Transfer Learning of Hydroprocessing Model from Fossil Feedstocks to Waste Plastic Pyrolysis Oil

Warumporn PEJPICHESTAKUL\*, Per Julian BECKER, Benoit CELSE

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

warumporn.pejpichestakul@ifpen.fr

Abstract

Hydroprocessing of waste plastic pyrolysis oil (WPPO) is a promising technology for upgrading low-quality pyrolysis oil in order to send it to a steam cracker. This unit operation is the first step to the chemical plastic recycle. However, developing predictive models for this process is challenging due to limited data availability. The aim of this paper is to show that the knowledge from fossil feedstocks can be transferred to plastic recycle. This study shows an application of transfer learning application to develop a naphtha density model for WPPO using data from fossil fuels. The Bayesian transfer learning approach effectively transferred knowledge from the source data to the target data. The cross validation at different g-prior was applied to obtain the optimal g value. The transfer model with optimal g value results in an accurate predicted naphtha density on the testing and unseen datasets. It outperforms the model trained solely on the target data while delivering comparable performance on the training dataset. This confirms the robustness and predictive capability of the transfer model.

**Keywords**: Transfer Learning, Hydrocracking, Pyrolysis Oil, Bayesian Statistics

* 1. Background

The need to address plastic waste management and reduce reliance on fossil feedstocks in the petrochemical industry requires sustainable solutions. One such solution is the upgrade of waste plastic pyrolysis oil (WPPO) through hydroprocessing. This approach has the advantage of using existing facilities in refineries. It furthermore provides a means to transform the low-quality and impure pyrolysis oil into more valuable products that can be used as chemical feedstocks for a steam cracker to produce precursors for new plastics.

The understanding of the hydroprocessing of this novel feedstock remains elusive, in contrast to the extensive knowledge accumulated over decades in the hydroprocessing of fossil feedstocks. To accelerate the industrialization of this process, there is a critical need for predictive models to support process development.

This work aims to develop a product property model for the hydroprocessing of WPPO by applying transfer learning based on the Bayesian approach from the fossil fuel data. The criteria of accuracy and robustness are used to evaluate the model.

* + 1. Process description

Hydroprocessing is a crucial process in petroleum refining, which converts the heavy fractions into more valuable products. Typically, it consists of one or more fixed bed reactor(s), where feedstock is fed over a catalyst with the presence of hydrogen. The classic hydroprocessing of fossil fuels includes the hydrotreatment (HDT) section, which removes impurities i.e., sulfur and nitrogen, followed by the hydrocracking (HCK) section, which breaks down heavy hydrocarbons into lighter, more valuable products. It is a flexible unit which can operate in different modes such as the maxi-middle distillate mode which aims to maximize diesel and jet yields and maxi-naphtha mode which aims to maximize naphtha yield. Further information on the maxi-naphtha mode is discussed elsewhere (Becker et al., 2023). The feedstocks for the maxi naphtha mode can be vacuum gas oil (VGO) or middle distillate.

The WPPO upgrade via hydroprocessing aims at transforming the impure WPPO into the steam cracker feed, which must adhere to strict feedstock specifications. Therefore, it requires several steps of impurities removal and hydrotreatment to produce naphtha. The products from hydrotreatment of WPPO might contain the diesel fraction that requires further conversion i.e., via hydrocracking to maximize naphtha.

This work focuses on the hydrocracking reactor, which is typically a single reactor in the pilot plant study as shown in Figure 1. The HCK reactor feed is first hydrotreated to remove impurities, particularly organic nitrogen that is a poison to zeolitic HCK catalyst. The catalysts are commercial bifunctional zeolite-base hydrocracking catalysts.

