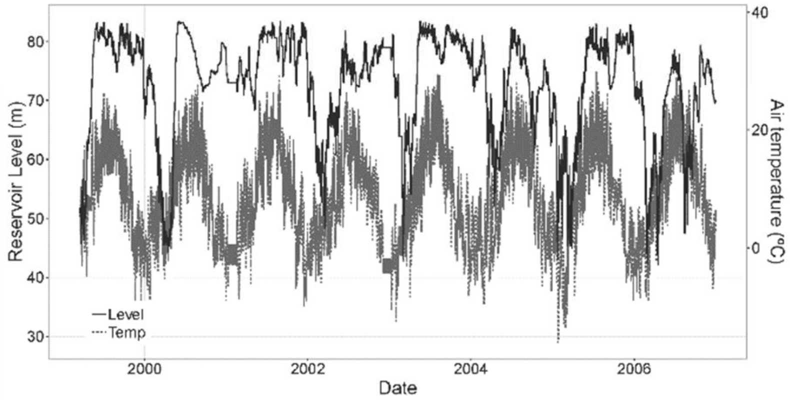


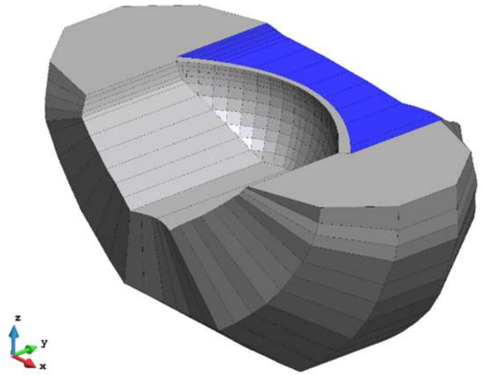


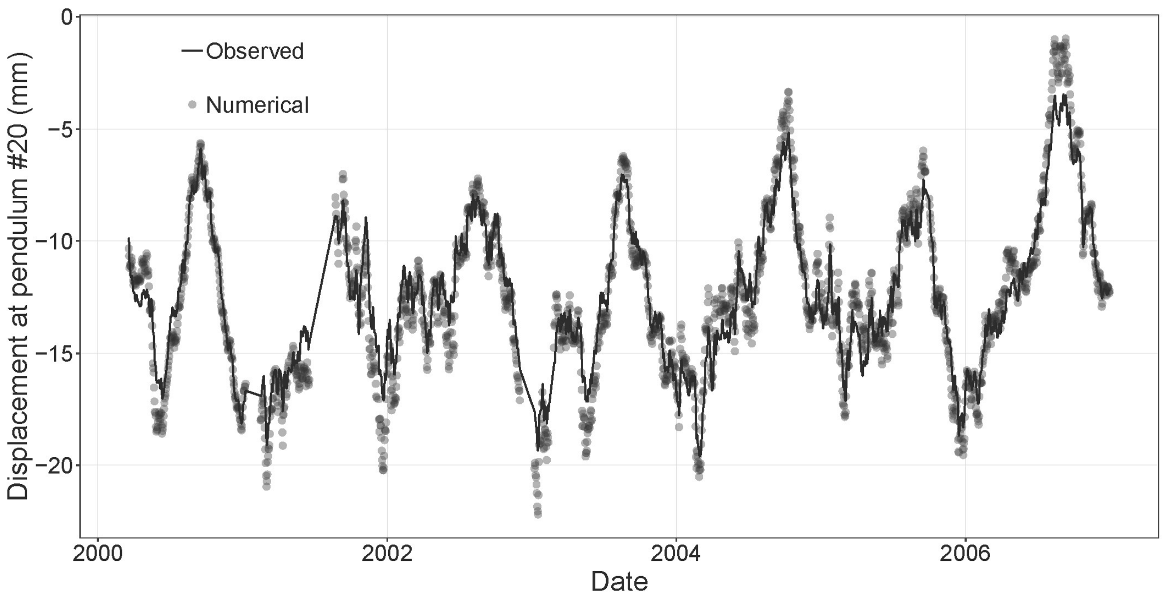


ICOLD-BW 2019
9th-11th September
MILANO
Numerical analysis of dams
CIMNE

2. METHODOLOGY





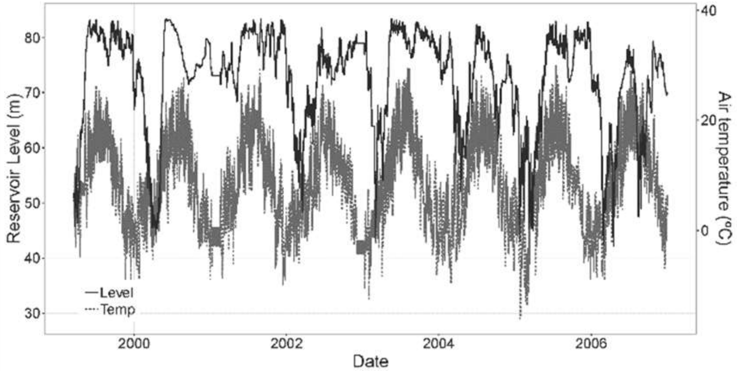


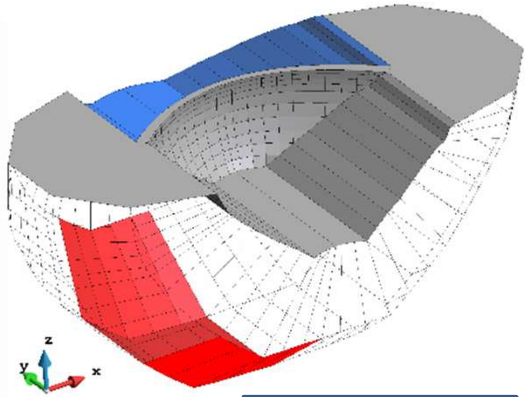
Calibration
Scenario 0
(reference)

CIMNE

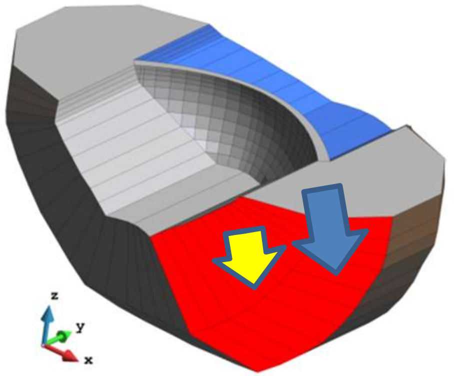
ICOLD-BW 2019
9th-11th September
MILANO
Numerical analysis of dams

