Stepwise Parameter Fitting to Combine Industrial and Pilot Plant Datasets.

Per Julian Beckera, Benoit Celsea

aIFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP 3, 69360 Solaize, France

[*per.becker@ifpen.fr*](mailto:per.becker@ifpen.fr)

Abstract

In kinetic model development for hydrotreatment processes industrial data are generally not used because deactivation must be taken into account, which is very difficult due to the high complexity of chemical phenomena. This is unfortunate because industrial data contains far larger feedstocks variation compared to pilot plant tests. The aim of this work is to propose an innovative method to include industrial data in kinetic model fitting. A stepwise parameter fitting method is proposed to use both pilot plant and industrial data. Pilot plant experiments provide robust data but with small feed variations, on the contrary industrial plant data provide huge feed variation. To obtain more robust models, a combined modeling framework for the kinetic reactor and deactivation models, solved simultaneously, is proposed for the HDN reaction in a hydrotreatment reactor. The kinetic parameters (reaction orders, activation energy) are calibrated on pilot plant points, while the empirical feedstock parameters as well as the deactivation model is calibrated on industrial points. This methodology leverages the strengths of each of the two datasets which results in more robust predictive models.

**Keywords**: Modeling, Catalyst Deactivation, Hydrotreatment, Industrial Data

* 1. Motivations & Objectives

Large datasets with thousands of points can be obtained from operating data of industrial units such as hydrotreatment units in fossil- and bio-refineries. This contrasts to the limited number of expensive and time-consuming pilot plant tests. Using industrial data for model fitting is highly desirable but remains challenging because (i) operating conditions generally remain close to the design conditions of a given unit and (ii) catalyst deactivation is very difficult to model due to the high complexity of the chemical phenomena. This leads to issues with the identifiability of kinetic parameters. Pilot plant data is, on the other hand, very well suited for calibration of kinetic models (operating condition variation) with carefully constructed designs of experiment covering a wide range of operating conditions and the absence of deactivation. However, feedstock variability is generally limited which leads to overfitting and consequently less robust models. The aim of this work is to propose an innovative method to include industrial data in kinetic model fitting. The idea is to combine the two datasets (pilot and industrial) in order to leverage their respective strengths. This work presents a methodology for combining industrial and pilot plant datasets.

The model presented in this work was implemented in Fortran and compiled to a shared library (dll) to be used with a high-level scripting language. Data pre-processing and parameter identification was done with R and post-processing was done with Python.

* 1. Process Description

Hydrotreatment and hydrocracking (HCK) is a very flexible process, which is extensively used in petroleum refining to convert the heavy Vacuum Gas Oil (VGO) fractions of crude oil into high-quality products (Becker 2016). Recently, there has been an increasing interest in this technology for upgrading and conversion of bio-sourced or recycled plastic feeds. A typical HCK unit consists of two fixed-bed reactors in series, under high hydrogen pressure (up to 160 bar), and operating at temperatures between 360 and 430°C. The catalyst of the first reactor is designed for removal of organic impurities via hydro-denitrogenation (HDN), hydro-desulfurization (HDS), or hydro-deoxygenation (HDO) reactions. The second reactor contains a zeolite catalyst, which performs the actual hydrocracking, i.e., breaking up of long-chained hydrocarbons, isomerization, and hydrogenation of aromatic rings. The reactors are followed by a distillation column.

