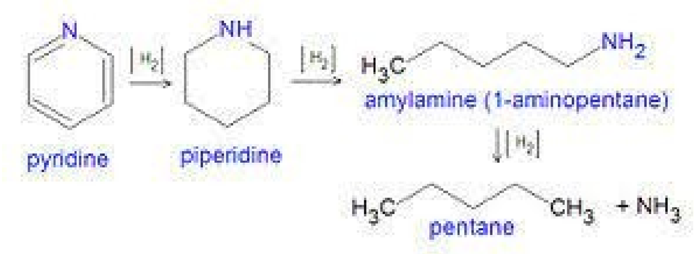
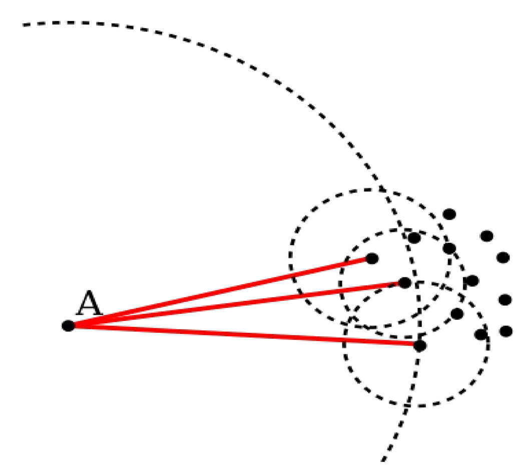
Institut Mines-Télécom

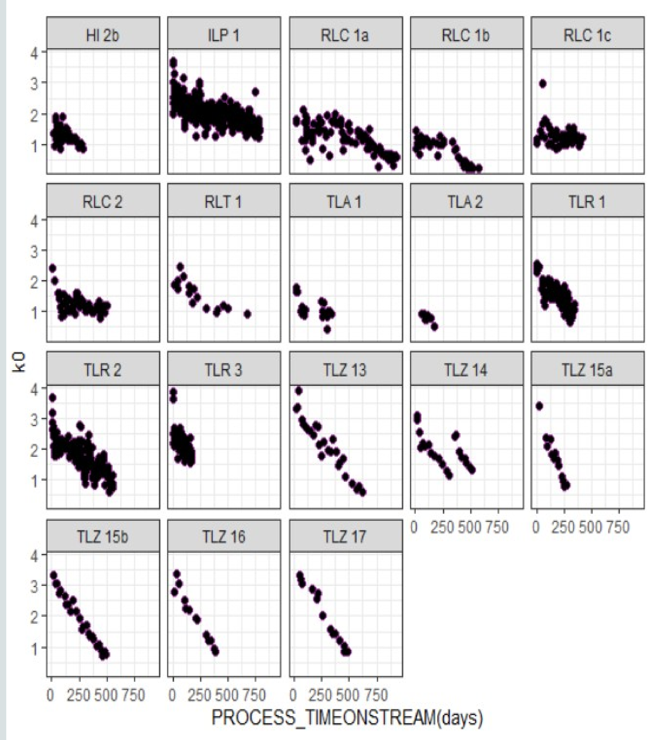
Estimation of a catalyst lifespan through machine-learning



