Course 3: Analysis methods of dam monitoring data

Cours n°3: Les méthodes d'analyse des mesures d'auscultation des barrages

Introduction to Machine Learning Methods

Introduction aux méthodes d'apprentissage automatique

Nathalie Rosin-Corre TRACTEBEL

Introduction to Machine Learning Methods

- Explanatory Data-based Models
- Multiple Linear Regression (HST) and its limitations
- Application of some Machine Learning methods to dam monitoring data analysis: Support Vector Regression (SVR), Random Forest (RF), Artificial Neural Network (ANN)

Explanatory Data-based Models

- Dam safety relies on the detection of irreversible changes in the dam behaviour possibly signalling a failure mechanism
- Explanatory data-based models can predict a response value Y based on explanatory variables $X_1, \ldots, X_p$:

$$Y = f\left(X_1, \ldots, X_p\right) + \varepsilon$$

  - Variables with reversible effects: reservoir level Z, seasonal angle θ (seasonal temperature), air temperature, rainfall…
  - Variable with irreversible effects: time t (ageing of dam)
  - Error ε : what is not taken into account by the model

Multiple Linear Regression (HST) and its limitations

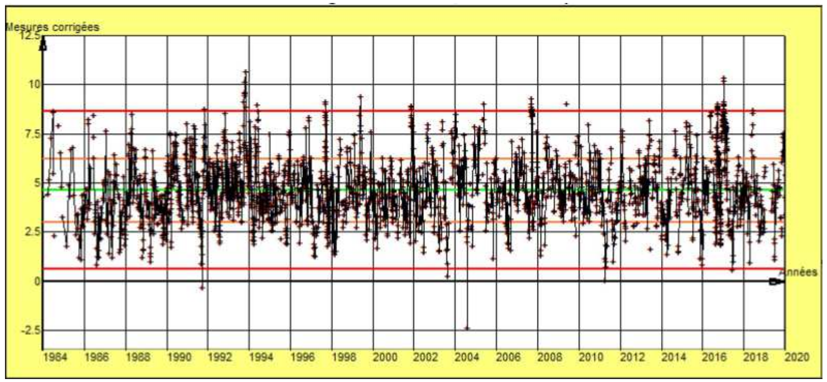
- Response of the dam assumed to be the sum of responses to each of explanatory variables:

$$Y = E_Z + E_\theta + E_t + \varepsilon$$

Where

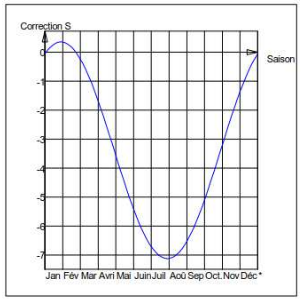
H effect $E_Z = a_1 Z + a_2 Z^2 + a_3 Z^3 + a_4 Z^4$
S effect $E_\theta = b_1 \cos\theta + b_2 \sin\theta + b_3 \cos 2\theta + b_4 \sin 2\theta$
T effect $E_t = c_1 t + c_1 e^{-t/t_0}$ (+step function)

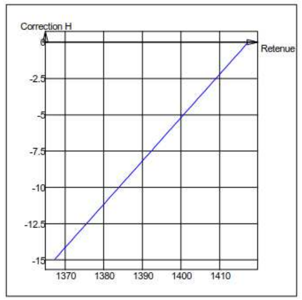
- Pros: analytical expression easy to handle
  diagnosis of monitoring data (normality thresholds)
- Cons: explanatory variables not independant
  non linear response