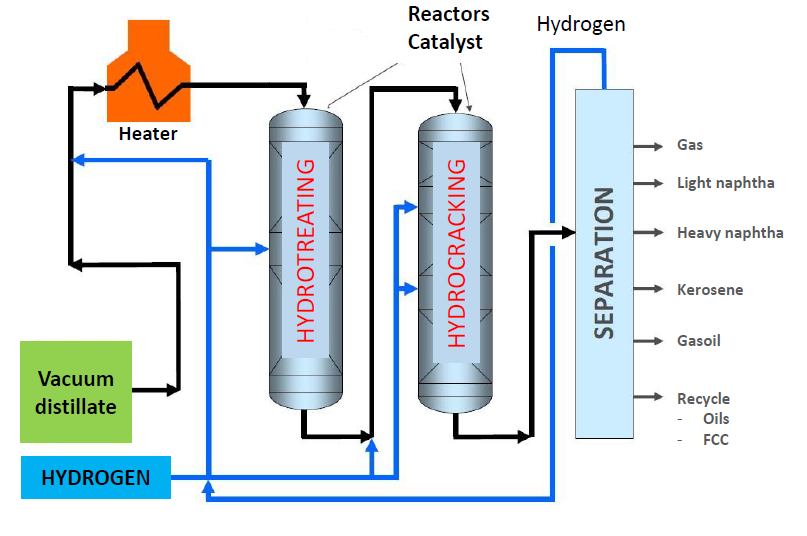


Figure 1 Schematic representation of the Hydrocracking process

The focus of this work is the HDN reaction in the hydrotreatment reactor. Removal of organic nitrogen impurities is important because they act as inhibitors on the zeolite catalysts in the second reactor. The design and operation of the hydrotreatment reactor is based on the required concentration of organic nitrogen (nitrogen slip) at reactor outlet, typically around 10 – 50 ppm. The reactor temperature is used as control variable during operation. A predictive model for the HDN reaction is required at the design phase and to be able to simulate the impact of eventual changes during operation of the unit.

* 1. Catalyst Deactivation

Hydrotreatment and hydrocracking catalysts deactivate over time (Forzatti 1999), principally due to coke formation, but occasionally also due to presence of heavy metals in the feed. The loss in activity is compensated by increasing reactor temperature. Once reactor temperature reaches the design limitation of the unit (typically < 430°C) the units must be shut down and the catalyst changed. Typical cycle duration is between 2 and 4 years, depending on feedstock characteristics and operating conditions. Shutdowns in large integrated refineries are very costly and need to be planned well in advance. A predictive deactivation model is necessary to correctly predict the required increase of temperature over time and subsequently the cycle lengths. This deactivation model must be combined with a kinetic model to de-correlate instantaneous changes.

* 1. Model Development
     1. Kinetic Model for HDN Reaction

The principal difficulty of hydrotreatment reactor modeling is the immense complexity of the system, with typical VGO feeds composed of hundreds of thousands of individual hydrocarbon species (Becker 2016, Chehadeh 2023). Detailed characterization of the chemical composition of feeds and effluents is not feasible with current analytical methods. While some purely data-driven approaches have recently been proposed (Pang 2024) for simulation of hydrocracking units, such methods are generally only applicable for a single refinery. These methods are therefore useful for process control but have poor predictive capabilities. The kinetic model proposed here considers organic nitrogen as a single chemical species and combines well-known reaction kinetics with an empirical term taking macroscopic, easily measured, feed characteristics into account (see Table 1). The kinetic model for the HDN reaction (1) is composed of the kinetic equation for an irreversible reaction with kinetic parameters combined with an empirical term, *gcorr*, which is used to model the impact of feedstock descriptors (*Xfeed*) on reactivity. In this work an explicit form of *gcorr* is used (2), however, it is possible to replace this term with a more complex data-driven model.

|  |  |
| --- | --- |
|  | (1) |
|  | (2) |

Table 1 Definition of descriptors and parameters used in the kinetic model.