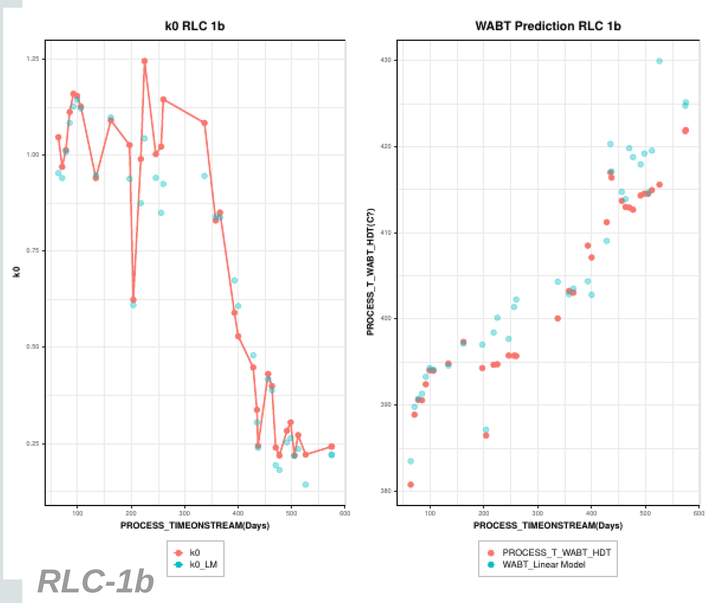
Examples of organic nitrogen compounds to be catalytically removed with hydrogen.



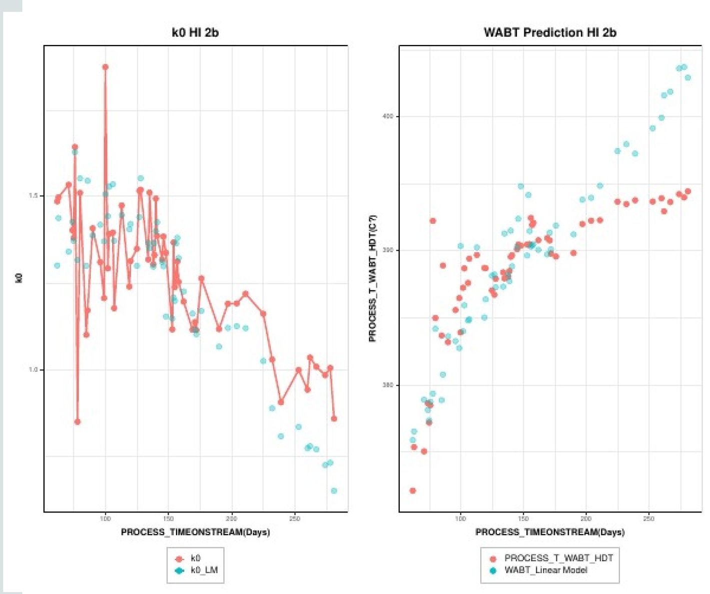
Basic idea of LOF: comparing the local density of a point with the local densities of its neighbors. A has a much lower density than its neighbors.



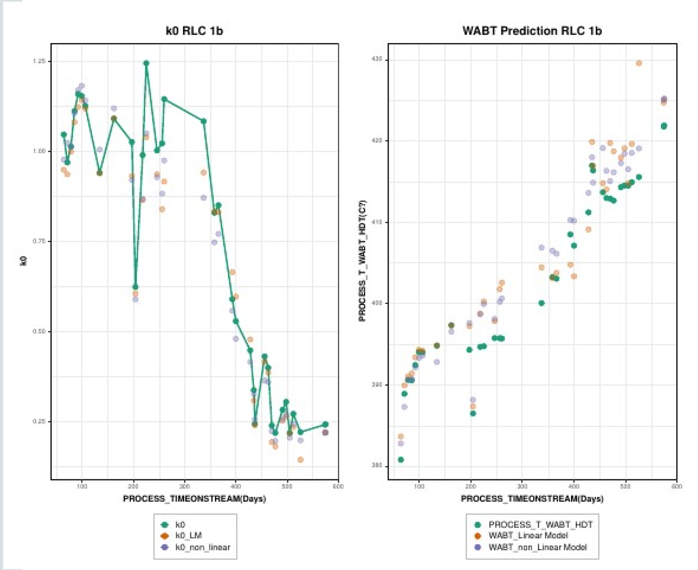
Experimental profiles of k0 for different catalysts as a function of process time..



