IASH 2019, the 16th INTERNATIONAL SYMPOSIUM ON
STABILITY, HANDLING AND USE OF LIQUID FUELS
Long Beach, California, USA
8-12 September 2019

Machine Learning Algorithms for Fuel Property Prediction in Comparison with State of the Art Physical Models

Clemens Hall,
Bastian Rauch, Uwe Bauder, Patrick Le Clercq, Manfred Aigner

DLR, German Aerospace Center, Institute of Combustion Technology, MAT, 70569 Stuttgart, Germany

Abstract

A reliable way to accurately predict the properties of fuels is essential to ensure their effective utilization. Furthermore, this could facilitate the development of new alternative jet fuels and help to streamline their approval process. This work investigates the application possibilities of the probabilistic Machine Learning (ML) algorithm Gaussian Process Regression (GauProReg) for the prediction of jet fuel properties. A critical evaluation is performed by comparing ML-based property predictions with the state-of–the-art physical-based discrete thermodynamics model.

A sub-set of properties from ASTM D1655 Table 1 was selected for this study. The detailed fuel data set from the CRC world fuel survey (CRC Report No. 647, June 2006 CRC, INC.) was used as the basic data set to represent the variability of conventional fuel properties. As input for both models, evaluated GCxGC measurements with 7 considered hydrocarbon families and carbon atoms in a range of 1 to 25 were utilized. The validity of the predicted values from the GauProReg was cross-validated in 4 folds with 25 % percent of test data.
The compared physical model was based on measured thermodynamic constants or derived from group-contribution methods. It received the same detailed compositional information as the ML algorithm. The two approaches were tested and compared with respect to their accuracy in the validated / trained data range and also their extrapolation behaviour on the basis of one conventional and two alternative fuels. Furthermore, the uncertainty of the GauProReg predictions was investigated systematically.

KEYWORDS: Aviation Fuel, Alternative Fuels, Machine Learning, Specification Properties, Modelling

ABBREVIATIONS

ASTM	American Society for Testing and Materials
CI	Confidence Intervals
CRC	Coordinating Research Council
DCM	Discrete Component Model
DLR	German Aerospace Center
GCxGC	Evaluated two dimensional gas chromatography measurements
GauProReg	Gaussian Process Regression
ML	Machine Learning
RMSE	Root mean squared error

INTRODUCTION

To ensure the safety and reliability of aircraft, jet fuels have to be certified in an approval process regulated by ASTM D4054. The number of properties and the broad spectrum of necessary tests make this process time consuming and very expensive, which restrains the development of alternative jet fuels. One solution is the introduction of a prescreening process, which assesses potential fuel candidates before entering the approval process and furthermore supports there optimization using mathematical models [1].
Traditional models are based on a mathematical description of the physical relationships. To model the relationship, an understanding of the underlying physical laws with their determining physical parameters and the actual determination of those parameters is necessary. Jet fuels are highly complex mixtures with hundreds of possible hydrocarbon combinations. This complexity makes the direct physical description of the fuel and its properties challenging. Furthermore, such relationships have to be developed for each property separately. State of the art property models for example, are based on approximating the distributions of fuel components in the different molecular families by distributions or by selecting representative molecules for each carbon number group in a first step, calculating their individual physical properties and finally calculating the desired property of the mixture with an appropriate mixture model. For properties of mixtures, where currently no sufficient models for properties or mixture exist, the correlations are approximated by empirical equations [1]. In recent years Machine Learning algorithms are more and more used for complex correlation models because of their flexibility in data input, the efficient training, testing and validation of the models [2]. Despite of their black box character and the fact that their correlation behavior is not completely comprehensive, they have proven to be reliable property models for chemical mixtures and even jet fuels [3] [4].

The scope of this work is a critical comparison of a state-of-the-art physical model and a probabilistic Machine Learning Algorithm with respect to their accuracy and their applicability to fuels inside and outside of the range of experience as recorded by the CRC world fuel survey [5]. For this purpose, we investigate the predictive performance of the models for three different fuels: A1 is a conventional Jet A, which represents an average crude-oil based Jet A1. B1 is an alternative fuel produced by the alcohol to jet (ATJ) consisting mainly of iso-alkanes, allowed in mixtures up to 50 vol% with conventional jet fuel. The third fuel is C1, a surrogate fuel with an for aviation fuels unusual high fraction of aromatics (20 vol%) and 34 vol% bi-cyclo-alkanes. The fuels were studied in the JETSCREEN project [6] and the fuel ID assumed for this work.

Since fuels are complex mixtures of hundreds of different hydrocarbon molecules, the composition can vary drastically, depending on the feedstock and the production pathway. Furthermore the identification of each molecule present in the fuel often not possible or too expensive [3]. Two-dimensional gas chromatography (GCxGC) is a comprehensive way of measuring the composition of fuels [3] [7]. The detected molecules are assigned to a matrix based on their hydrocarbon family and their number of carbon atoms. For this work hydrocarbon molecules from seven different families in a range of 1-25 carbon atoms were considered. Table 1 shows all families together with a characteristic molecule. Measured tri-cyclo-alkanes are lumped into the family of bi-cyclo-alkanes. Since their fraction is marginal (in all cases below 10 mass%) the resulting error can be neglected. For the course of this work each single parameter of the GCxGC data is called feature, the GCxGC data of one fuel is called feature set and the desired properties are called labels.

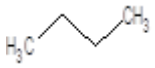
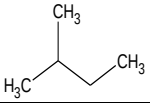
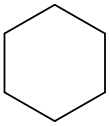
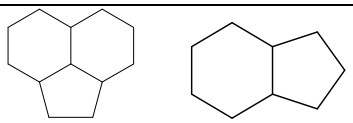
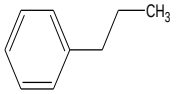
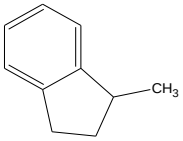
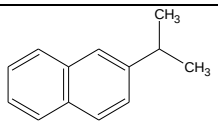
Fuel family	Molecular formula	Characteristic molecule
n-alkanes	n-CnH2n+2	
iso-alkanes	i CnH2n+2	
mono-cyclo-alkanes	CnH2n	
bi-cyclo-alkanes	CnH2n-2	
mono-aromatics	CnH2n-6	
cyclo-aromatics	CnH2-8	
di-aromatics	CnH2n-12	

Table 1*: Considered fuel families with their corresponding molecular formula and a representative molecule*

TEST FUELS

A plot comparing the composition of the three test fuels A1, B1 and C1 is shown in Figure 1. Table 2: Table of considered fuels with their composition summed up to the corresponding familiesTable 2 lists the tabular summary of the mass fractions.