|  |  |  |
| --- | --- | --- |
| *N* | Nitrogen concentration [ppm] |  |
| *T* | Temperature (WABT) [K] | ref = 648.15 K (375°C) |
| *t* | Time dimension, contact time [s] |  |
| *ppH2* | Hydrogen partial pressure [bar] | ref = 128.21 bar |
| *Xfeed* | Feedstock descriptors |  |
| *Res* | Feed resins [%] |  |
| *TMP* | Weighted mean boiling point of feed [°C]  = (DS5 + 2\*DS50 + 4\*DS90)/7 | ref = 479.2°C |
| *gcorr* | Correction term for feed descriptors |  |
| *d154* | Feed density [g/cm3] | ref = 0.9175 g/cm3 |
| *KWatson* | aka KUOP [-] | ref = 11.687 |
| *A0, v, d, k* | Feed Parameters: *βfeed* |  |
| *k0, Ea, n, m* | Kinetic Parameters: *βkin* |  |

* + 1. Deactivation Model

Catalyst deactivation is principally due to the formation of coke which accumulates on the catalyst particles and blocks the active sites, thus reducing activity over time (Rodriguez 2018). It is well known that coke formation increases with temperature and tends to be slowed down by increasing hydrogen partial pressure and/or hydrogen-to-hydrocarbon flow ratio (Rodriguez 2018). Feed asphaltene and carbon conradson content of the feedstock are good indicators for the tendency of increased coke formation. A detailed kinetic model for coke formation could not be developed because of the absence of relevant measurements. An semi-empirical deactivation model (3) and (4) was therefore developed, based on the considerations outlined above. Table 2 details the descriptors and models parameters.

|  |  |
| --- | --- |
|  | (3) |
|  | (4) |

Table 2 Definition of descriptors and parameters used in the deactivation model.

|  |  |  |
| --- | --- | --- |
| *T* | Temperature WABT\* [°C] |  |
| *t* | Time on Stream [days] |  |
| *CC* | Feed Carbon Conradson [%] | lower limit = 0.8 % |
| *AS* | Feed Asphaltenes [ppm] | lower limit = 10 ppm |
| *ppH2* | Hydrogen partial pressure [bar] | std. value = *f(KWatson, TMP)* |
| *H2HC* | H2 gas to hydrocarbon ratio [L/L] | std. value = 5*\*ppH2std + 100* |
| *a, b, αCC, βCC, αAS, βAS, a0, βppH2, βH2HC* | Deactivation parameters: *βdeac* |  |

\*WABT = weight averaged bed temperature

* 1. Datasets

Industrial data was obtained from a total 28 cycles from 19 hydrotreatment units, in addition to 117 pilot plant points performed at IFPEN. The principal differences, with respect to the development of a combined kinetic and deactivation model are summarized in Table 3.

Table 3 Comparison of industrial and pilot plant datasets

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | data points | Feed variability | Operating Conditions | βkin identifiability | βfeed identifiability | Catalyst De-activation |
| Industrial | Large 10,343 | High, daily variations | Fixed set-point per cycle | Poor | Good | Yes |
| Pilot Plant | Small 117 | Low, discrete feeds | Varied according to DOE | Good | Poor | No |

Pilot plant tests are done on relatively short runs of up to 1.5 months, deactivation is therefore negligible compared to long industrial cycles, the deactivation model cannot be calibrated on this data. Due to careful experimental designs these points are, however, well suited for identification of the kinetic parameters. Data from industrial cycles, on the other hand, allow for the deactivation effect to be observed, but identification of the kinetic parameters is problematic. This is mainly because the nitrogen content in the effluents vary generally very little from the set-point of a particular unit, while temperature is adjusted to correct for loss in catalyst activity, this makes it very difficult to identify the correct reaction order and activation energy. Pilot plant test runs are performed with a limited number of available feedstocks, resulting in a small number (20) of discrete points, which often leads to over-fitting of the *βfeed* parameters in the empirical term of the kinetic model. The feedstock characteristics of the industrial dataset varies at an almost daily rate. This is shown by comparing TMP and d154 in Figure 2. The pilot plant feeds generally fall within the range covered by the industrial data, except for heavier HCGO (feeds 1 and 16), higher boiling point DAO (feeds 13 and 17), and the light feed 15. The feed of plant R cycle 2 is atypical due to higher boiling point than for the other industrial cycles.

