Development of a Novel Machine Learning Methodology for the Generation of a Gasoline Surrogate Laminar Flame Speed Database under Water Injection Engine Conditions

20XX-01-XXXX

Development of a novel machine learning methodology for the generation of a gasoline surrogate laminar flame speed database under water injection engine conditions

Leonardo Pulga, Gian Marco Bianchi, Matteo Ricci, Giulio Cazzoli, and Claudio Forte
Affiliation (Do NOT enter this information. It will be pulled from participant tab in MyTechZone)

Abstract

The water injection is one of the technologies assessed in the development of new internal combustion engines fulfilling new emission regulation and policy on Auxiliary Emission Strategy assessment. Besides all the positive aspects about the reduction of mixture temperature at top dead centre and exhaust gases temperature at turbine inlet, it is well known that the water vapour acts as a mixture diluter, thus diminishing the reactants burning rate. A common methodology employed for the RANS CFD simulation of the reciprocating internal combustion engines turbulent combustion relies on the flamelet approach, which requires the knowledge of the laminar flame speed and thickness. Typically, these properties are calculated by mean of correlation laws, but they do not keep into account the presence of water mass fraction. A more precise methodology for the definition of both the laminar flame speed and thickness is thus required. The interrogation of a previously computed look-up table of such properties during run time seems to be a suitable and more accurate method than using correlations. In order to generate a database with all the possible combinations of chemical and physical properties that can be reached during the simulation of internal combustion engines, including the presence of a given mass fraction of water vapour and exhaust gases, a very high number of detailed chemical kinetics simulations needs to be performed. The present work aims to introduce a new methodology for the fast generation of laminar flame characteristics look-up tables that account also for the presence of water vapour in the reacting mixture. By using this new approach, engine designers will have the possibility to generate look-up tables of laminar flame characteristics for different fuels with the same computational cost that is currently required to generate a single table.

1. Introduction

1.1 Water injection in internal combustion engines

The water injection in reciprocating internal combustion engines is viewed by many researchers as one of the technical solutions to the critical issues introduced to reduce CO_2 and NO_x emissions, such as stoichiometric combustion, downsizing and higher compression ratios. In fact, most og these strategies lead to an increase in knock and pre-ignition tendencies inside the engine that must be carefully avoided, and to an increase in TiT (Turbine Inlet Temperature), in the case of a turbocharged powertrain [1].

Besides pure thermodynamics considerations and engine cycle analyses [2], the addition of water has a large impact on the combustion process. In particular, the addition of a diluent is expected to reduce the LFS (Laminar Flame Speed) of the reacting mixture, which is an essential property required to calculate the reaction rate in most combustion models such as: ECFM-3Z [3] and G-Equation [4], which are based on the flamelet assumption. State of the art combustion models, however, rely on experimental correlations for the LFS [5], that present several shortcomings with respect to the sensitivity to the chemical properties of the mixture, and do not keep in to account the presence of water vapour as a diluent. This is the reason why several authors [6][7] have proposed to refer to databases of LFS generated by mean of detailed chemical simulations, rather than to classical correlations. The generated datasets can be used during simulation by direct interpolation [6] or with the use of new correlations fitting the new data [7].

1.2 Aim of the activity

The computational cost required to generate a database of LFS at given conditions is directly proportional to the chemical kinetics scheme adopted, and to the number of single values (breakpoints) considered for each variable, i.e. Pressure, Temperature of the unburnt mixture, Equivalence ratio between fuel and fresh air, EGR (Exhaust Gas Recirculation) and Water

mass fractions. It is straightforward to notice that reducing the number of breakpoints even for only one variable can lead to a significant time reduction, but also to a lack in the accuracy of the method. The focus of the present work is to analyse the properties of a database of laminar flame speed, where the effect of water addition is accounted for, and define a new strategy for the more rapid development of new look-up tables, based on such observations. In particular, machine learning algorithms will be adopted to account for the effect of water vapour addition to the unburnt gas mixture definition, in order to strongly reduce the number of simulated points required to capture the relationship between LFS and water mass fraction.