2. METHODOLOGY



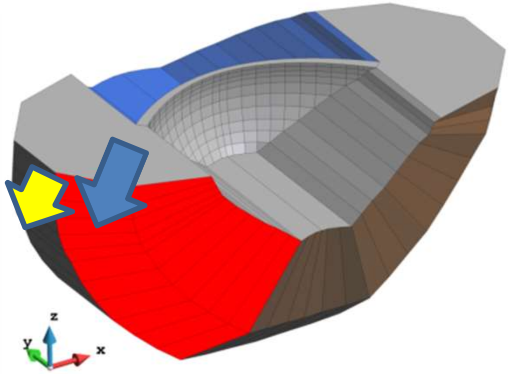


S5 – 1 mm



S1 – 1 mm

S2 – 0.5 mm



S3 – 1 mm

S4 – 0.5 mm

5 Anomalies

CIMNE

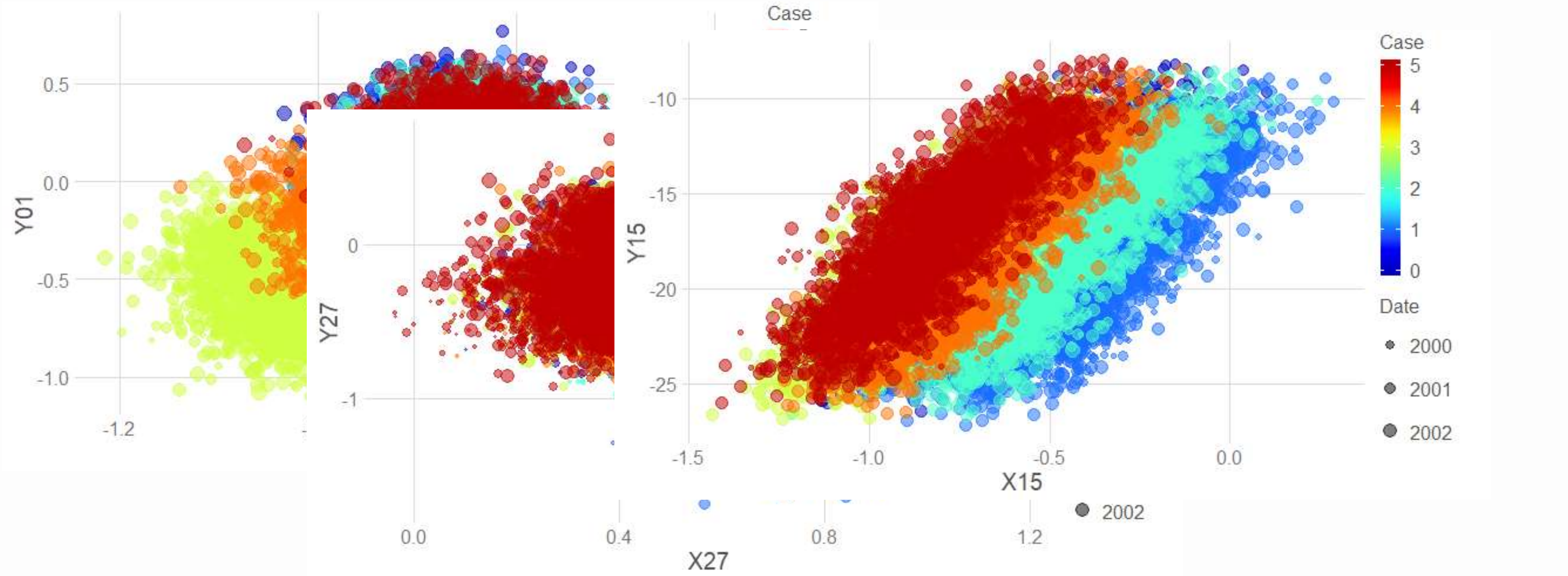
Numerical analysis of dams

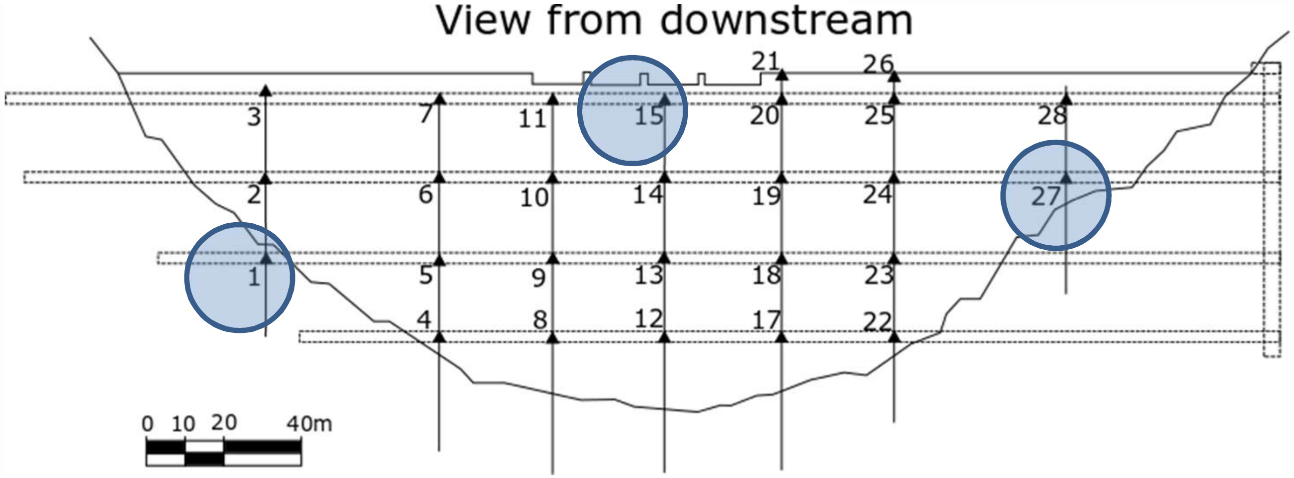
ICOLD-BW 2019

9th-11th September

MILANO

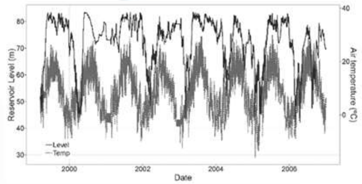
2. METHODOLOGY

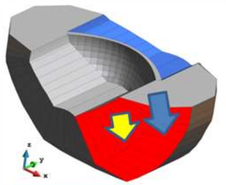




CIMNE

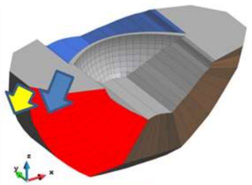
2. METHODOLOGY





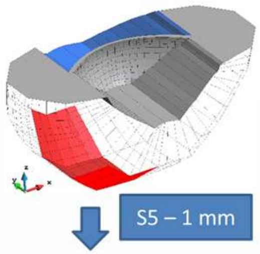
S1 – 1 mm

S2 – 0.5 mm



S3 – 1 mm

S4 – 0.5 mm



S5 – 1 mm

5 Anomalies

Noise added

58 inputs

1 output

Level	Temperature	Disp. Y (28)	Disp. X (28)	Label (scenario 0-5)