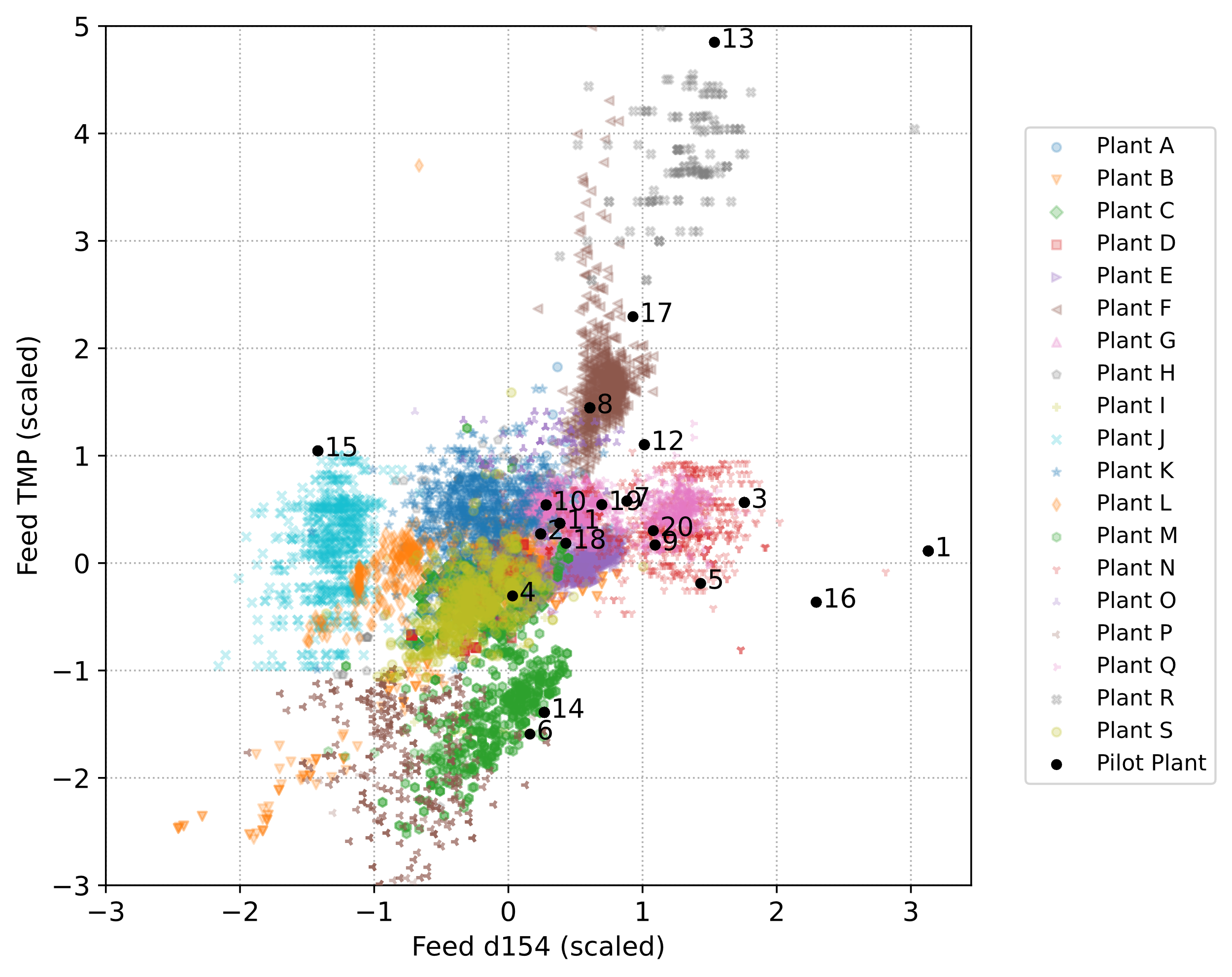


Figure 2 Comparison of feed density and TMP for pilot plant and industrial datasets.