RLC-1b



HI-2b



Comparison between linear and non-linear regressors

Partners

MINES Saint-Étienne — Une école de l'IMT

IFP Energies nouvelles

Auteurs

Marc Fischer*
Quentin Lemay*
Benoît Celse**

Partenaires

*Mines Saint-Etienne, Univ Lyon, CNRS, UMR 5307 LGF, Centre SPIN, F-42023 Saint-Etienne France,

**IFP Energies Nouvelles, Rond Point de l'échangeur de Solaize, BP 3, 69360 Solaize, France,

1. Introduction

- **Context:** catalytic hydrodenitrification of crude oil distillates.
- **Aging of the catalyst:** increase of the necessary reaction temperature until 450 °C is reached → replacement of the catalyst.
- **Goal:** predict the evolution of a catalytic reactor's temperature based on data from similar reactors (**machine learning**).

2. Pre-processing

- **Variable selection :** 12 most relevant variables selected based on expert knowledge: $t, p, p(H_2)_{out}, q_{feed}, q_{outlet}, reflux-ratio, p(N)_{out}, p(S)_{in}, p(resines), p(N)_{in}, density_{in}, p_{H_2,in}$
- **Missing values:** removal of physically aberrant and missing values.
- **Computation of cumulative variables :** $CUMSUM = \sum_{i=1}^{n-1} \frac{h_i}{2}(X_{i+1} + X_i)$.
- **Identification and removal of outliers:** use of the local outlier factor (LOF).
- **Remaining data:** 23 cycles and 1885 time points.

3. Methodology

- **Chemical kinetic model:** $$\frac{dC_N}{dt} = \frac{-k_0.exp\left(-\frac{E_a}{R}\left(\frac{1}{T}-\frac{1}{T_0}\right)\right)\left(\frac{ppH_2}{ppH_{2,ref}}\right)^m.C_N^n}{(1+A_0.Res_0)\left(1+C_0\frac{C_{N,0}}{C_{S,0}}\right)}$$
- **Prediction** of the catalyst efficiency represented by k_0.
- **Computation** of the experimental k_0 from the temperature measurements and computation of the predicted T from the predicted value of k_0.
- **Design mode**: we suppose that we know the profiles of all other variables than T and use them to predict $k_0(t)$ and in turn $T(t)$ at a given time t.
- **Multiple linear model**: based on 10 regressors, we seek to predict $\log(k_0)$.
- **Cross-validation:** the evolution of $T(t)$ for a cycle is predicted based on all other cycles.

4. Results

- **Global metrics:**

Mean RMSE	4.847
Standard deviation	1.774
Median	4.27

- **Satisfying average prediction error: ->** the average prediction error for T is 4.8 °C.
- Some cycles such as RLC-1b are well predicted from the beginning to the end.
- For others such as HI-2b there is a systematic gap towards the end.
- Linear relationships are not enough to fully grasp the system's behavior!

5. Addition of non-linear terms

- Addition of non-linear terms to the regressors : $X^2, log(X), \sqrt{X}, \frac{1}{X}, XY$
- We now want to select the 10 best regressors out of 552 potential regressors.
- Doing an exhaustive search would be **computationally prohibitive**.
- Use of an adaptive random search to find a good combination of regressors.
- A modest improvement could be reached:

Metrics	Linear model	Non-linear model
Mean RMSE	4.847	4.186
Standard deviation	1.774	1.817
Median	4.27	3.541

6. Conclusion and outlook

- While multiple linear regression is conceptually very simple, a reasonable agreement was reached.
- To better grasp the non-linear relationship, methods like neural networks and support vector regression could prove quite fruitful.
- It is also important to make sure that the machine learning method does not deliver **unphysical** predictions.
- Introduction of **weights** depending on the proximity of the cycle to the cycle to be predicted.
- **What if mode**: regression model based on both the other cycles and the beginning of the cycle → creation of a **digital twin**.

Contact : marc.fischer@emse.fr

Novembre 2015 Colloque de l'Institut Mines Télécom Paris