2. Laminar flame speed modelling

2.1 Correlation law for laminar flame speed

The LFS of a reacting mixture is defined in literature [8] as a function of the physical properties P (Pressure) and T_u (Temperature of the unburnt gas) and chemical characteristics ϕ (Equivalence ratio) and X_{EGR} (mass fraction of the diluent) of the fresh mixture with the profile of Equation 1:

$$s_L = s_L^{\,0} \cdot \left(\frac{T}{T_0}\right)^{\alpha} \cdot \left(\frac{P}{P_0}\right)^{\beta} \cdot \left(1 - k \cdot X_{EGR}\right) \quad (1)$$

$$s_L^{\,0} = B_m - B_f \left(\phi - \phi_m\right)^2 \quad (2)$$

$$\alpha = \alpha_0 - \alpha_1 \left(\phi - 1\right) \quad (3)$$

$$\beta = \beta_0 + \beta_1 \left(\phi - 1\right) \quad (4)$$

where T_0 is the reference unburnt gas temperature and P_0 the reference pressure values at which $s_L^{\,0}$ was calculated (the values of T_0 and P_0 are usually taken at ambient conditions). The factor $k \cdot X_{EGR}$ is a correction term introduced to account for the presence of EGR as inert (k has values found in literature between 1.7 and 2.3 [8]). The value ϕ_m corresponds to the equivalence ratio at which the maximum LFS was found, while the other coefficients in Equations 2, 3, 4 must be modified as a function of the chosen fuel and correlation (usually Metghalchi and Keck [9], Heywood [8] or Gülder [10]).

2.2 Effect of the addition of EGR and water

Recently, a new correlation was proposed for keeping into account also the effect of water vapour as a diluent, based on detailed chemical simulations performed on relevant conditions that are reached during engine operation [11]. The results of the correlation, in terms of sensitivity to the diluent effect of water addition at relevant conditions, were adequately in agreement with the values of the detailed chemical simulations. On the other hand, it resulted that the use of a literature-standard power-law function for higher P and T_u values led to the prediction of unphysical values of LFS for ϕ greater than stoichiometric. The effect of the addition of water and EGR in reducing the LFS was, however, captured well by using a linear correlation, in the form of Equation 5:

$$\frac{s_L}{s_{L0}} = \left(1 - k_{EGR} \cdot X_{EGR}\right) \cdot \left(1 - k_{wat} \cdot X_{H2O}\right) \quad (5)$$

where s_L is the actual LFS, s_{L0} indicates LFS at the same P, T_u and ϕ but with $X_{EGR} = 0$ (EGR mass fraction) and $X_{H2O} = 0$ (water vapour mass fraction), and k_{EGR} and k_{wat} are the correlation coefficients.

3. Chemical simulations of laminar flames

3.1 Characteristics of the simulations

The detailed chemical simulation for the definition of laminar flame speed and thickness at each given condition is performed in a one-dimensional domain with successive mesh refinement. The free flame is considered adiabatic, planar and steady, reached by a mass flow of fresh mixture, whose velocity corresponds to the displacement speed of the reaction zone [5]. Since the simulation is performed for an unstrained steady flame, the displacement speed can be considered corresponding to the reaction speed, and therefore, its value represents the LFS required for the combustion models [5]. The definition of the species that constitute the unburnt mixture is performed based on the equivalence ratio between fuel and air (composed only by O_2 and N_2) and on the presence of EGR and water vapour addition. The composition of EGR is calculated based on a complete

stoichiometric combustion, therefore from a combination of O_2, N_2 and H_2O whose mass fractions only depend on the fuel formulation.

The adopted methodology for the database generation and the chemical simulations is based on the work by Cazzoli et al. [6], which relies on the Cantera [12] implementation in Python and requires to choose other three key aspects of the simulations: the chemical kinetics mechanism, the fuel surrogate and the breakpoints for all the accounted variables.