* 1. Parameter Fitting Algorithm

The combined parameter fitting algorithm consists of two steps: 1) Fit *βfeed* and *βkin* on the pilot plant dataset, and 2) fit *βfeed* and *βdeac* on the industrial dataset keeping at *βkin* the value from 1). In the first step the full kinetic model is calibrated only on pilot plant data, this ensures good estimation of the kinetic parameters *βkin*, while *βfeed* might be poorly estimated or over-fitted. In the second step the empirical feed parameters *βfeed* and the deactivation model *βdeac* are calibrated on the industrial dataset. This step improves the robustness of the empirical term in the kinetic model without compromising the good estimation of *βkin* from step 1). The “optim” function of R with the “Nelder-Mead” algorithm was used for parameter fitting.

* 1. Results & Discussion

Results for the combined kinetic/deactivation model with the stepwise parameter fitting procedure (case 2) are shown in Figure 3 with results obtained by either fitting the kinetic model exclusively on pilot plant (case 1) or industrial data (case 3). The target value is an RMSE of ±5°C. Atypical pilot plant feeds are marked in red. The aim here is not to compare the datasets but to show that using both industrial and pilot plant data leads to a more robust model with respect to feed variation thanks to the industrial data, and more robust with respect to operating condition (T, P, LHSV), thanks to the pilot plant data.