8 years
2 rows/day
6 scenarios
34,152 rows

2. METHODOLOGY. The algorithm

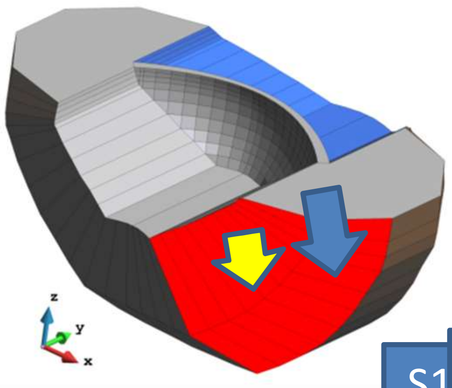
Random Forests

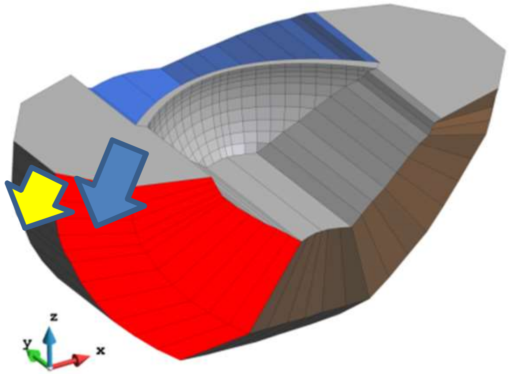
- Automatic input selection
- Good performance when many, highly correlated inputs
- Robust w.r.t. model parameters and training
- Automatic classification of inputs

3. RESULTS

Table 4: Example of confusion matrix. Model C (all inputs and training period 1999-2002)

		Actual scenario					
		0	1	2	3	4	5
Predicted scenario	0	2898	0	42	0	97	67
	1	0	2781	8	0	0	0
	2	8	141	2872	0	0	0
	3	0	0	0	2789	4	0
	4	14	0	0	133	2818	20
	5	2	0	0	0	3	2835





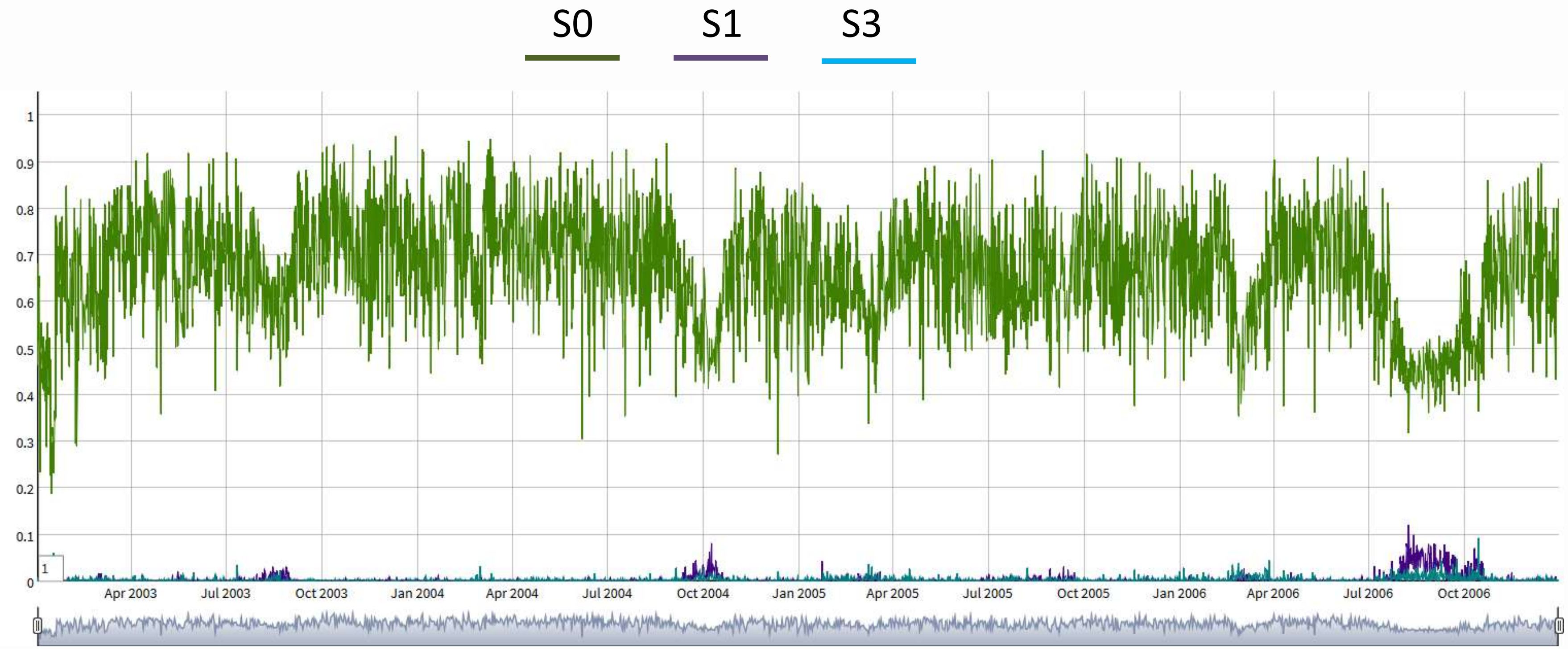
3. RESULTS

Table 4: Example of confusion matrix. Model C (all inputs and training period 1999-2002)