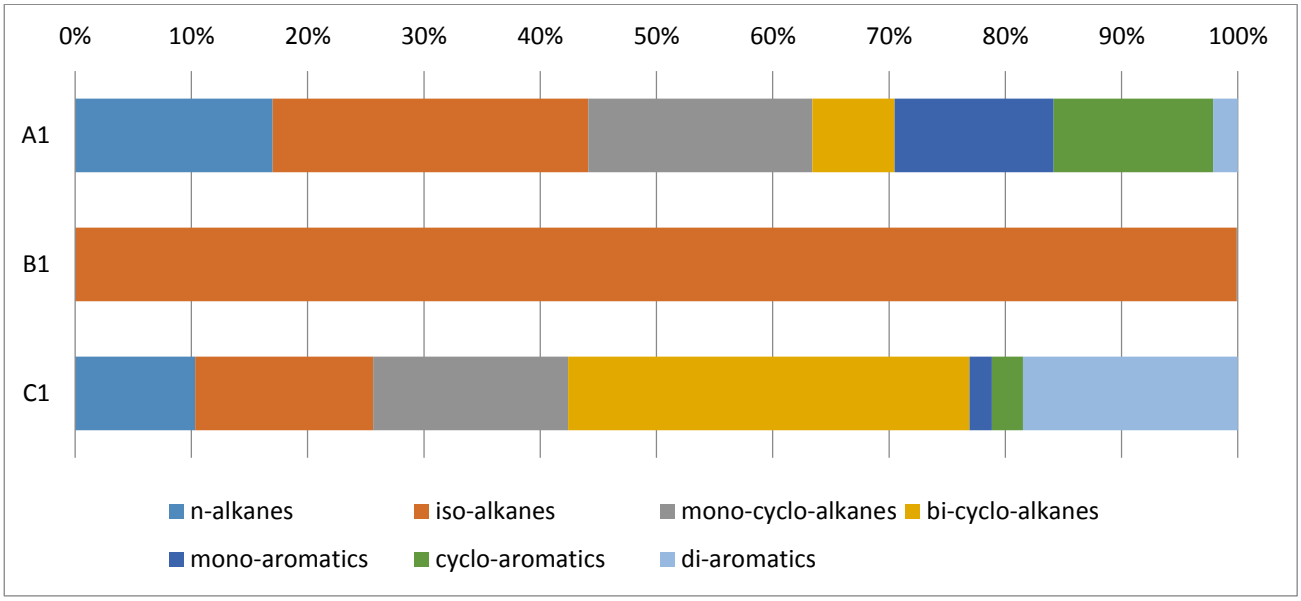


Figure 1: Comparative Plot of the considered fuels with their composition summed up to the corresponding families in %mass

Fuel	**n-alkanes**	**iso-alkanes**	**mono-cyclo-alkanes**	**bi-cyclo-alkanes**	**mono-aromatics**	**cyclo-aromatics**	**di-aromatics**
A1	19,2	30,7	21,8	8	15,5	2,4	1,7
B1	0	99,9	0,1	0	0	0	0
C1	10,1	15	16,4	33,8	1,9	2,6	20,1

Table 2: Table of considered fuels with their composition summed up to the corresponding families

The three specification properties from ASTM D1655 Table 1 selected for this study are: physical the fuel density, kinematic viscosity and distillation curve. To evaluate the adequacy of the models used in this study, the inferred model accuracy is compared to the ASTM D1655 property limits as displayed in Table 3.

Property		**Value**
Density at 15 °C [kg/m^3]		775-840
Viscosity at -20 °C [cSt]	Max	8
Distillation temperature		
10 % recovered [°C]		150-205
50 % recovered [°C]		165-229
Final boiling point [°C]	Max	300

Table 3: Selected ASTM D1655 property limits for jet fuels

Furthermore, the temperature dependent behavior of the properties is predicted to evaluate the models adequacy in predicting the fuel fit-for-purpose properties (ASTM D4054, Tier 2). Density and viscosity prediction are performed for a temperature range from -40 to 140 °C and for distillation range predictions are performed from 0-100 % recovered Volume in steps of 10% respectively. A prediction of the distillation curve of the fuels based on physical models is not considered in this study.

Physical Model

The Discrete Component Model (DCM) developed at the DLR uses pure component thermodynamic constants and mixing rules to predict relatively complex hydrocarbon mixture properties as a function of temperature. When measured critical properties, boiling point, and acentric factor of pure components identify by GCxGC-MS are not available (e.g. heavy cyclo-alkanes or cyclo-aromatics) then the Group Contribution Method (GCM) of Constantinou and Gani [8] is used for predicting these thermodynamic constants. The aforementioned relative complexity lies in the number of chemical families (see Table *1*) and the number of species per families (up to 25). However, to keep the complexity contained, we assume jet fuel to be a quasi-ideal liquid mixture of hydrocarbons, which means that the mixture density can be predicted using a linear mixing rule and molecular transport properties, which are dependent upon molecular interactions, because of the similarity in the hydrocarbon molecules can be predicted by classical non-linear mixing rules (e.g. Chueh and Prausnitz [9]. The proposed method was developed with the deliberate intention of generality. Actually, it can predict a large variety of crude-based as well as synthetic alternative jet fuels and does not rely on fitted parameters to be determined experimentally or with regression methods, which necessitate a large number of data. The generality comes with the acceptable loss in accuracy. For the liquid mixture density we followed the recommendation of Poling *et al* [10] and used the expression of Spencer and Danner as in [10] for the mixture molar volume:

$$V_m = R\left(\sum_i \frac{x_i T_{ci}}{P_{ci}}\right) Z_{RAm}^{\left[1+(1-T_r)^{0.2857}\right]}$$

$$Z_{RAm} = \sum x_i \, Z_{RAi}$$

with Racket's compressibility factor of each species given by the expression of Yamada and Gunn [10]:

$$Z_{RAi} = 0.29056 - 0.0877\omega_i$$

For the viscosity of pure liquid hydrocarbons as a function of temperature, we use the single-parameter equation of Mehrotra [11]:

$$\log(\mu + 0.8) = 100(0.01T)^b$$

where the value of the parameter b was regressed together with the molar mass M according to Mehrotra [11], for five families of light and medium liquid hydrocarbons:

$$b = B_0 + B_1[\log M] + B_2[\log M]^2$$

The mixing rule used for the viscosity of the final mixture of hydrocarbons is a generalization of the method of Katti and Chaudhri [12]:

$$\ln \mu_m \, V_m = \sum x_i \ln \mu_i V_i$$

where we've neglected the interaction term.

Machine Learning – Gaussian Process Regression

Machine Learning is a generic term enclosing a broad spectrum of mathematical concepts and

algorithms. The prediction of a property or a specific value is a typical regression problem from a Machine Learning point of view. The Gaussian Process Regression (GauProRreg) is a proven algorithm for regression problems with a high-dimensional data input. As a probabilistic model it furthermore brings the advantage of providing the standard deviation for each predicted point [11].

The GauProReg can be interpreted as an algorithm that approximates a distribution of regression functions $f(x)$ for training data $D = \{(x_i, y_i)|i = 1, \dots n\}$ composed of features X and labels Y with their corresponding probabilities [11]. Thereby not only the solution with the best fit or highest probability is given as a result, but also the standard deviation over all the functions. In the following plots, the standard deviation is chosen as confidence interval (CI). This means, that the model also covers the uncertainty quantification of the epistemic error, e. g. not enough measurement for the prediction of a certain point. The process is initialized by a distribution of random functions. The distribution is determined by a so called covariance function $k(x, x')$ or kernel. This kernel can be composed out of different sub-kernels. In the scope of this work a combination of three different kernels, with different weightings are used, which are listed below:

Radial-basis kernel: $$k(x, x') = \exp\left(-\frac{(x - x')^2}{l_{RBF}}\right)$$ *Equation 1*

White kernel: $$k(x, x') = \delta,\ 0 \text{ if } x = x'$$ *Equation 2*

Rational Quadratic: $$k(x, x') = \left(1 + \frac{(x - x')^2}{2\alpha * {l_{RQ}}^2}\right)^{-\alpha}$$ *Equation 3*

The Radial-basis kernel is a universal kernel for GauProReg applicable for most of the non-periodic regression problems. The White kernel deals with potential noise of data. The Rational Quadratic kernel is selected due to its characteristic of modeling smooth functions that do not vary too quickly. This kernel is equivalent of adding multiple Radial-basis kernels together. The weighted sum of all three kernels creates a covariance function that is able to cover the complexity of high-dimensional problems, handle the noisiness problem and return a smooth function as to be expected from physical behaviors.