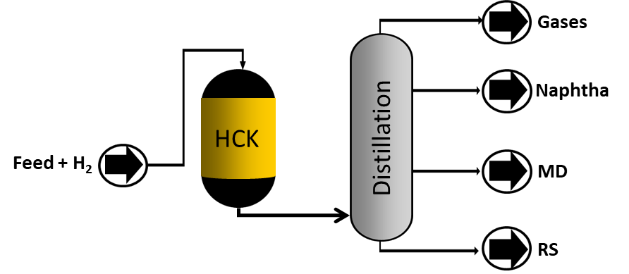


Figure 1 Schematic representation of hydrocracking reactor used in the pilot plant study.

* + 1. Product and Feedstock properties
       1. Fossil fuels

Petroleum feedstocks and products are complex mixtures of long-chained hydrocarbons. The detailed characterization of gas oil (GO) or heavier cuts remains currently impossible, as they are composed of thousands of individual species. Petroleum cuts are characterized by Simulated Distillation SimDist (ASTM D2887), and density (ISO 12185), as well as the concentration of nitrogen and sulfur. Thanks to its extensive knowledge in petroleum feedstocks, some properties such as aromatic carbon (CA), naphthenic carbon (CN) and paraffinic carbon (CP) can be estimated using the ndM method (ASTM D3238).

* + - 1. Waste Plastic Pyrolysis Oil

WPPO is produced from the pyrolysis process of plastic waste. Its composition is different from that of fossil fuels as it contains unsaturated compounds and contaminants that are derived from the additives of the plastic. Its lower heating value is comparable to diesel. Hydrotreated WPPO is used as a feedstock in this study. Therefore, olefins are saturated and contaminants i.e., silicon, chlorine and metals have been removed. The macroscopic properties such as SimDist, density, sulfur, and nitrogen are characterized using the same standards as for fossil fuels. However, the CA content is evaluated using 13C NMR analysis, as the ndM method is not adapted to this new feedstock.

* + - 1. Naphtha product

Full range naphtha is a mixture of C5-C12 hydrocarbon with a boiling range between 30-200 °C. The liquid composition is characterized using gas chromatography with flame-ionization detection (GC/FID). It provides the concentration of each family (n-paraffins, iso-paraffins, olefins, naphthenes and aromatics). and carbon number. These measurements are combined with the analysis of the gas produced from the reaction. This recombined data allows for the calculation of the density and distillation curve of the cuts. It also allows for the creation of naphtha cuts with different cut points i.e., C5-150 or   
C5-175 °C. In this work, the product property of interest is naphtha density.

* 1. Methodology

The product property models are typically linear models as they are simple, robust, and interpretable. The models have been developed for the hydrocracking of fossil fuels, which are used as the source model. The source model is developed using an ordinary linear square (OLS) method. The features are selected using exhaustive feature selection (Section 2.1). Then it is transferred to the target data, the hydrocracking of WPPO, while maintaining identical features, using a Bayesian transfer learning (Section 2.2). The data is scaled according to the median for centering and interquartile range for reducing. The cross validation at different g-priors (Section 2.3) was performed to estimate the optimal value to be used for the transfer model. This transfer learning method is applied to different feedstocks (fossil to WPPO) having distinct matrices. Additionally, the g-value cross-validation is applied to guarantee the transfer model's predictability and robustness.

* + 1. Feature Selection

Exhaustive feature selection was performed using all available macroscopic descriptors (feed properties, operating conditions and SimDist of the naphtha cut) discussed in Section 1.2. The retained parameters are SimDist of the naphtha cuts at 30 % and 70 %, CA content of feed, and conversion.

* + 1. Bayesian Transfer Learning

The methodology has been discussed in detail elsewhere (Iapteff et al., 2021), a brief overview is provided here. The linear model for the target data is , where *βt* is model parameters of size *np*, ***X*** is the design matrix, and *y* is the target variable. In the transfer approach (Bouveyron and Jacques, 2010), *βt* is treated as a random variable with a prior density π(βt). The Bayes’ theorem is then applied to derive the posterior of *βt* with respect to target observations.