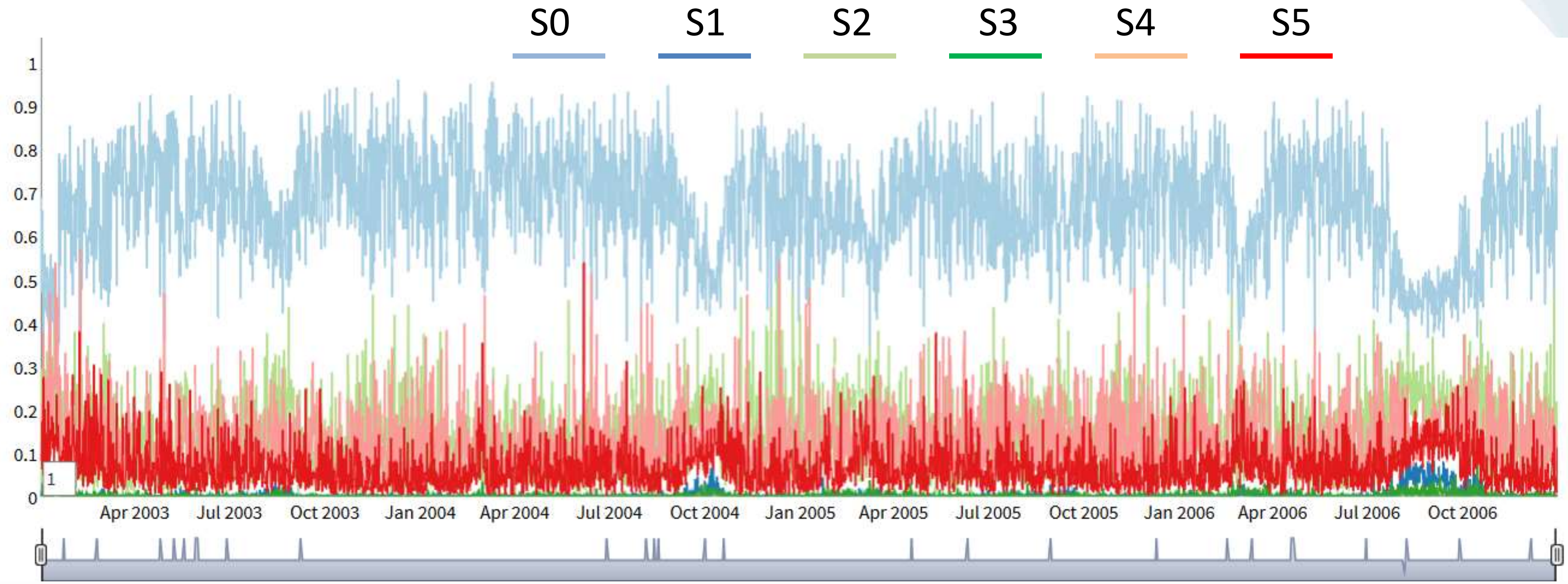
		Actual scenario					
		0	1	2	3	4	5
Predicted scenario	0	2898	0	42	0	97	67
	1	0	2781	8	0	0	0
	2	8	141	2872	0	0	0
	3	0	0	0	2789	4	0
	4	14	0	0	133	2818	20
	5	2	0	0	0	3	2835

3. RESULTS. Sc.0

Model output: Class probability [0,1]