The hyper parameters $l_{RBF}, l_{RQ}, \delta, \mathrm{C}$ and α, as well as the individual weights or maximum covariance are parameters that have to be adjusted during the training of the regression process. This task is typically accomplished by an optimizer, maximizing a likelihood function [11]. In this work the "L-BFGS-B" optimizer [12] and the log-marginal-function are used [11]. After the determination of the hyper parameters in the optimization loop, the predictions of test data features are compared with the test data labels. Based on their accordance, the covariance function is adjusted via Bayes theorem. The prediction of values for regions, where no training data is available, is possible via the assumption of a multidimensional Gaussian distribution N as shown in Equation 4. The unknown *(u)* value y_U from the test set can be calculated with known values *(k)* from the trainings set via consideration of the mean function μ and the covariance matrix which is composed out of the matrices of the covariance for the trainings set K_{kk}, the trainings and the test set K_{ku}, as well as the test set K_{uu} itself. The mean function and the corresponding variance at this location can be calculated by Equation 5

$$\begin{pmatrix} Y_k \\ y_u \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_k \\ \mu_u \end{pmatrix}, \begin{pmatrix} K_{kk} & {K_{ku}}^T \\ K_{ku} & K_{uu} \end{pmatrix}\right)$$ *Equation 4*

$$y_u | Y_k \sim N\left(K_{ku} {K_{kk}}^{-1} Y_k, K_{uu} - K_{ku} {K_{kk}}^{-1} {K_{ku}}^T\right)$$ *Equation 5*

The start parameters for the hyper parameter optimization as well as the individual weights of the kernel functions have to be adjusted before the actual regression task. There are different approaches to solve the optimization problem. The chosen Bayesian Optimization approach uses Gaussian Process which works with the same principles described for the property regression. In contrast, for the property prediction a Matern kernel from Equation 6 and the negative expected improvement function as loss function are used [13].

Matern:

$$k(x,x') = \sigma^2\left(1+\frac{\sqrt{3}(x-x')}{l_{Matern}}\right)\exp\left(-\frac{\sqrt{3}(x-x')}{l_{Matern}}\right) \qquad \textit{Equation 6}$$

This protocol was repeated until the convergence criterion or the maximum number of iteration is reached. Inside of this optimization loop a fourfold cross validation is executed, that randomly divides the data in four different training and test sets with a test fraction of 25%. The mean squared error of the cross validation is used as optimization objective for the Gaussian Process. The maximum calls of the optimizer are set to 30, which proved by experience to be a good iteration restriction.

TRAINING DATA

The performance of a Machine Learning algorithm is highly determined by the dataset that it was trained and tested on. Number and variance of the datasets over the range of interest is important for a good predictive capability. Furthermore, it is essential to provide data that has been measured with identical or comparable methods. In the scope of this work the DLR database with 90 GCxGC measurements, mainly composed out of data of the CRC world survey [5] as well as internal measurement campaigns, is used. The data thereby mostly consisted out of crude-oil based Jet A /Jet A-1 fuels. It has to be mentioned, that not for all fuels the same amount of property data was available Table 1 shows the number of fuels and measurements that were available for the training and testing of each property model.

Property	**Fuels**	**Measurements**
Density	81	419
Kinematic Viscosity	68	191
Distillation	73	479

Table 4: Overview of the number of fuels and measurements available for each property

For the application to the GauProReg, the training data can considered to be sparse, and concentrated on specific parameter combinations, e.g. most of the measurements were only available for certain temperature values. Figure 2 shows bar charts for each property and the respective number of data points available. .

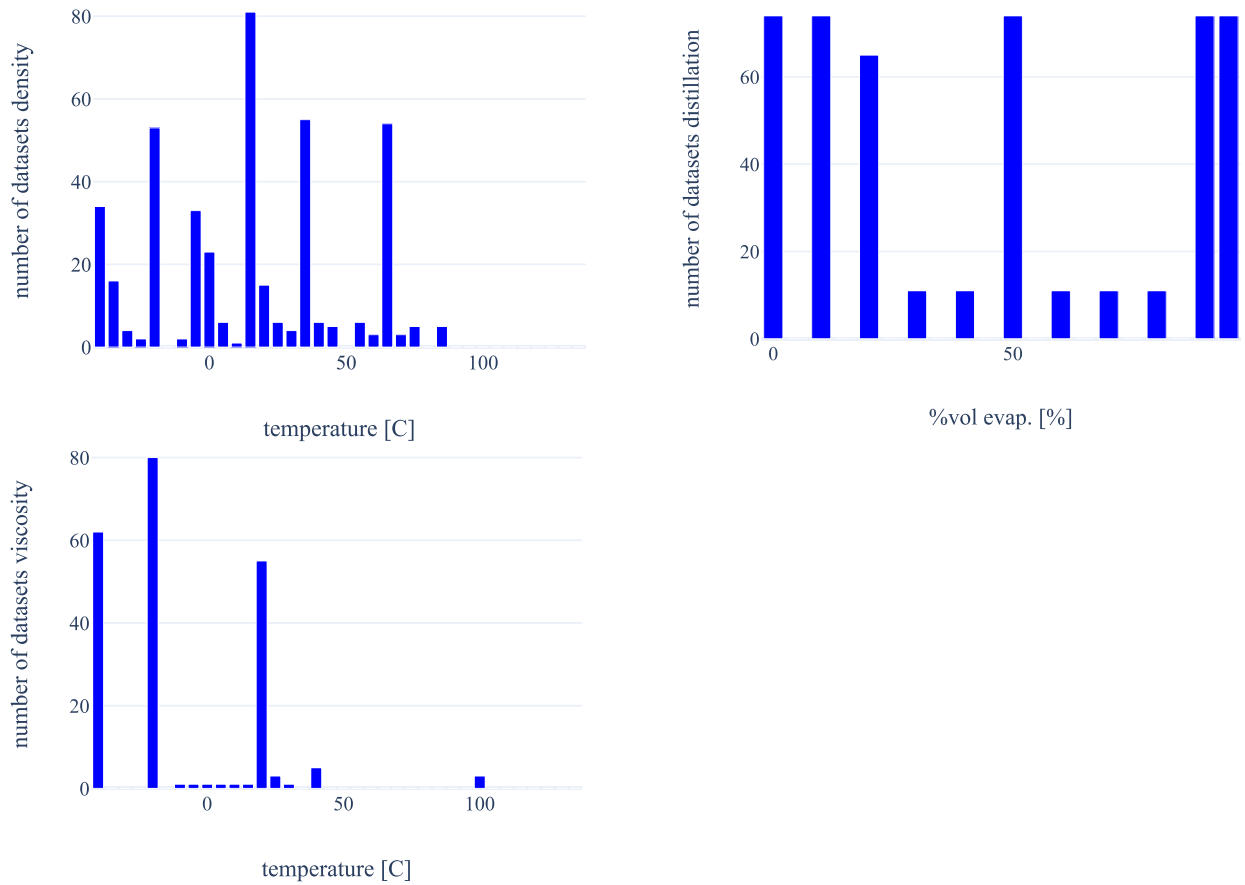


Figure 2: Bar chart of number of available datasets with respect to property and its dependency: density (upper left), viscosity (lower left) and distillation (upper right)

The amounts and numbers of training data as well as their sparsity and similarity with the tested fuels have to be considered, when interpreting the results.

RESULTS

GauProReg Training and Testing

Figure 3 to Figure 5 show the results of the four fold cross validation of the data for density, viscosity and distillation. The predicted values are plotted with respect to the true measured values. If prediction and measured value are in perfect accordance to each other, the dot lies on the unity line, which is indicated in black. The results for the training are displayed on the left, the results for the testing on the right. The standard deviation calculated by the GauProReg for each point is plotted as CI and corresponds to a certainty of 68 %. As a quantitative figure of merit the root mean squared error (RMSE) was calculated for each property, which is shown in Table 5 for the training and the testing.