3.2 Choice of the chemical kinetics scheme

The selection of the chemical kinetics scheme must keep into account three parameters:

- the presence of reactions for all the species of the chosen surrogate;
- the number of species and reactions considered, under the point of view of the completeness of the mechanism
- the computing time required to run a simulation.

The chemical kinetics schemes considered in the present work for the generation of the database are reported in Table 1 and they were developed for simulating the reacting features of gasoline under both high and low temperature conditions. The observations reported by Cazzoli et al. [11], and an initial testing to perform a time estimate for the simulations have led to the decision of referring the present activity to results obtained with the complete POLIMI [13] scheme.

The data related to computing times are obtained with a benchmark performed on ten successive simulations on a computer with Intel Xeon Platinum with 3.0 GHz, 36 cores, 144 GB ram, Cantera version 2.3.0 and Python version 3.7.2. The time required for the computation is proportional to the number of species and reactions present in the chemical kinetics mechanism, but also to the physical and chemical properties of the simulated point, since the simulation might require more iterations to converge. On average, a set of simulations performed using the POLIMI scheme [14] has required 1000 seconds per point.

Table 1 Description of the tested chemical kinetics mechanisms.

	Time/Sim	Species	Reactions
POLIMI red. [14]	0.15 x REF	156	3465
POLIMI [13]	REF	451	8153
LLNL [15]	>3 x REF	1387	10481

The LLNL mechanism [15] would have required unaffordable computing resources, while the POLIMI reduced scheme [14] was found to produce less accurate results with respect to experimental data, in terms of laminar flame speed, as reported in Figure 1.

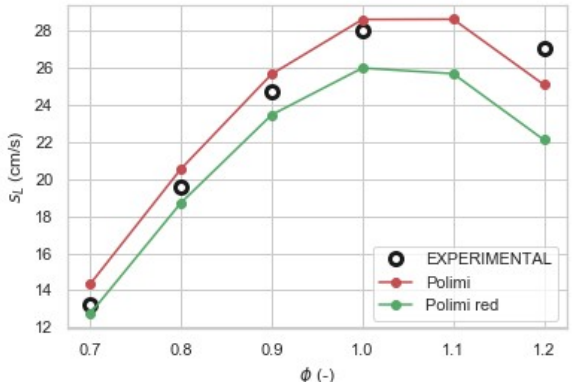

Figure 1 Comparison of Polimi and Polimi reduced mechanisms with experimental data [15] for the prediction of LFS of PRF87 at P=20 bar, T=353 K.

3.3 Choice of the surrogate fuel

Since the aim of the activity is to generate a database of LFS for the simulation of gasoline combustion, the analysis is mainly conducted on the fuel surrogate named *TAE7000* (composition reported in Table 2) which was experimentally designed to provide similar values of reaction speed to a TOTAL commercial gasoline [16].

For validating the robustness of the methodology, with respect to the adopted surrogate, it was chosen to perform the same computations also for other two fuels with different composition, namely PRF87 [17] and TRF95-2 [18], described in Table 2.

Table 2 Description of the composition of the analysed surrogates.

SURROGATE	COMPOSITION	RON
TAE_7000 [16]	i-C_8H_{18} (42.9%) n-C_7H_{16} (13.7%) C_7H_8 (43.4%)	95.1
PRF87 [17]	i-C_8H_{18} (87.0%) n-C_7H_{16} (13.0%)	87
TRF95-2 [18]	i-C_8H_{18} (82.04%) n-C_7H_{16} (7.96%) C_7H_8 (10%)	95

The values of LFS for a reference operating point and different equivalence ratios are reported in Figure 2, to emphasise the differences between the chosen fuel surrogates.