An improvement of Zellner’s prior (Zellner, 1986), also known as g-prior, was proposed by (Iapteff, 2022). Only the mean of the prior distribution is affected by the source data when using such a prior. The target data determines the structure of the prior covariance of *βt*. The impact of the prior distribution can be adjusted using a scalar parameter *g*. The posterior mean (Eq. 1) tends to the maximum likelihood estimator learned on only the source data () for large values of g and to the prior mean () for small values. Iapteff (2022) proposed a heuristic method for estimating for optimal g value in Eq. 3, which was tested for diesel density prediction. It is the inverse of the diagonal elements of Σs as given in Eq. 2. The parameters with subscription s correspond to the source data.

|  |  |
| --- | --- |
|  | (1) |
|  | (2) |
|  | (3) |

* + 1. Cross Validation at Various g-prior

As discussed in Section 2.2, the g-prior is crucial for the transfer. The sensitivity analysis on the g-prior was carried out by varying the g values at different orders of magnitude with respect to the suggested value by Iapteff (2022). The cross validation (CV) of the target data with 5-fold cross validation was carried out. The optimal g value is then chosen based on the lowest mean RMSE on the testing dataset of the cross validation.

* 1. Datasets
     1. Source Data

The source data is from the hydrocracking of VGO and GO in maxi-naphtha mode. The VGO data comprises two sets: stage 1 (the hydrocracking of hydrotreated VGO) and stage 2 (hydrocracking of the middle distillate and residue fractions of stage 1 effluent). The detailed composition of naphtha cuts makes it possible to create different virtual cuts, which are C5-150 °C and C5-175 °C. By using these two virtual cuts, twice the number of data points are obtained which can be used for the development of a generic model for naphtha cuts. 263 observations were retained after the outlier removal using local outlier detection (LOF). The data is split into training and testing datasets using the Kennard-Stone algorithm. The features were then selected as discussed in Section 2.1. Table 1 summarizes the performance of the training and testing datasets of the source data. The model performances are well below the acceptance criteria of ±0.005 g/cm3.

Table 1 Statistical indicators for source model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | Nb. points | MAE | RMSE | ± 0.01 (%) | ± 0.005 (%) | ± 0.0025 (%) |
| Train | 194 | 0.0031 | 0.0039 | 97.9 | 82.5 | 49 |
| Test | 69 | 0.0023 | 0.0031 | 98.6 | 94.2 | 66.7 |

* + 1. Target Data

The target data for WPPO hydrocracking contains 6 experimental points from one feedstock and two liquid hourly space velocities (LHSV). The compositions of the total liquid were analyzed using GC/FID and two-dimensional gas chromatography (GCxGC). This makes it possible to construct virtual cuts beyond 175 °C. Due to the limited information, the training and testing datasets were split by LHSVs. Table 2 summarizes the datasets of the target data. The training and testing datasets are the naphtha cuts with the same as the source data (C5-150 °C and C5-175 °C). The WPPO data at other cut points, i.e., C5-160, C5-200, C5-225, and C5-250 °C are then used to validate the models and considered an unseen dataset.

Table 2 Summary of the target data.

|  |  |  |  |
| --- | --- | --- | --- |
| Datasets | Nb. points | Cuts | LHSV (h-1) |
| Train | 8 | C5-150, C5-175 °C | A |
| Test | 4 | C5-150, C5-175 °C | B |
| Unseen | 24 | C5-160, C5-200, C5-225, C5-250 °C | A & B |

* 1. Results and Discussion

The application of transfer learning to the naphtha density model, transferring knowledge from fossil data to WPPO, is examined and compared with a model trained solely on the target data (without transfer). Different g values in the transfer models are evaluated against an estimation by Iapteff (2022) , which is 0.05. In Figure 2a, the mean RMSE from cross validation is compared across various g values, indicating an optimal g value of 0.0005. A small g value indicates that the posterior mean tends to the prior mean. With the optimal g value, the signs of all normalized coefficients from the source model are conserved, as depicted in Figure 2b.