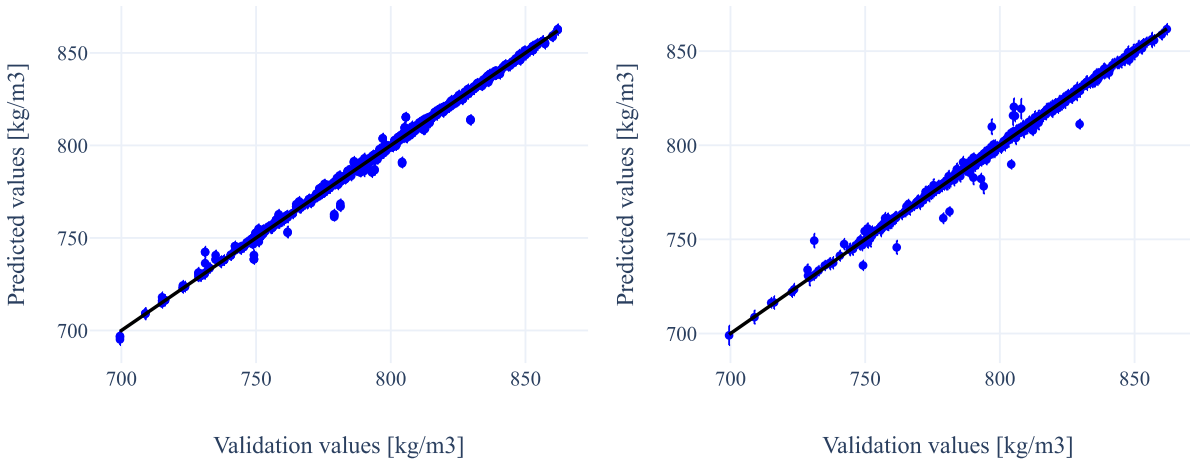


Figure 3: Results four fold cross validation for density

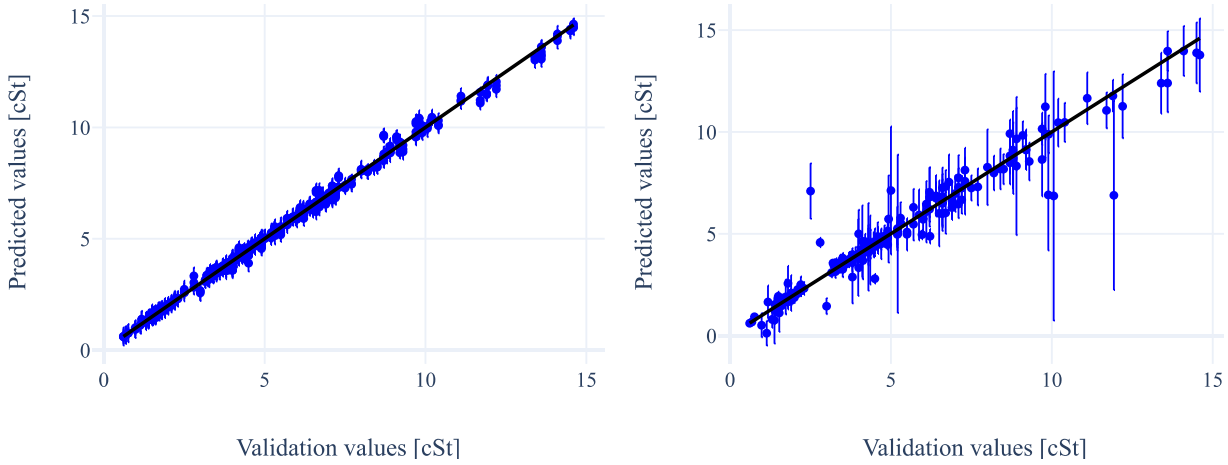


Figure 4: Results four fold cross validation for kinematic viscosity

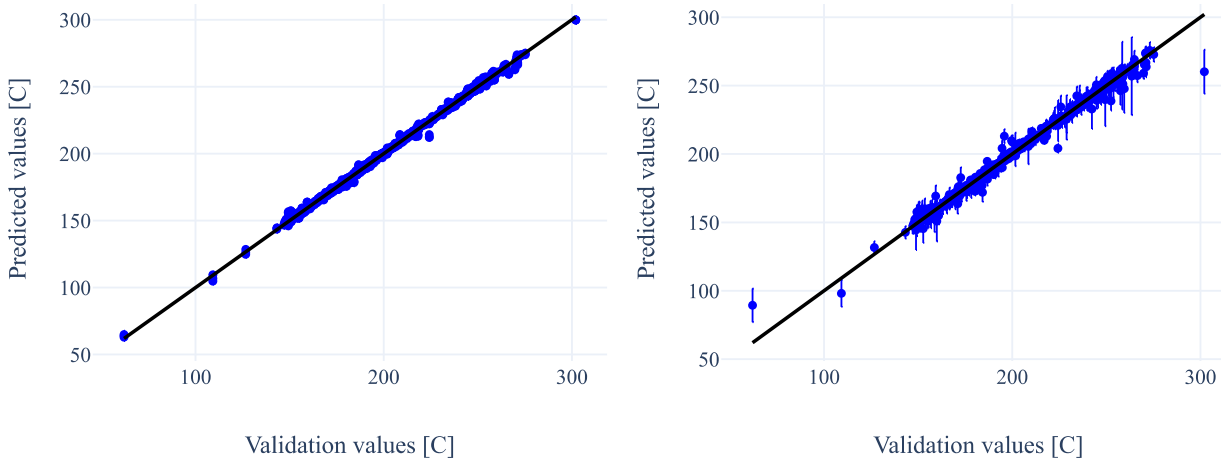


Figure 5: Results four fold cross validation for distillation

Property	**RMSE Training**	**RMSE Testing**
Density [kg/m^3]	2.206	2.91
Kinematic Viscosity [cSt]	0.159	0.706
Distillation [°C]	1.469	1.82

Table 5 Mean RMSE of four fold cross validation for all properties

For all properties sufficient prediction accuracies are reached in the training and testing. The values

lie close to the unity line, which indicates no systematical errors and which demonstrates the capability of GauProReg for the modelling of physical properties. For the test sets a similar trend is observed, but a few outliers are still visible. The deviation of the outliers is probably due to the composition of the fuels. They differ from the majority of the CRC Jet A fuels and are considered to be extreme fuels, partially unique in the database. This observation conforms to the RMSE values in Table 5. The RMSE values of training and testing are similar to each other, which indicates a sufficient fit and no overfitting.

Uncertainty Quantification: Calculation of Confidence Intervall for GauProReg predictions

The properties for each considered fuel are calculated over ranges appointed in the indroduction, for density and viscosity the temperature interval of -40 to 140 °C and for the distillation the range of evaporated volume from 0 to 100. Figure 6 shows the prediction of the viscosity for fuel A1 with GauProReg. The overall trend of the viscosity is sufficiently covered. The CI, as the measure of the uncertainty in the GauProReg prediction, on the left show the typical Gaussian bell-curve-like behavior of a GauProReg with a Radial-basis and a Rational quadratic kernel [11]. The CI increases with the distance to the last base point, at which training data is available. The sparsity of available data for the whole temperature range is thereby responsible for the bell shape of the CI. Since the trend of all fuel viscosities decreases with temperature, we introduced a correction of the CI based on the shape of viscosity to correct the CI in hindsight. Therefore, the following two assumptions are made: 1). The uncertainty correlates with the slope of the viscosity with respect to the temperature$K\frac{d\nu}{dT}\sim CI$, where K is a constant that has to be adjusted to output of the cross validation. 2). The overall trend is covered by the algorithm. In order not to distort the output of the cross validation, the constant K is fitted with Equation 7. Since the chosen CI is equivalent to the standard deviation of the GauProReg which mathematically corresponds to the formula of the RMSE for large enough datasets, K is adjusted so that the corrected RMSE and the one determined in the cross validation are the same. Therefore, the overall uncertainty is maintained but shifted to regions with higher gradients. The slope $\frac{d\nu}{dT_i}$ is calculated from values computed by the GauProReg with 5 K difference.

$$K\frac{1}{n}\sum_{i}^{n}\frac{d\nu}{dT_i} * CI_i = RMSE \qquad \textit{Equation 7}$$

The result of the physical correction of the CI is displayed on the right side of Figure 6. A clear decrease for temperature regions with low slopes is visible. This concept is applied for all considered properties. For the distillation lines, the slope is calculated with respect to the distillated fraction.