Corrected data (same H and S conditions) and normality thresholds

S effect

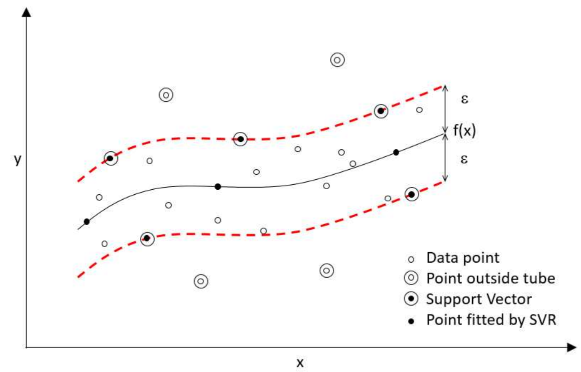




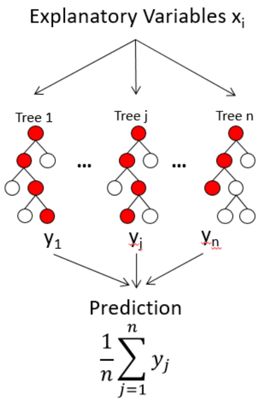
H effect

Application to dam monitoring data analysis

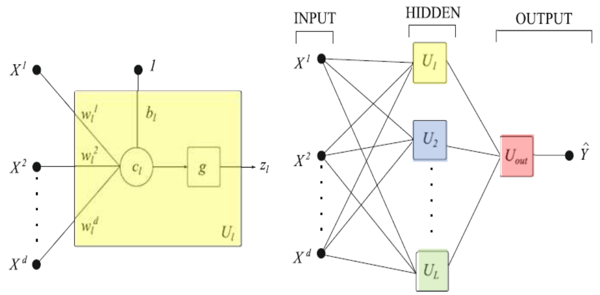
- Three Machine Learning methods:
  - Support Vector Regression (SVR)
  - Random Forests (RF)
  - Artificial Neural Network (ANN)
  - Limited number of hyperparameters
- Applications to dam monitoring data analysis:
  - Movement of a gravity dam
    (≈ 600 data over 48 years)
  - Leakage flow of an arch dam
    (≈ 2000 data over 40 years)
  - Piezometric levels in an embankment dam
    (≈ 500 data over 20 years)



SVR



RF



perceptron

artificial neuronal network

ANN

TRACTEBEL ENGIE

Cfbr comité français des barrages et réservoirs

Application to dam monitoring data analysis

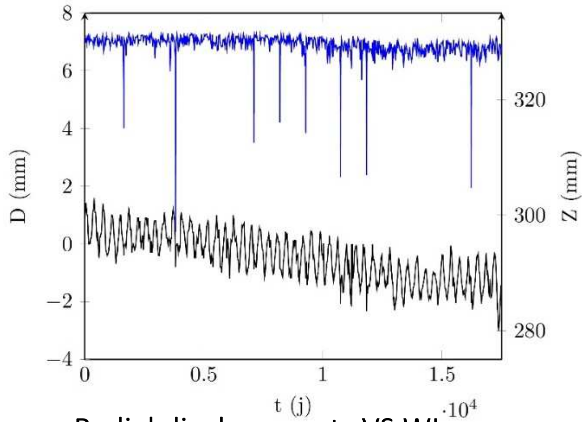
- Four input explanatory variables: reservoir level Z, seasonal angle ($\cos\theta$, $\sin\theta$), time t (same as HST models)
- Calibration and validation:
  - Database without considering the last three years
  - Calibration: 75% ; validation: 25% (random draw)
- Predictive capacity assessed over the last 3 years
- Coefficient of determination = part of the dispersion of the observed measurements explained by the model = statistical learning capacity of the model
- Comparison with HST models

Hyperparameters	Gravity dam movement	Arch dam leakage rates	Embankment dam piezometric levels
SVR			
K	radial	radial	radial
ε	0.1	0.1	0.1
C	3	3	2
RF			
n_a	100	100	80
n_v	2	2	2
ANN			
g	sigmoid	sigmoid	sigmoid
m	(4,4)	(6,6)	(4,8,4)
n_{rep}	6	5	10