A smaller g value of 5E-05 results in a higher mean RMSE than the optimal value (0.05), indicating a similarity to the model without transfer (Target model). Notably, the model without transfer results in a negative coefficient for SimDist at 30 %, unlike the source model, contradicting the expectation that higher boiling point generally correspond to higher density. This contradiction of the model without transfer results from the insufficient training data and implies that without transfer learning the model suffers from significant overfitting.

The models with and without transfer are first applied to the training and testing datasets of the target data, the WPPO data. Then, they are applied to predict the unseen dataset, naphtha at other cut points. This challenges the model as it is an extrapolation to cut points higher than the ones used in the training dataset. Figure 3 shows the parity plot of the model predictions using the model with transfer (blue) and the model without transfer (orange) on the unseen dataset.

|  |  |
| --- | --- |
| a) mean RMSE from CV | b) Normalized coefficients of all models.  Transfer model with a g value of 0.0005. |

Figure 2 Panel a: mean RMSE from cross validation at different g values. Panel b: the comparison of the normalized coefficients of different models.

Table 3 presents a summary of the statistical indicators for models applied to the training, testing, and the unseen datasets of the target data. The model without transfer, calibrated solely on the target data, marginally outperforms the transfer model using a g value obtained from cross validation (0.0005). However, the transfer model provides better performance overall on the testing and unseen datasets, surpassing the model developed solely on the target data (without transfer) while delivering comparable performance on the training dataset. This confirms the robustness and predictive capability of the transfer model.

Transfer learning is a powerful tool that can facilitate model development for new conditions and feedstocks. Cross validation at various g values is used to selecting the optimal value for g-prior. Due to limited experimental data, assessing its robustness across different WPPO feedstocks poses a challenge.