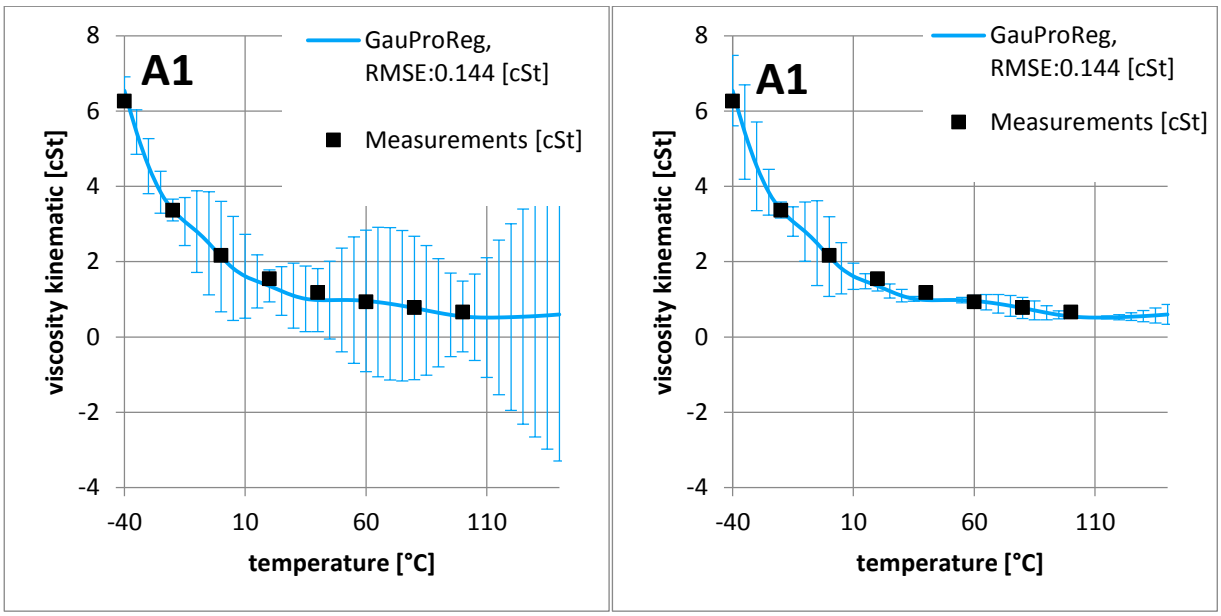


Figure 6: Viscosity prediction of the GauProReg and DCM for fuel C1 with no correction of the CI (left) and with CI correction (right)

Model comparison

The results of the density prediction for all three considered fuels with the adjusted CI can be seen in Figure 7. As an additional reference, the minimum and the maximum densities of fuels in the CRC world fuel survey are plotted. The lines of the CRC minimum and maximum density are fitted to a first order polynomial, using the available values at 15 and 65 °C.

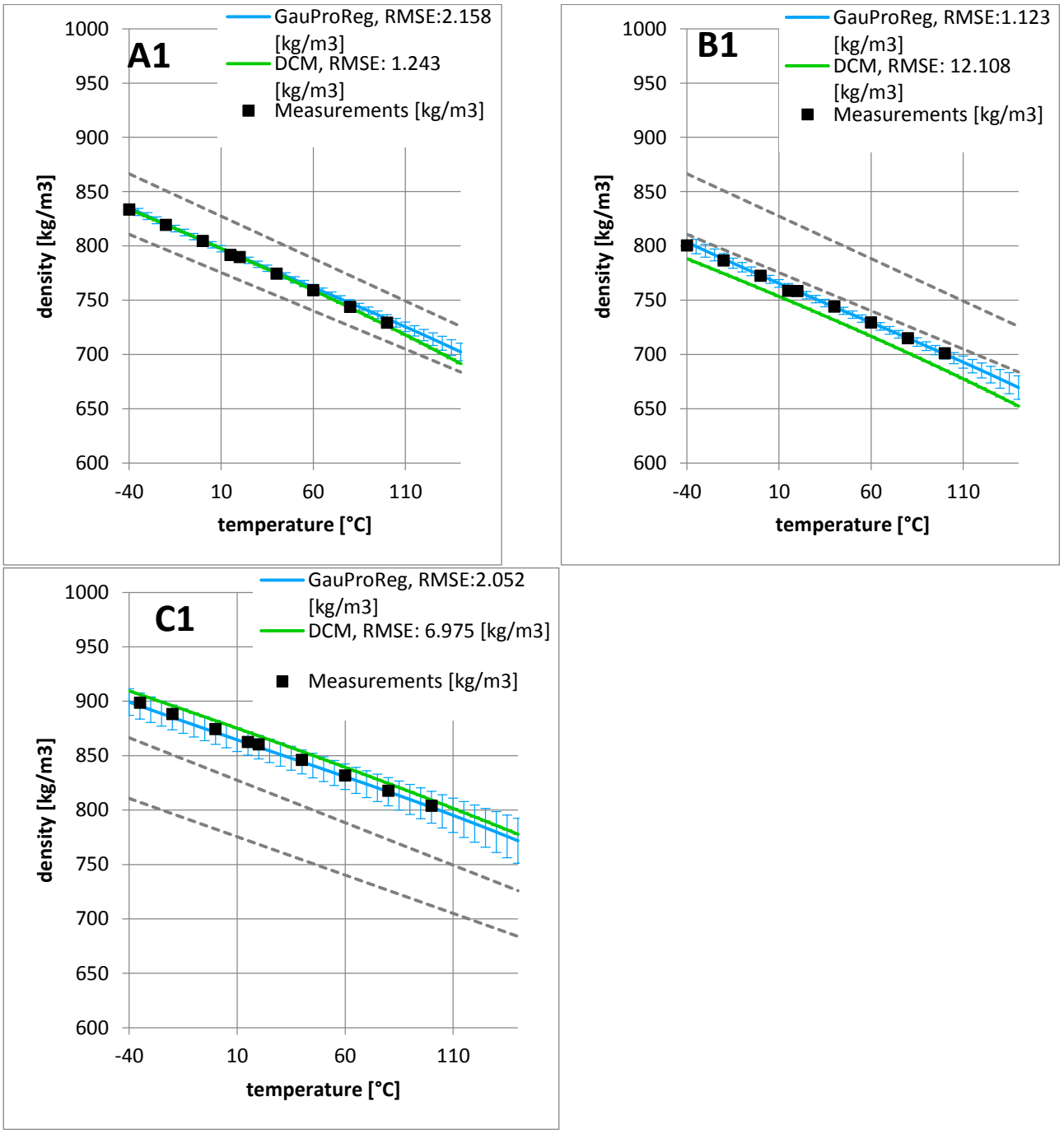


Figure 8: Results of the density prediction for all three considered fuels.

For all three fuels the predictions of both models are in good agreement with the measurements. For the extreme fuels B1 and C1, the DCM model underpredicts and overpredicts the density, respectively. This deviation is probably due to the wrong selection of representing molecules. The high accuracy of the GauProReg shows, that is able to extrapolate and correctly predict the values, if the property follows a linear trend in the range of interest. Furthermore the existence of a ATJ fuel in the training data explains the accuracy for fuel B1, mainly composed of iso-alkanes.

Figure 9 shows the results of the viscosity prediction for all three fuels. Again the minimum and maximum viscosities of the CRC world fuel survey are plotted, this time using a logarithmic equation using values at -40 and 40 °C.