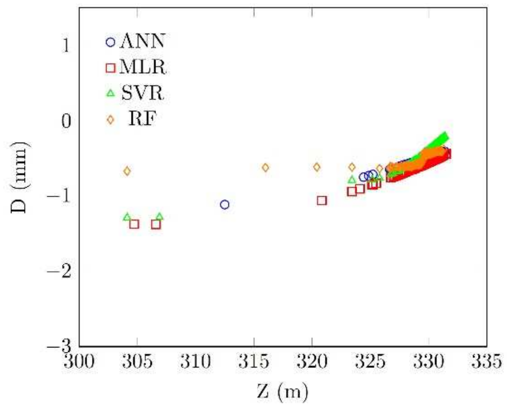
SVR : K kernel function, ε tolerance margin on regression error, C penalisation of error
RF: n_a number of trees, n_v number of variables drawn at each branch
ANN : g activation function, m number of perceptrons per layer, n_{rep} number of re-run (learning base quite limited)

Gravity Dam Movement (Pendulum)

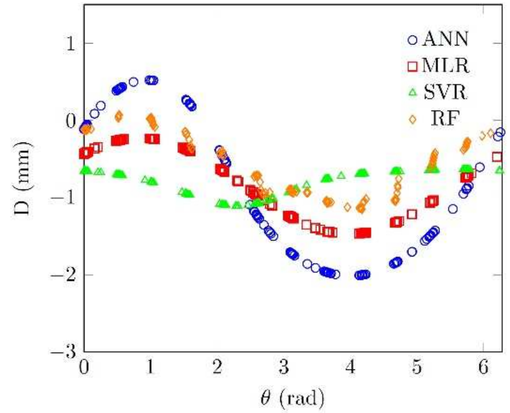
- Successive drawdowns with slight upward trend
- $R^2 \approx 0.9$ for all models
- Differences in Z, S and time effects (drawn with other explanatory variables at their median values):
  - Z effect with threshold near 325 for RF
  - S effect amplitude and phasing
  - time effect negligeable for SVR



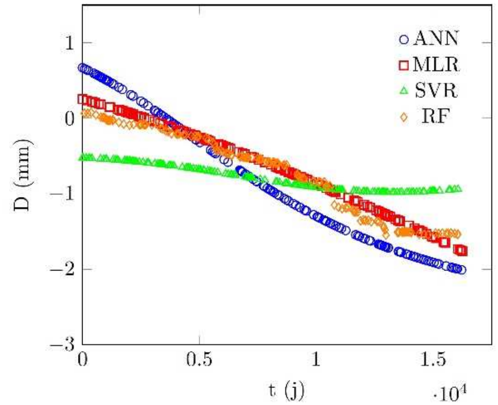
Radial displacements VS WL



Z effect (S, t constant)



S effect (Z, t constant)

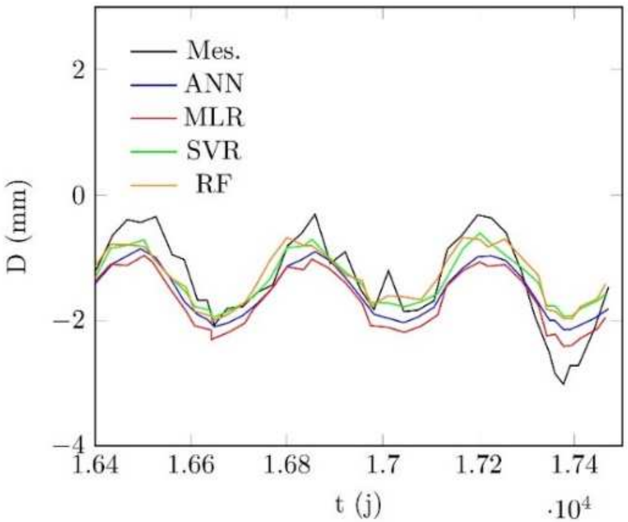


t effect (Z, S constant)