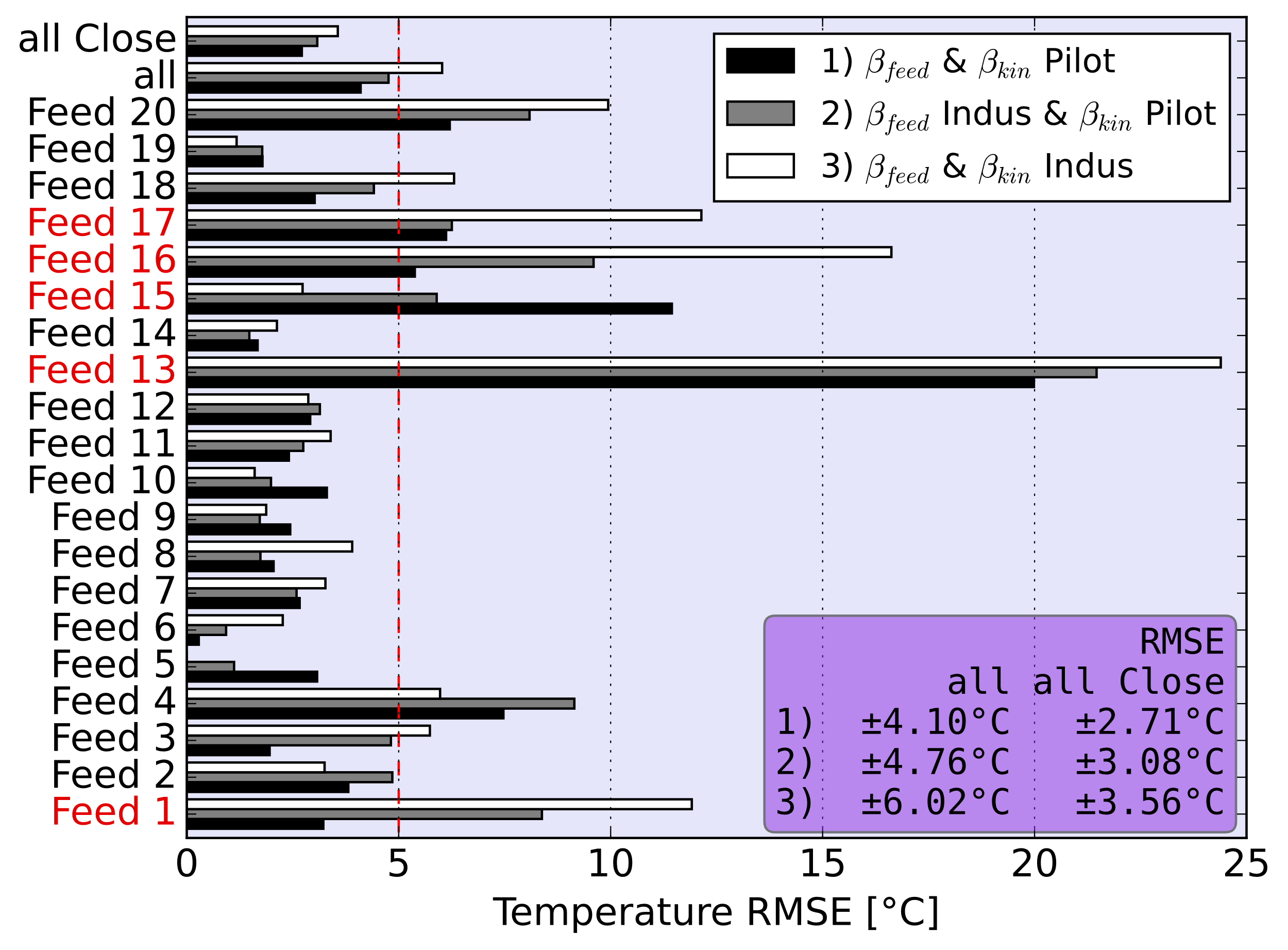
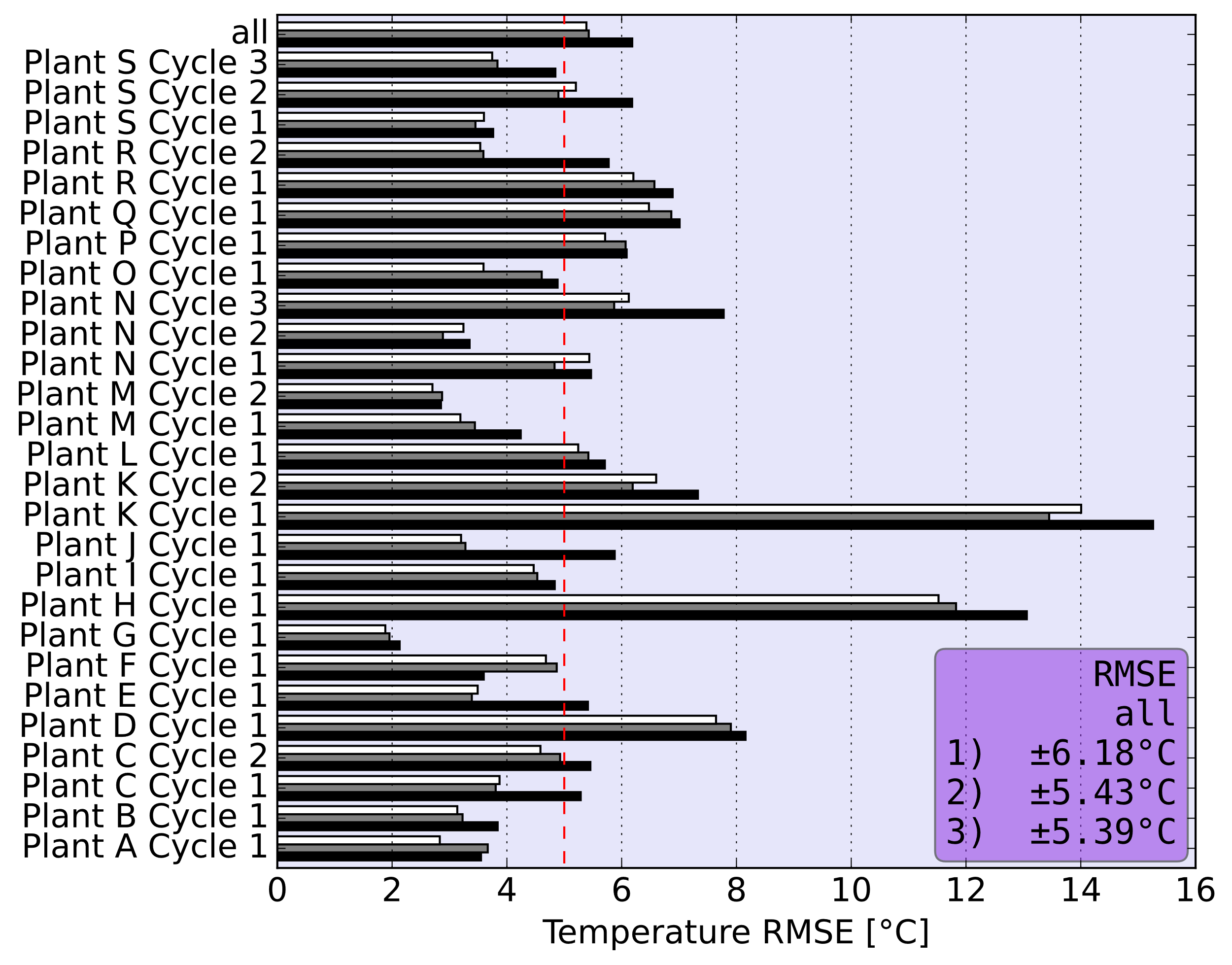