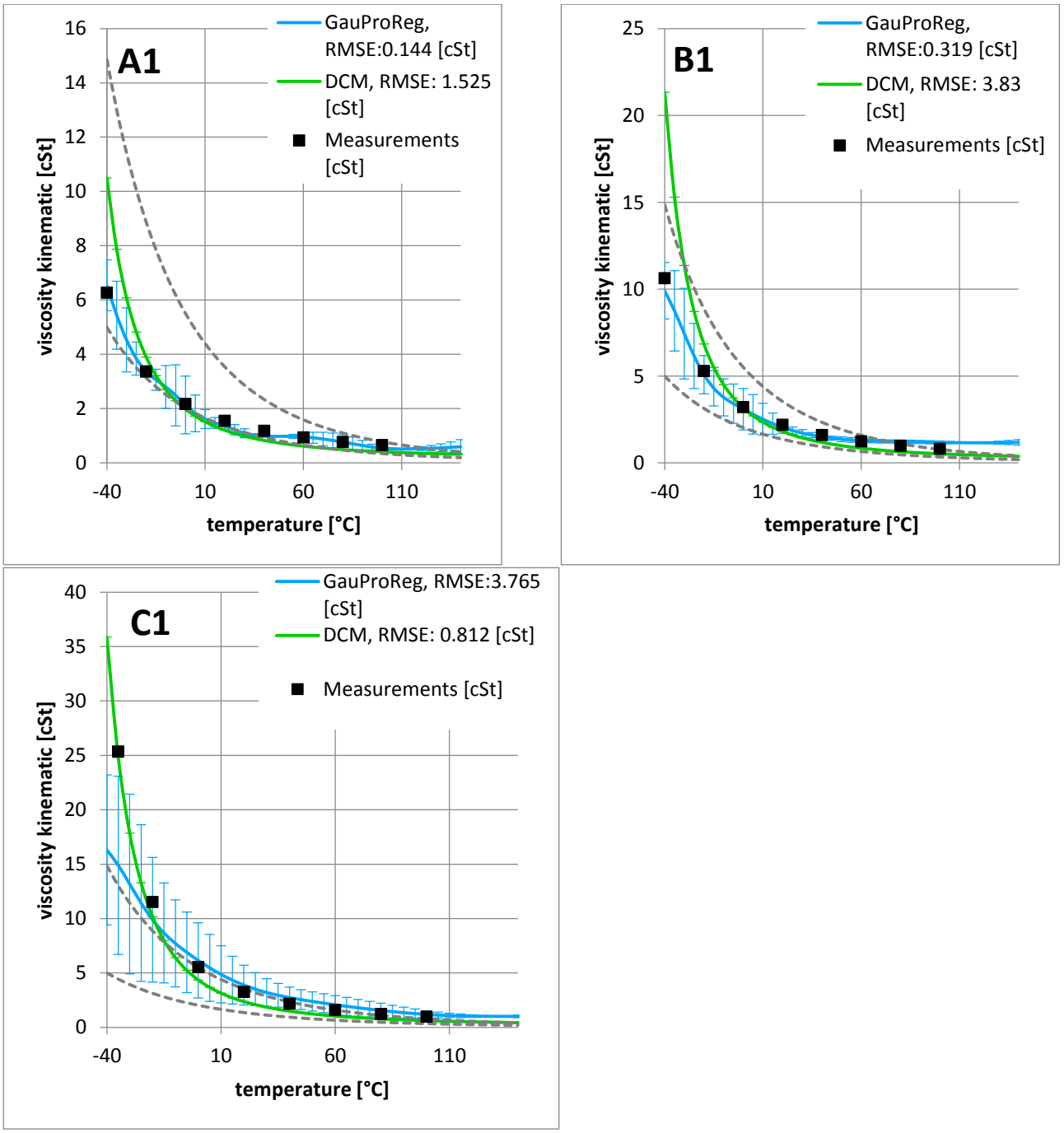


Figure 9: Results of the viscosity prediction for all three considered fuels

The prediction of the GauProReg show good accuracy for fuel A1 and the B1 over the whole temperature range. For C1 the prediction deviates at -20 °C. This observation can be explained by the available data of both regular Jet A fuels and the ATJ fuel in the database. The absence of a fuel similar to C1, results in significant deviations. The potential error is indicated, by the drastic increase of the CI, revealing the lack of data and the uncertainty due to the nonlinear trend of fuel viscosities at low temperatures. The slightly oscillating trend of the A1 prediction for the interval between 50 and 100 °C is likely due to the sparsity of the data with respect to the temperature, see *Figure 2*. Considering the very sparse data with respect to temperature available for the GauProReg training, the results nonetheless show a convincing physical trend. The DCM prediction of A1 and B1 deviates at low temperatures and overpredicts the viscosity. The overpredictive trend of the DCM however is likely due to the wrong selection of the representing molecules. The prediction of C1 in contrast turns out better for the DCM, even at the low temperature range.

The results of the prediction of the distillation lines are displayed in Figure 10. The minimum and maximum values of the CRC world fuel survey are again plotted as interpolated lines using the values at 0, 10, 50, 90 and 100 %vol evaporated. For A1 and B1 the predictions are in adequate accordance with the measurements. At higher %vol evap., the mean prediction of the GauProReg for B1 do not meet the measurement, but still lie inside the 68% CI. For the other extreme fuel C1, the measurements are only met in the intermediate %vol evaporation range. The measurements still lie inside the CI, but drastically away from the mean function. This deviation is probably again due to the absence of a similar fuel in the training data and the strong non-linearity of the distillation curves.