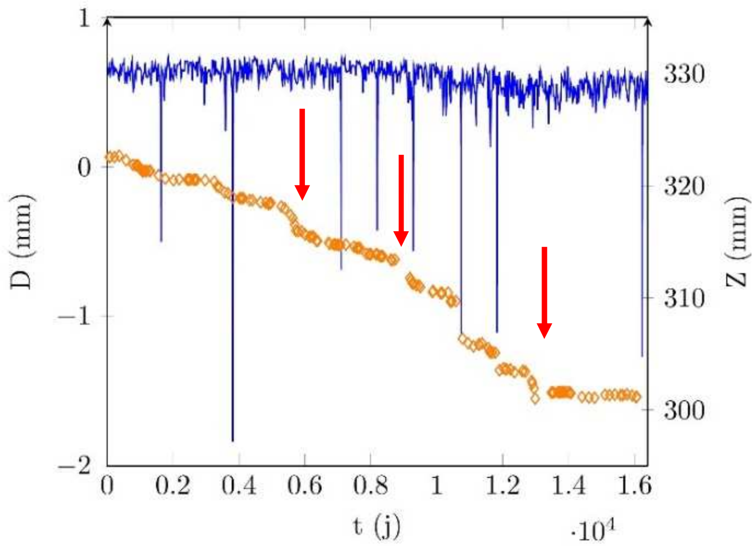
TRACTEBEL ENGIE

Cfbr comité français des barrages et réservoirs

Gravity Dam Movement (Pendulum)



Monitoring data and predicted values

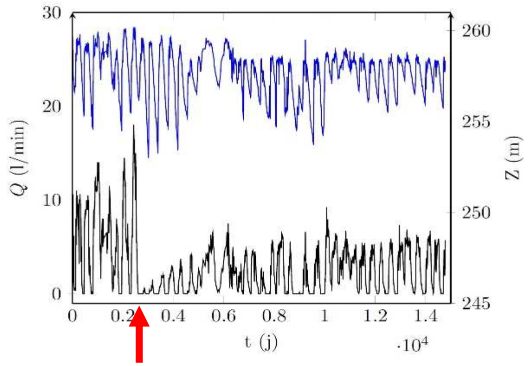


RF Time drift VS WL (red arrows = warmer summers)

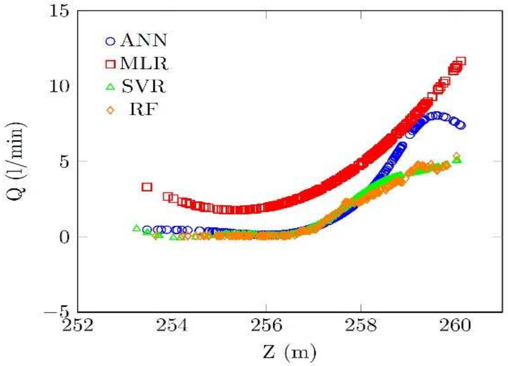
- Similar predicted behaviours for all models
- SVR and RF best-performing for prediction though amplitude of last year not replicated
- RF time drift with steps correlated to drawdowns and warmer summers

Arch Dam Leakage Rate

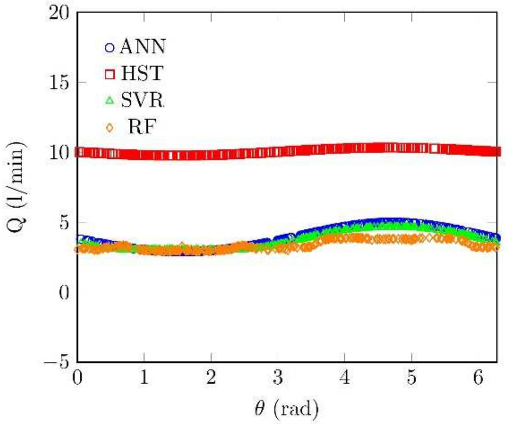
- Leakage of a lift joint, grouting at 2200 days, increase and stabilization
- $R^2 \approx 0.8$ (SVR, MLR) to 0.9 (ANN, RF)
- Z effect:
    - Threshold effect at the lift joint level
    - MLR unrealistic
- Time effect:
    - Negative values by ANN
    - No clear separation before/after grouting by MLR, SVR