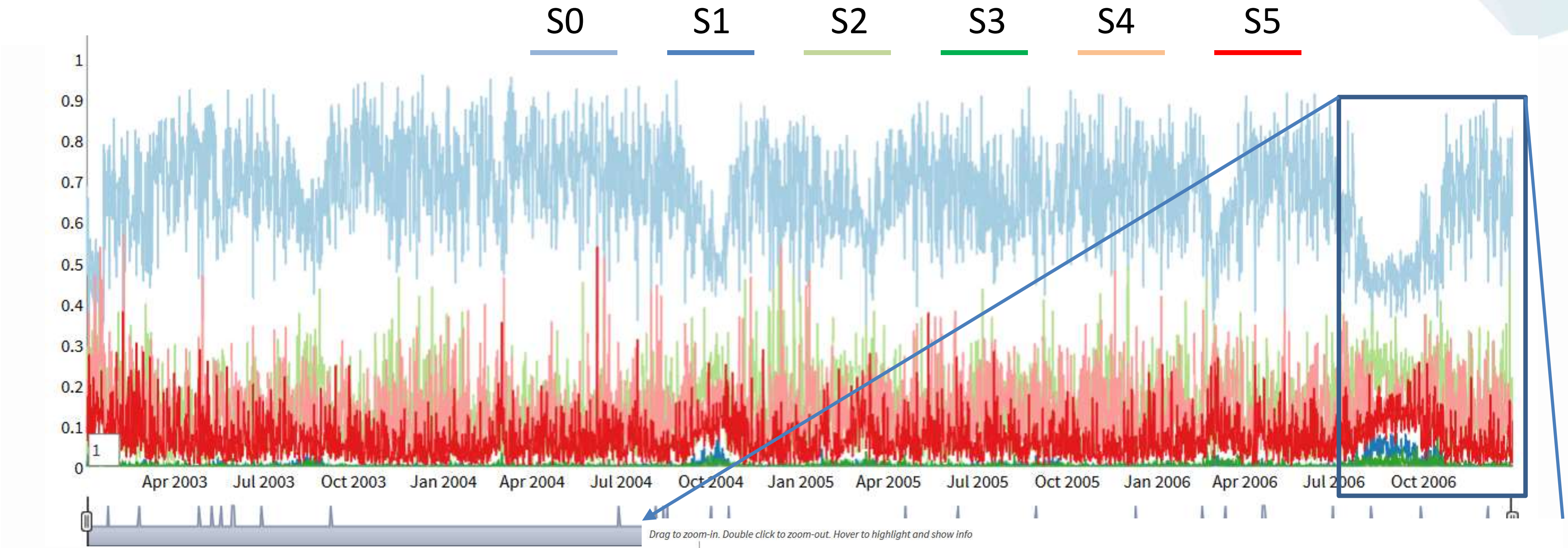
3. RESULTS. Sc.0

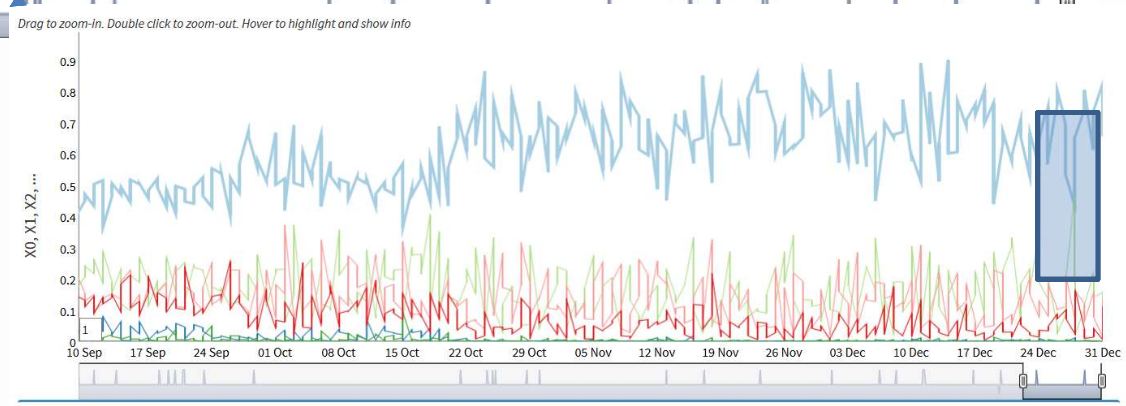


CIMNE

ICOLD-BW 2019
9th-11th September
MILANO
Numerical analysis of dams

3. RESULTS. Sc.0

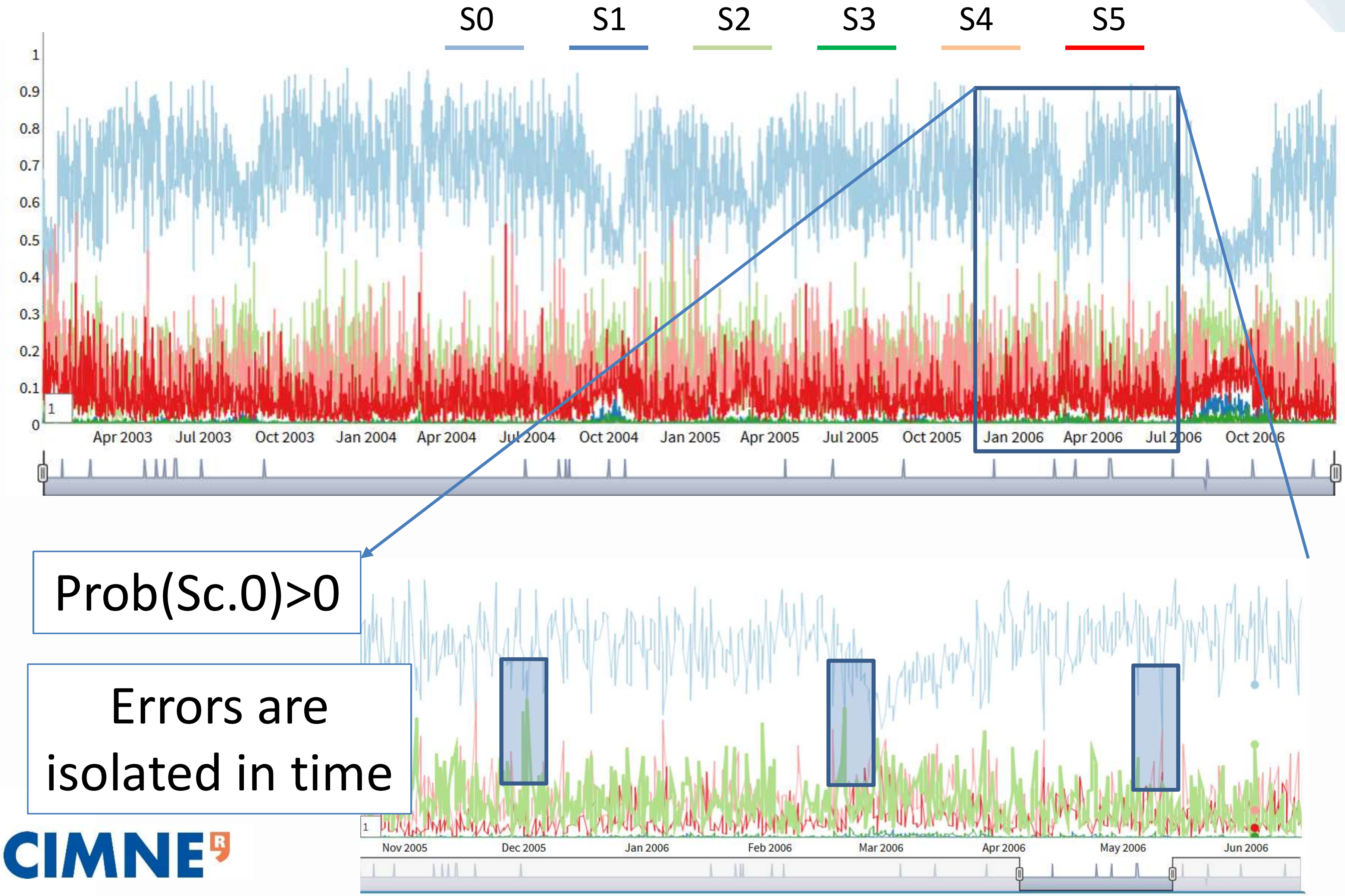




Prob(Sc.0)>0

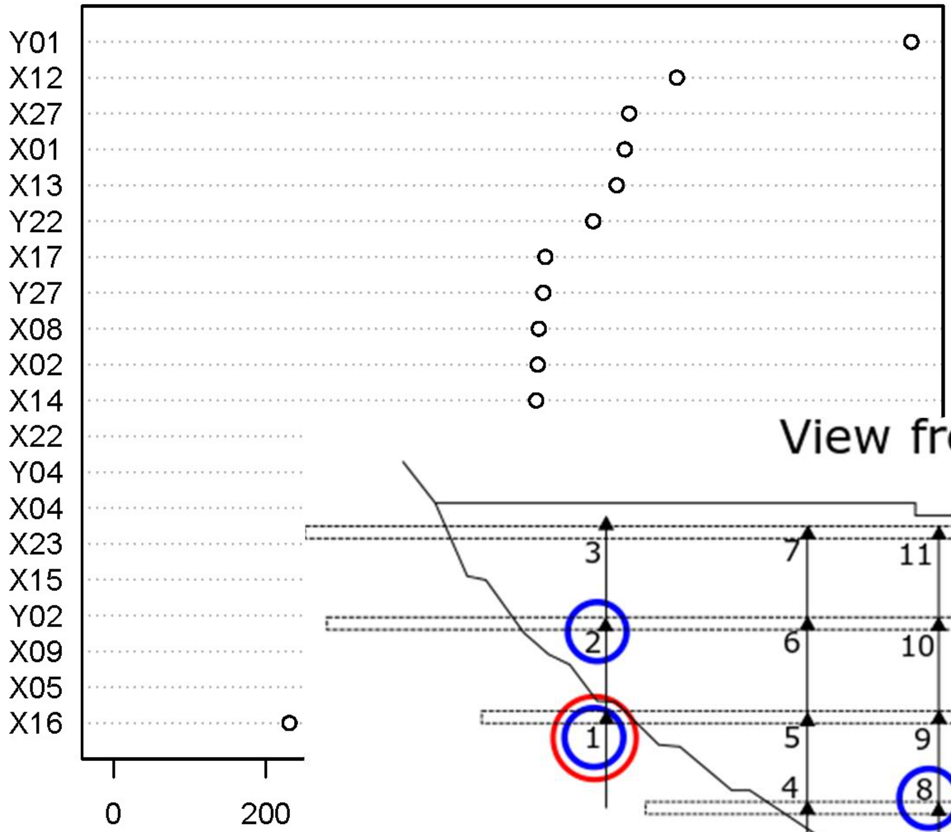