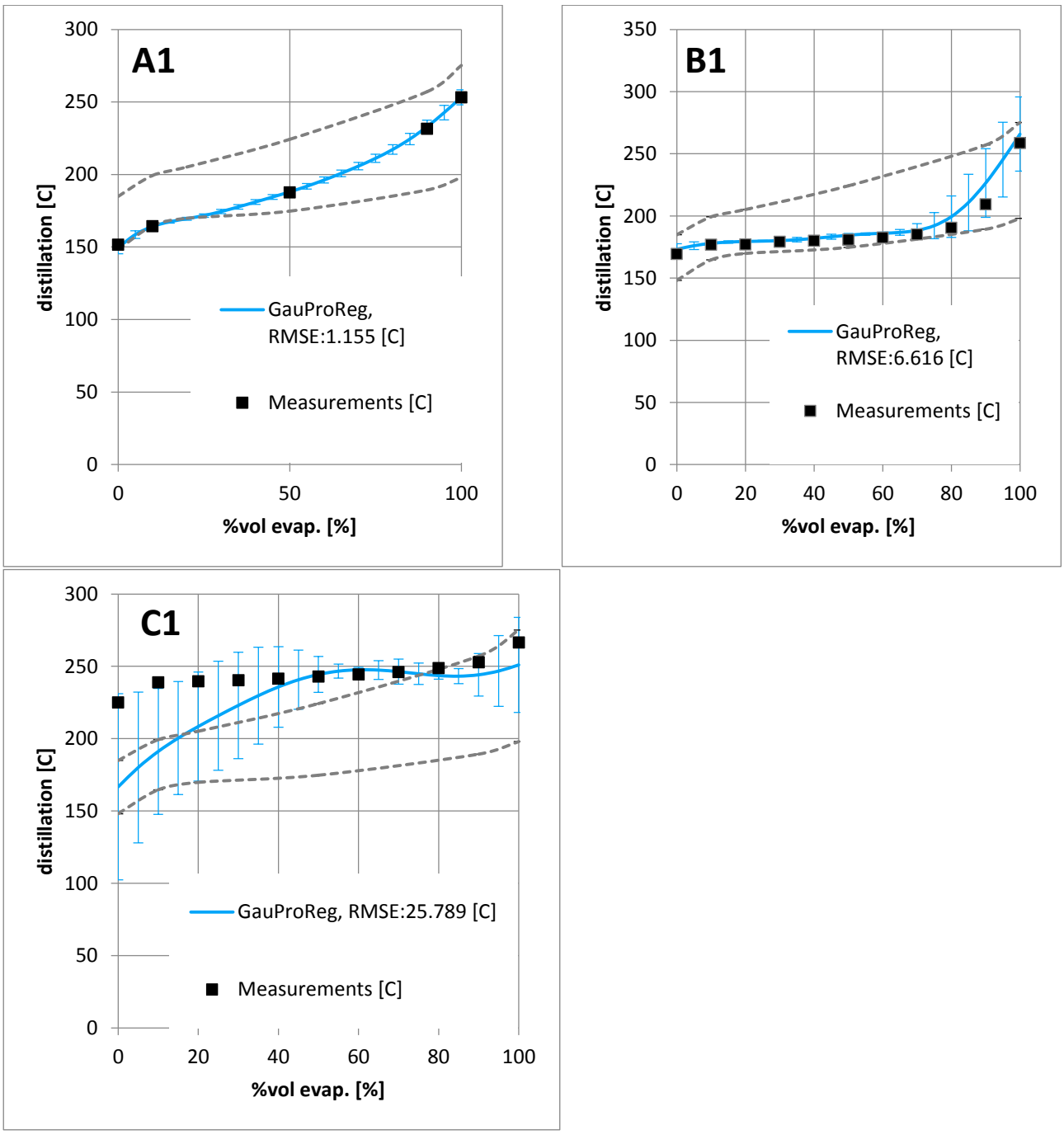


Figure 10: Results of the distillation lines for all three considered fuels

Overall the evaluations of the results show sufficient accordance for all considered properties of the fuels A1 and B1. This can be can be explained by the presence of similar fuels. The observed deviations of predicted properties with nonlinear behavior (distillation and viscosity), observed by C1, is likely due to the absence of a similar fuel in the training data. To investigate the influence of a similar fuel in the training data, we introduced C3, a fuel similar to C1, but with a lower fraction of di-aromatics. A comparison of their composition with respect to their hydrocarbon families is shown in Figure 11, the values are displayed inTable 6.

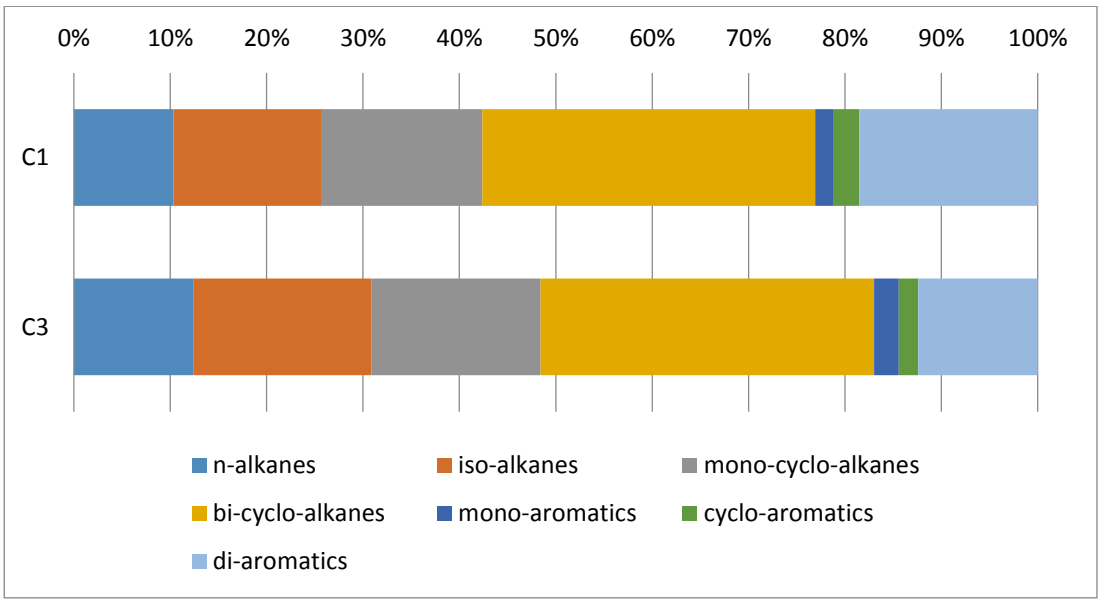


Figure 11: Comparison between the fuel composition of fuel C1 and C3 with respect to their hydrocarbon families in %mass

Fuel	**n-alkanes**	**iso-alkanes**	**mono-cyclo-alkanes**	**bi-cyclo-alkanes**	**mono-aromatics**	**cyclo-aromatics**	**di-aromatics**
C1	10,1	15	16,4	33,8	1,9	2,6	18,1
C3	12,2	18,2	17,3	34,1	2,5	2	12,2

Table 6: Detailed composition of fuels C1 and C3 with respect to their hydrocarbon families

The results of the property prediction for fuel C1 with C3 in the training set are shown for all properties in *Figure 12*, the results of the GauProReg with C3 in the training set is thereby plotted in orange. For viscosity and distillation, a drastic enhancement in accuracy was achieved. The predictions lie close to the measurements. For density a slight increase in the RMSE is observed, probably due to the overfitting to the introduced fuel C3. Notable is the reduction of the CI, which indicates the better confidence of the GauProReg for the predicted values. All in all this underlines the importance of similar fuel in the training set.

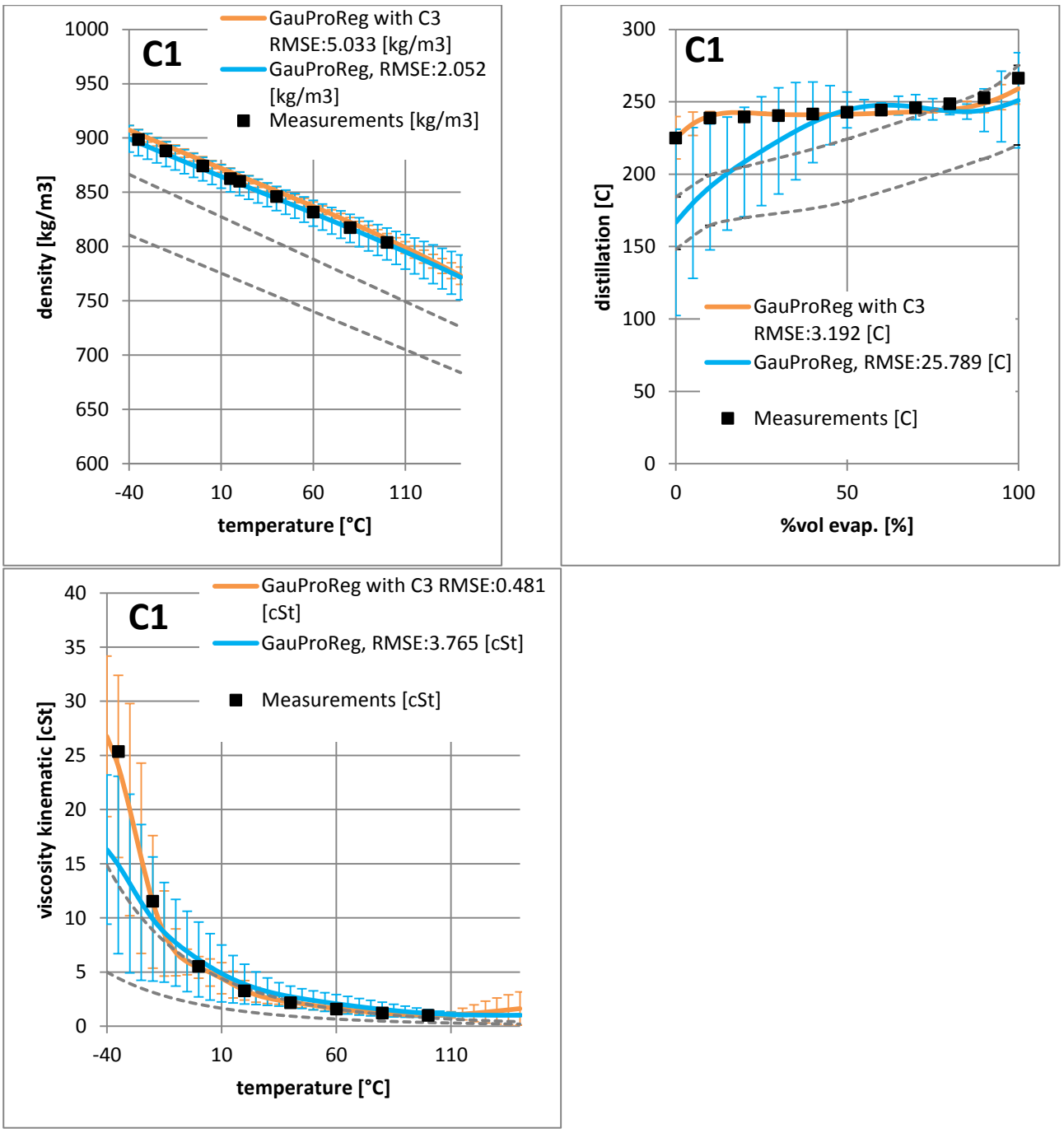


Figure 12: Property prediction of density (upper left), viscosity (lower left) and distillation (upper right) with fuel C3 in the training set