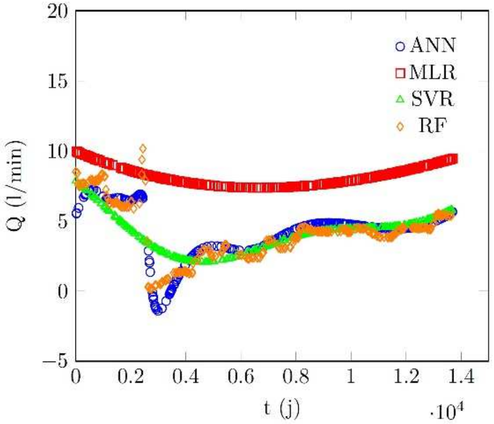
Leakage VS WL (red arrow = grouting)



Z effect (S, t constant)



S effect (Z, t constant)

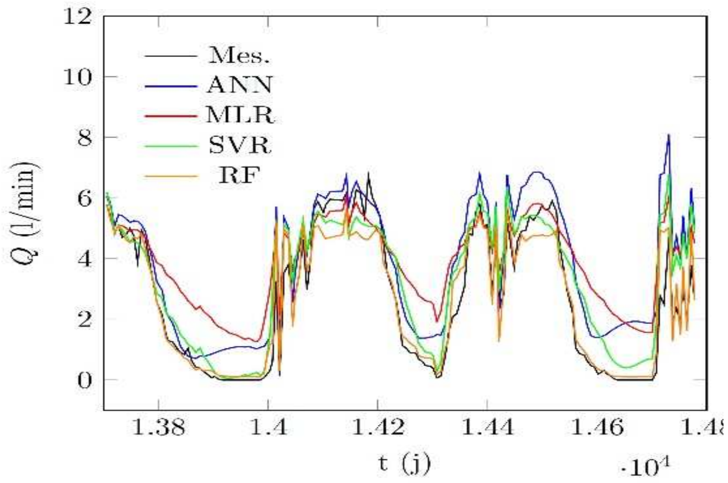


t effect (Z, S constant)