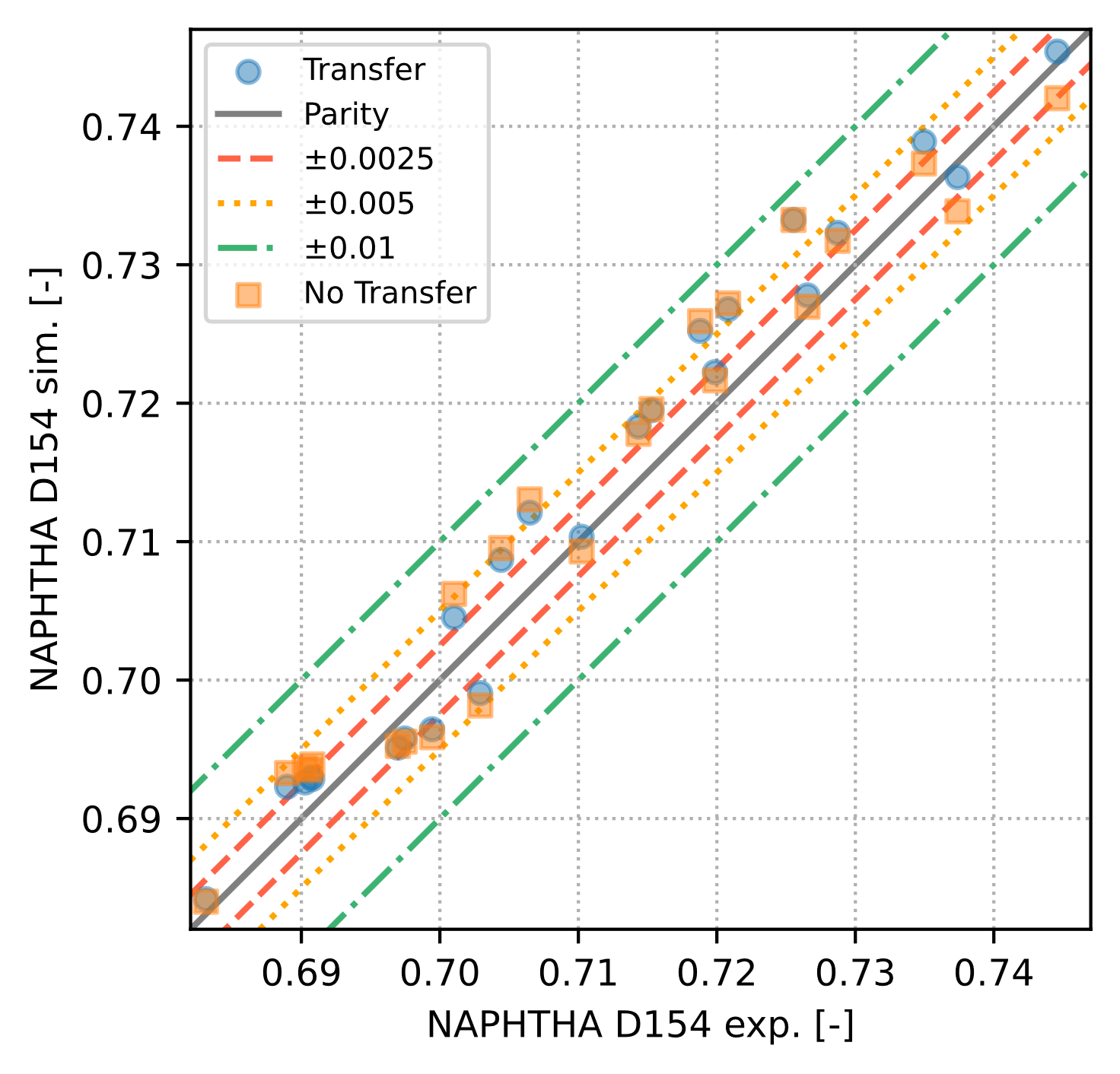


Figure 3 Parity plot comparing the model with (blue) and without transfer (orange) on the unseen dataset.

Table 3 Statistical indicators of models for target data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset/ model | MAE | RMSE | ± 0.01 (%) | ± 0.005 (%) | ± 0.0025 (%) |
| Train without Transfer | 0.0006 | 0.0007 | 100.0 | 100.0 | 100.0 |
| Train with Transfer | 0.0007 | 0.0009 | 100.0 | 100.0 | 100.0 |
| Test without Transfer | 0.0024 | 0.0027 | 100.0 | 100.0 | 50.0 |
| Test with Transfer | 0.0013 | 0.0017 | 100.0 | 100.0 | 75.0 |
| Unseen without Transfer | 0.0036 | 0.0041 | 100.0 | 75.0 | 29.2 |
| Unseen with Transfer | 0.0032 | 0.0037 | 100.0 | 83.3 | 45.8 |

* 1. Conclusion

This study has applied transfer learning using the Bayesian transfer learning approach to develop a model for predicting naphtha density of the hydroprocessing of WPPO, using data from the hydroprocessing of fossil fuels. The cross validation at different g-priors was applied to obtain the optimal value used for the transfer model. The model with transfer accurately predicted naphtha density on the testing and unseen datasets and outperforms the model trained solely on the target data while delivering comparable performance on the training dataset. This confirms the robustness and predictive capability of the transfer model.

References

P.J. Becker, L. Iapteff, B. Celse, Improving Model Robustness with Transfer Learning for Product Property Models, 33rd European Symposium on Computer Aided Process Engineering, 52, p. 1053–1058 (2023)

C. Bouveyron, J. Jacques, Pattern Recognition Letters ,31, 2237–2247 (2010)

L. Iapteff, Transfer Learning for Smart Predictive Analytics. PhD Thesis (2022)

L. Iapteff, J. Jacques, M. Rolland, B. Celse, J. R. Stat. Soc. Ser. C. Appl. Stat. ,70, 1344–1364 (2021)

A. Zellner, Bayesian Inference and Decision techniques, 233–243 (1986)