SUMMARY AND OUTLOOK

This work investigated the possibility of the probabilistic Machine Learning algorithm Gaussian Process Regressor (GauProReg) for the prediction of jet fuel properties and compared it the with the state-of-the-art Discrete Component Model (DCM). As a reference case, the prediction of selected the ASTM D1655 relevant fuel properties density, kinematic viscosity and distillation were chosen. The prediction was carried out for three fuels: A1 (traditional Jet A fuel, crude oil based, ASTM D4054 certified), B1 (alternative fuel produced by alcohol to jet (ATJ) production) and C1 (fuel with an unusual high amount of aromatics and bi-cyclo-alkanes). Both models had the same data input and processed the fuel composition in the form of evaluated GCxGC measurements 25x7 matrix with 7 considered hydrocarbon families and considered molecules with up to 25 carbon atoms. The GauProReg was trained using a data base of 81 fuels and 419 measurements for density, 68 fuels and 191 measurements for the kinematic viscosity and 73 fuels and 479 measurements for the distillation. The covariance function of the GauProReg was composed out of three weighted sub-kernels: Radial-basis, Rational Quadratic and White kernel. This covariance function was used for the prediction of all considered properties. The hyper parameter optimization was carried out using an outer loop Gaussian Process with a Matern kernel. The GauProReg Model was validated in a four fold cross validation. The cross validation proved that all considered properties could be modelled by the GauProReg on the basis of the GCxGC measurements. In training and testing similar accuracies were reached and no systematic errors were detected. The characteristic of GauProReg as probabilistic model is computing a standard deviation for every predicted value indicating the epistemic error certainty of the prediction. In this work the computed standard deviation was utilized for the Uncertainty Quantification of the algorithm as certainty intervals (CI) and figure of merit for the credibility of the predicted value. The CI was furthermore improved using an approach to shift the uncertainty from regions with low variance in the property to regions with higher ones, based on physical understanding of the properties. Thereby a physical understanding was integrated into the model. The results showed a sufficient fit and a convincing physical trend.

The subsequent comparison of the DCM and the GauProReg on the basis of the three test fuels produced similar accuracy for the property density, except for fuel B1, there the GauProReg overtrumped the accuracy of the DCM significantly. For prediction of the kinematic significant difference were observed, between the algorithms. The DCM displayed an over predicting behavior of viscosity for fuel A1 and B1 at low temperatures. For C1 in contrast, the predictions were in good alignment with the measurements. The GauProReg on the other hand produced good predictions for A1 and B1, for C1 howeber the prediction deviated at very low temperatures. The prediction of the DCM was only carried out with GauProReg. The results showed again a sufficient accuracy for fuel A1 and B1, the values of C1 deviated significantly from the predicted mean, but still lay inside the CI. All in all the GauProReg showed good predictive accuracy for fuels similar to the ones in the training set. The predictive capability of the GauProReg was traced back data available for the training of the algorithm. For the fuel A1 and B1 similar fuels were present in the training data, whereas a for C1 non was provided. This absence resulted in large errors. After the introduction of fuel C3, which is similar to C1, into the training set, the accuracy of the prediction for C1 improved drastically.

This work showed the potential of the GauProReg for the prediction of jet fuel properties, for traditional and alternative jet fuels, as well as surugates.. The accuracy was similar, if not better compared to the DCM. Furthermore, the intrinsically calculated standard deviation can be used as a measure for the credibility of the prediction and the potential error. The limitations of the GauProReg result from the fuels it had been trained on. For unseen fuels with unfamiliar composition the predictions can deviate and lie outside of the CI.

The adaptability and predictive capability of the GauProReg for all considered properties should be tested on more fuel properties and fuel behavior as well as data sets describing test rig or engine tests. The used covariance function proved to be versatile, but the output had to be corrected to

achieve a shape that represented the physical trend. With respect to the predictive capability of the models, further research is needed to correctly estimate the capability of the model and correlate the certainty with the potential error.

LITERATURE

[1] B. Rauch, Systematic Accuracy Assessment for Alternative Aviatin Fuel Evaporation Models, Stuttgart: DLR Stuttgart, 2017.

[2] G. Liu and L. W. e. al., "Artificial neural network approaches on composition-property relationships of jet fuels based on GC-MS," *Fuel,* 11 2007.

[3] Vozka et al., "Jet fuel density via GC × GC-FID," *Fuel,* pp. 1052-1060, 30 08 2018.

[4] Creton, Alonso et al., "Prediction of Density and Viscosity of Biofuel Compounds Using Machine Learning Methods," *Energy and Fuels,* pp. 2412-24-26, 20 03 2012.

[5] Coordinating Research Council, "CRC Report No. 647," 2006, 06.

[6] Jetscreen, "https://www.jetscreen-h2020.eu/," ARTTIC , [Online]. Available: https://www.jetscreen-h2020.eu/. [Accessed 03 08 2019].

[7] Hammond et al., "Navy Fuel Composition an Screening Tool," *Naval Research Laboratory,* 10 05 2014.

[8] B. E. Poling, J. Prausnitz and J. P. O'Connell, The Properties of Gases and Liquids, McGraw-Hill, 2004.

[9] P. K. Katti and M. M. Chaudhri, "Viscosities of Binary Mixtures of Benzyl Acetate with Dioxane, Aniline, and m-Cresol," *J. Chem. Eng. Data,* pp. 442-443, 1 July 1964.

[10] A. K. Mehrotra, "Generalized one-parameter viscosity equation for light and medium liquid hydrocarbons," *Ind. Eng. Chem. Res.,* pp. 1367-1372, 1 June 1991.

[11] C. Rasmussen, Gaussian Processes in Machine Learning, Berlin: Springer, 2003.

[12] F. Pedregosa, G. Varoquaux and A. Gramfort, "scikit-learn," *Journal of Machine Learning Research,* pp. 2825--2830, 12 2011.

[13] scikit-optimize developers, "scikit-optimize," scikit-optimize developers, [Online]. Available: https://scikit-optimize.github.io/. [Accessed 2 06 2019].

[14] ASTM International, "Standard Practice for Evaluation of New Aviation Turbine Fuels and Fuel Additives," ASTM International, West Conshohocken, PA United Stares, 2019.

[15] ASTM International, "Standard Specification for Aviation Turbine Fuel Containing Synthesized Hydrocarbons," ASTM International, West Conshohocken, PA United States, 2019.

[16] E. Jones and T. Oliphant and P. Peterson et al., "SciPy.org," 2001. [Online]. Available: https://www.scipy.org/. [Accessed 27 06 2019].

[17] A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques for Building Intelligent Systems, O'Reilly, 2017.

[18] G. Flora, S. T. Kosir, et al., "Properties Calculator and Optimization for Drop-in Alternative Jet Fuel Blends," in *SCITECH Forum*, San Diego, California, 2019.

[19] Dernoncourt, Franck & Young Lee, Ji, Optimizing Neural Network Hyper Parameters with Gaussian Processes for Dialog Act Classification, 2016.