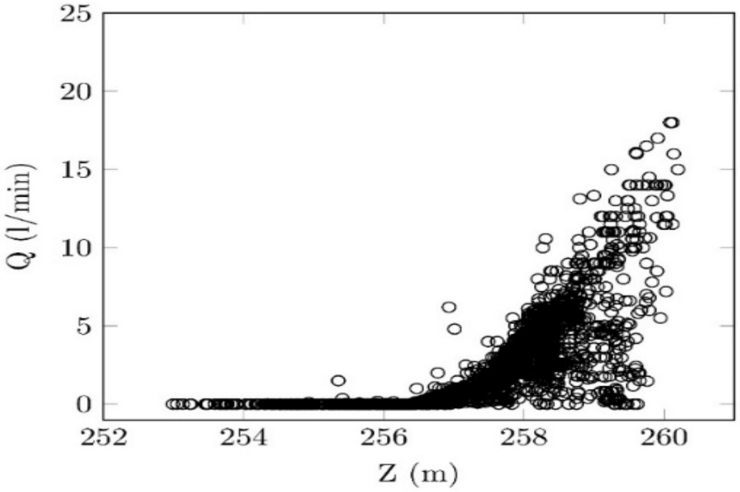
TRACTEBEL ENGIE

Cfbr comité français des barrages et réservoirs

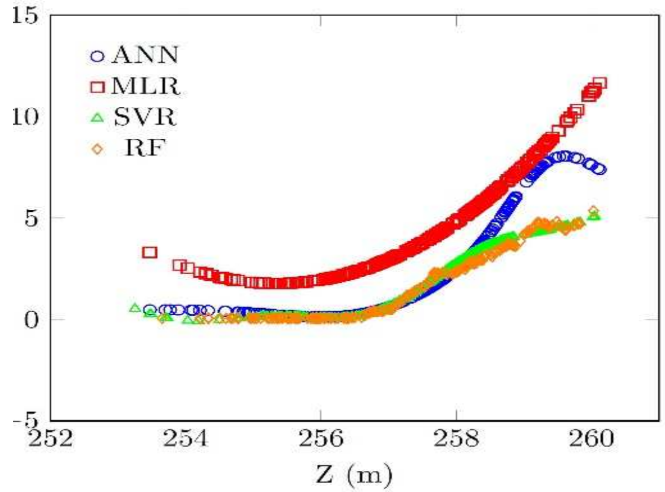
Arch Dam Leakage Rate



Monitoring data and predicted values



Monitoring data VS WL

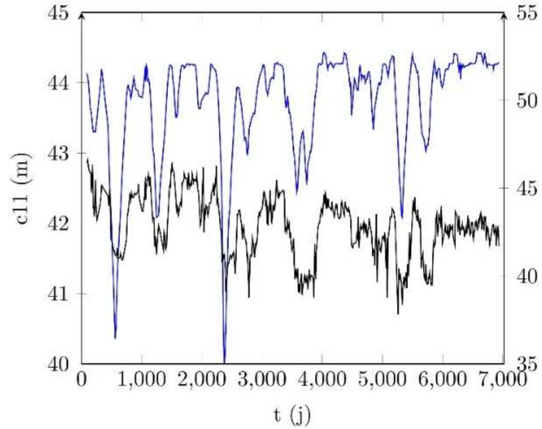


Z effect (S, t constant)

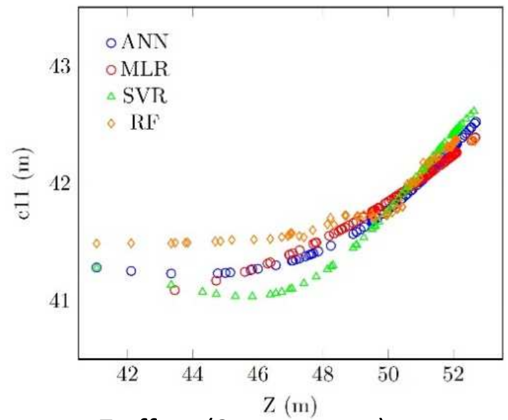
- RF model best-performing for prediction
- Closure of crack at higher WL by SVR, RF, ANN?

Embankment Dam Piezometric Levels

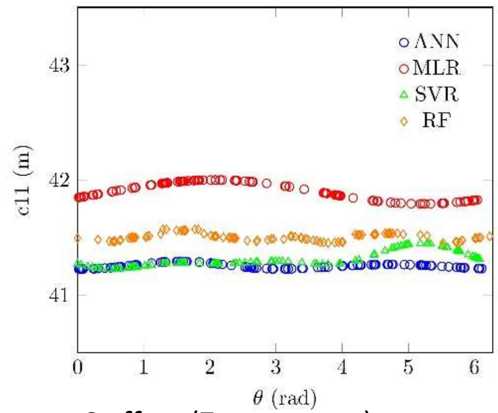
- VW cell near mid-height of core (D/S third)
- $R^2 \approx 0.8$ (MLR, ANN) to 0.9 (SVR, RF)
- Z effect:
    - Same for WL>50
    - Threshold effect near 46-48 (SVR, ANN, RF)?
- Low S effect for all methods
- Same trend for t effect
    - SVR unrealistic inflexion after 4500 days