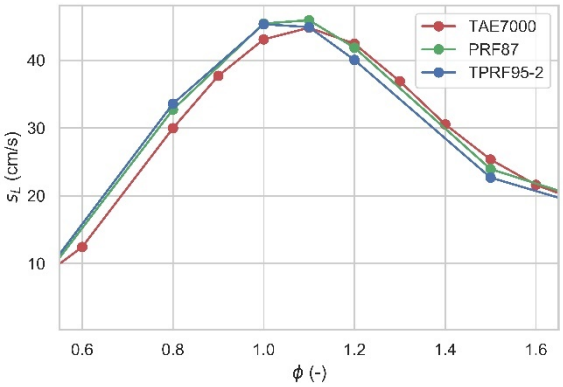

Figure 2 LFS of the different fuel surrogates at P=50 bar, T=600 K, EGR=0% and XH2O=0%.

3.4 Choice of the simulated points

A complete LFS database must present values for all the conditions that can be encountered during the simulation of the combustion process in terms of P, T_u, ϕ and mass fraction of EGR and Water. Since the range of mixture equivalence ratio inside the combustion chamber might be wide, and that the look-up table should be applicable to both low load and full load engine conditions, the simulations were performed for all the combinations of the five variables and their limits are reported in Table 3. The number of break points for each variable was chosen to adequately capture the main trends, with a focus on those regarding the mass fraction of each mixture component.

Table 3 Description of the breakpoints used for generating the dataset of LFS.

	Min	Max	# Points
¿	4	290	8
P ¿			
$T_{un}(K)$	358	700	5
− ¿	0.3	1.8	14
ϕ ¿			
$X_{egr}(\%)$	0	20	10
$X_{H2O}(\%)$	0	6	7

The water mass fraction is limited to 6% because it was reported in [1] that, during engine operations, it is not feasible to use a higher amount of water than of fuel at stoichiometric conditions, mainly because of the evaporation times of the liquid droplets. The parameter *s*, defined as the ratio of injected water mass and stoichiometric fuel quantity should remain under 1.0, which corresponds, assuming C_8H_{18} as fuel representative, to a water mass fraction of approximately 6.23%, not including the EGR mass fraction.

4. Effect of water addition on LFS

The analysis reported in this section refers to the simulations performed on the *TAE7000* surrogate but can be qualitatively extended to the other fuels. It can be noticed that the effect of the addition of water vapour to the unburnt mixture definition leads to a decrease in terms of LFS. This effect, as reported for two combinations of P and T_u in Figure 3 and 4, is higher for near stoichiometric reacting mixtures. This behaviour was found also in literature [19] and was represented by Cazzoli et al. [11] as a linear dependence on the water mass fraction.

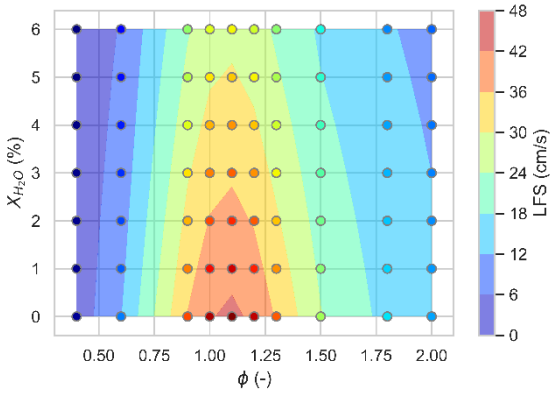

Figure 3 LFS function of Phi and X_{H2O} at P=50 bar, T=600 K and EGR=1%.

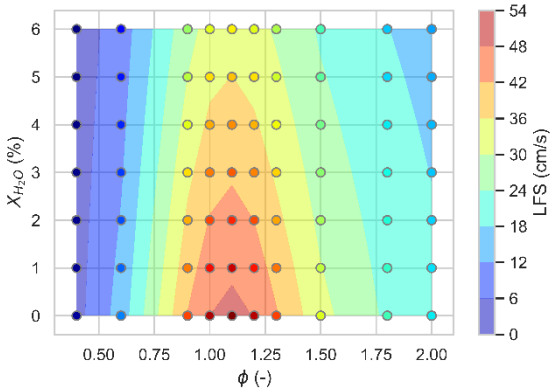

Figure 4 LFS function of Phi and X_{H2O} at P=90 bar, T=700 K and EGR=3%.