Errors are isolated in time

3. RESULTS. Sc.0

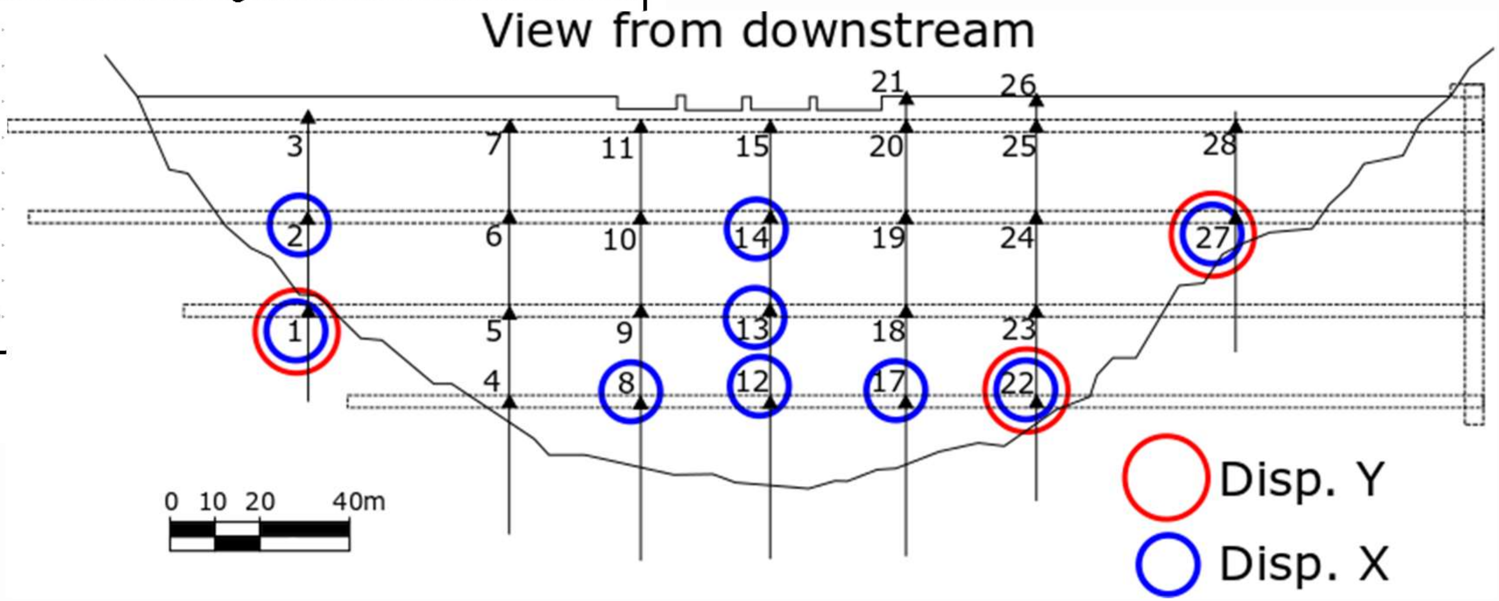


CIMNE

3. RESULTS. Variable importance



Top 12 variables



4. CONCLUSIONS

- Machine learning models can identify behaviour patterns jointly using a set of data series from multiple monitoring devices
- Anomalies need to be defined and susceptible to be reproduced with numerical models

Ongoing work

- Different dam typologies and kind of response variables
- Undefined anomalies

Thank you

F. Salazar, A. Conde, D.J. Vicente

fsalazar@cimne.upc.edu

aconde@cimne.upc.edu

djvicente@cimne.upc.edu