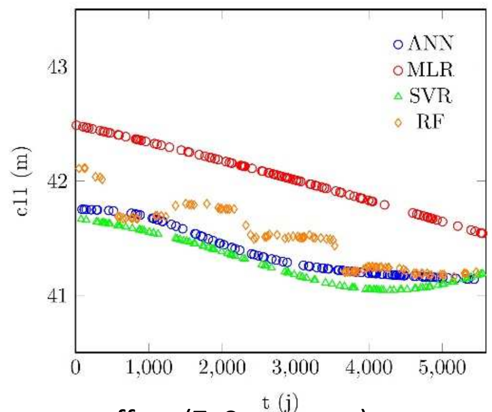
piezometric level VS WL



Z effect (S, t constant)



S effect (Z, t constant)

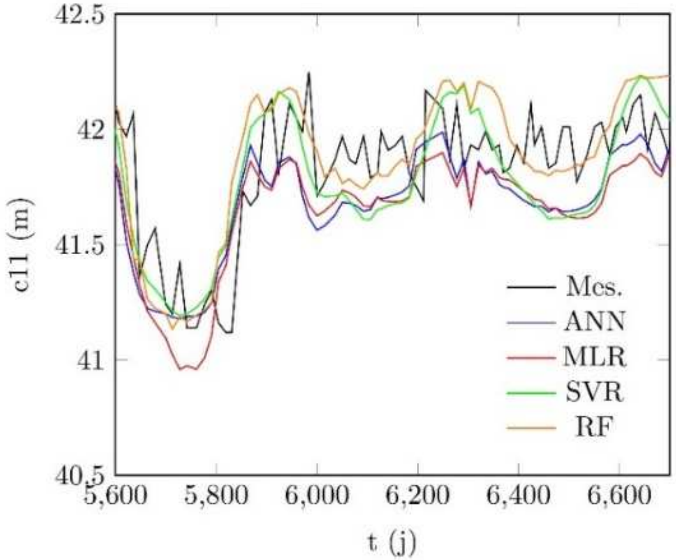


t effect (Z, S constant)

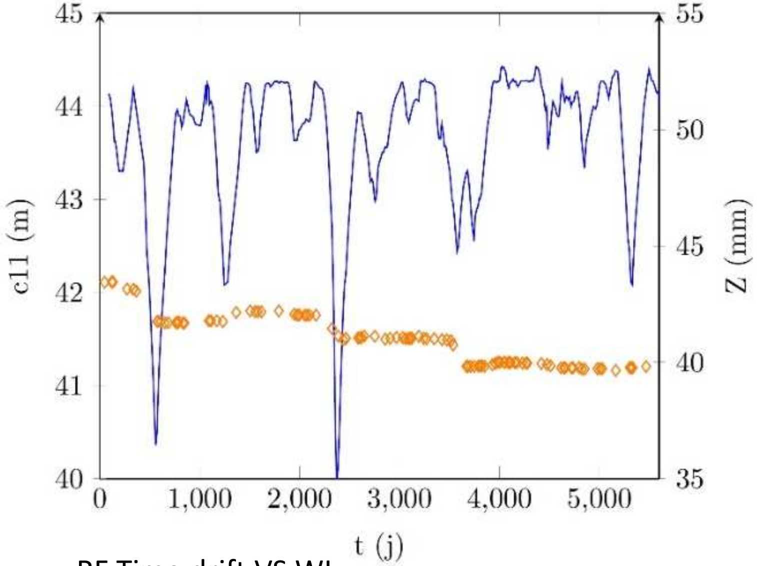
TRACTEBEL ENGIE

Cfbr comité français des barrages et réservoirs

Embankment Dam Piezometric Levels



Monitoring data and predicted values



RF Time drift VS WL

- RF model best-performing for prediction
- RF irreversible time drift correlated with significant drawdown?

Conclusions

- All data-based methods sensitive to data quality and sampling bias: measurement errors, too small populations, statistically non-homogeneous distributions
- Machine Learning methods can provide interesting additions to the analysis of dam monitoring data, especially when conventional linear regression models have low coefficients of determination
- Engineering judgment is needed to assess if use of these methods, requiring significant technical skills and expertise, is appropriate