4.1 Evaluation of the hypothesis of linear dependence

From Equation 5, the effective value of k_{wat} can be calculated from the simulated values of laminar flame speed, as in Equation 6, where s_L represents the actual LFS with a mass fraction of water vapour equal to X_{H2O} and s_{L0} the LFS at the same conditions of P, T_u, ϕ and X_{egr} without water addition:

$$k_{wat} = \left(1 - \frac{s_L}{s_{L0}}\right) / X_{H2O} \qquad (6)$$

In Figures 5, 6 the behaviour of the actual k_{wat} calculated from Equation 6 is represented, with respect to water mass fraction and equivalence ratio at specific conditions of P, T_u and EGR. The two surfaces are qualitatively representative of the behaviour of the full database and are characterized by:

- a local constant value, near stoichiometric conditions and values of X_{H2O} smaller than 4%;
- a value function only of ϕ for rich conditions (i.e. $\phi \geq 1.5$);
- a value function of both ϕ and X_{H2O} for lean conditions (i.e. $\phi \leq 0.75$).

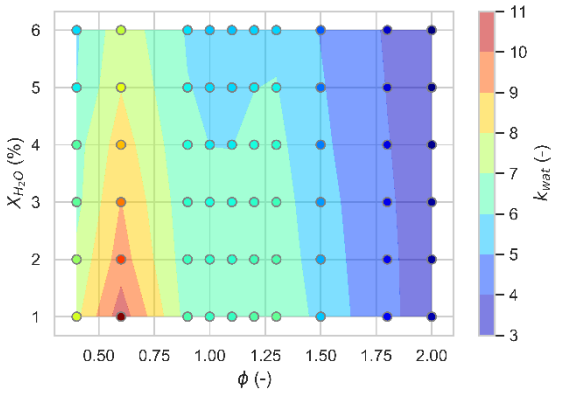

Figure 5 k_{wat} function of Phi, X_{H2O} at P=50 bar, T=600 K, EGR=1%.

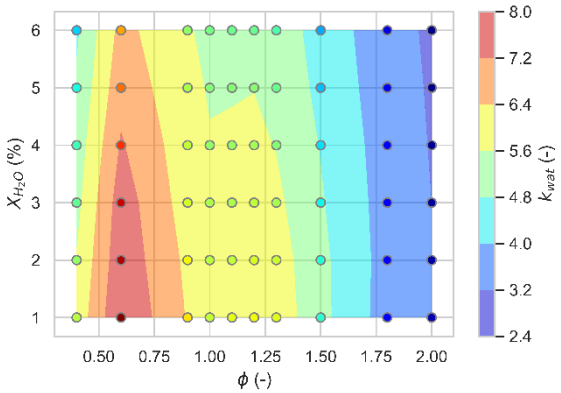

Figure 6 k_{wat} function of Phi, X_{H2O} at P=90 bar, T=700 K, EGR=3%.

This behaviour underlines how the hypothesis of a linear correction to account for the presence of water in the reacting mixture represents a good approximation, if the mixture equivalence ratio is near stoichiometric and the value of the water mass fraction is limited to 4%. The distribution of the relative error committed in calculating the LFS on the full database, using a single value of k_{wat} with respect to the output of the simulations is reported in Figure 7. It can be noticed that the peak of the distribution is near 0, which derived from the choice of the best fitting value of k_{wat} , but in more than 50% of the simulated points the effect of water addition is overestimated.

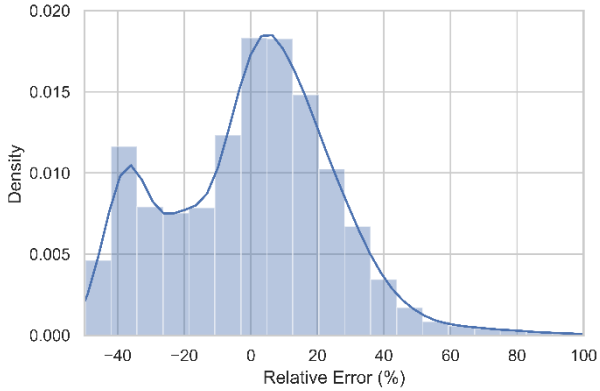

Figure 7 Distribution of the relative error committed on the evaluation of k_{wat} considering a fixed value.