Figure 3 Errors for industrial (left) by cycle and pilot plant (right) by feed.

For the industrial dataset the model error improves by 0.75°C between case 1) and case 2), while the error of the pilot plant dataset is degraded by 0.66°C (0.36°C excluding atypical feeds). No further improvement can be observed when also fitting all parameters of the kinetic model on industrial data (case 3), while the pilot plant error is further degraded by 1.26°C (0.49°C excluding atypical feeds). This suggests that fitting *βkin* on industrial data leads to over-fitting and these parameters should be identified on pilot plant data. Concerning *βfeed*, the degradation of the pilot plant error is well within the acceptable range, except for the atypical feeds, while significantly improving the error of the industrial data. This method presents a good compromise, for combining the strengths of the two datasets.

* 1. Conclusions & Perspectives

The proposed parameter fitting algorithm allows both pilot plant and industrial data to be used to fit the parameters of a combined kinetic and deactivation model. Using both datasets for parameter fitting was found to improve model performance compared to the case where only one of the two datasets is used. This is an effective method for combining two very different datasets with different strengths and weaknesses in terms of parameter identifiability. An explicit term was used for the empirical term in the kinetic model, future work includes the implement different data-driven Machine Learning (ML) models for this term. The parameter fitting procedure presented here allows such a hybrid model to leverage the larger industrial dataset for the ML term which generally requires larger volumes of data.

References:

P. Forzatti, L. Lietti, 1999, *Catalyst deactivation*, Catalysis Today. 52, 165-181.

D. Chehadeh, X. Ma, H. Al Bazzaz, 2023, *Recent progress in hydrotreating kinetics and modeling of heavy oil and residue: A review,* Fuel. 334, 126404

E. Rodriguez, et al., 2018, *Modeling of hydrotreating catalyst deactivation for heavy oil hydrocarbons*, Fuel. 225, 118-133

P.J. Becker, B. Celse, D. Guillaume, V. Costa, L. Bertier, E. Guillon, and G. Pirngruber, 2016, *A continuous lumping model for hydrocracking on a zeolite catalysts: model development and parameter identification*, Fuel, pages 73–82.

P.J. Becker, N. Serrand, B. Celse, D. Guillaume, H. Dulot, 2016, *Comparing hydrocracking models: Continuous lumping vs. single events*, Fuel 165, 306-315.

Z. Pang, et al., 2024, *Data-driven prediction of product yields and control framework of hydrocracking unit*. Chemical Engineering Science, 1193386