5. Strategies considered for generating laminar flame speed lookup table

5.1 Computational cost analysis

As reported in Section 3, the dataset of LFS for applications to Internal Combustion Engine simulations must represent all the combinations of the five variables that the flame front can encounter during combustion. A first approach would be to perform detailed chemical simulations for the full array of combinations, which corresponds to 39200 simulations, for the number of reference points reported in Table 3. Several hypotheses can be assumed in order to reduce the time requirements for the generation of the dataset:

- some combinations of P and T_u and EGR are unrealistic when applied to engine operating conditions [7], thus the calculation of the LFS in those points (high pressure and low temperature, or the opposite) can be omitted,
- the solution can be reached in a fraction of the computing time if using a reduced chemical kinetics scheme.

The first hypothesis would reduce the computing time by a limited fraction, while on the other hand, the second option would lead to a reduced accuracy of the results. One additional assumption can be applied, considering the effect of water to have a linear relation with its mass fraction, in the case of the generation of an LFS dataset accounting for the presence of water vapour in the reacting mixture. In this way, just two different values of the variable X_{H2O} would be required, in order to compute the value of k_{wat} to be used for the definition of the generation of the intermediate combinations. This assumption would reduce the number of necessary simulated points by more than 70%, but would induce several errors, like those reported in Figure 7.

A different approach will be outlined in next sections, where the effect of the water addition is predicted by mean of machine learning algorithms trained on reference conditions. The aim of employing such methodologies is to further reduce the time requirements to generate the dataset while maintaining a high level of accuracy.

5.2 Integration of machine learning algorithm into the workflow

The integration between machine learning algorithms and the dataset of LFS with the aim of reducing the number of simulations required for the its generation can be performed following several strategies. In particular, it was chosen to focus on two of them, which require the generation of a look-up table of LFS for all values of P, T_u, ϕ, X_{EGR} without accounting for the addition of water, and a limited number of additional reference points where the mixture contains a fraction of water vapour. The additional computing effort required for the calculation of the LFS at the reference points with water vapour is expected to represent a fraction of the simulation time of the full dataset, leading to a time reduction of approximately 80%.

The analysed strategies are summarised in Equations 6, 7, where $\widetilde{s_L}$ represents the predicted LFS, s_{L0} is the value of LFS at $X_{H2O}=0$, η_{ML} is the machine learning predicted effect of the water mass fraction and k_{ML} is the proportionality coefficient of the effect of X_{H2O}, also predicted by a machine learning model.

$$\widetilde{s_L} = s_{L0} \cdot \eta_{ML}\left(P, T_u, \phi, X_{EGR}, X_{H2O}\right) \quad (6)$$

$$\widetilde{s_L} = s_{L0} \cdot \left(1 - k_{ML}\left(P, T_u, \phi, X_{EGR}, X_{H2O}\right) \cdot X_{H} \quad (7)$$

Even if the algorithms to be followed for the application of both methodologies are similar, the ideas underlying each of them are deeply different. In fact, in strategy #1, the effect of the water addition is entirely predicted by the algorithm, which must, therefore, be reliable on the full dataset extension. This aspect implies that, for example, in the case of the absence of water mass fraction the algorithm might still provide values that are slightly lower than the actual LFS. Obviously, this condition can be avoided by performing an adequate training of the machine learning model, but the results in the case of low water mass fractions might still be underestimated.

As far as the strategy #2 is concerned, the additional hypothesis of a linear correlation of LFS with X_{H2O}, as reported in Section 4, is applied. Differently from the previous works [11] the proportionality coefficient k_{ML} is defined as a function of the operating characteristics and of the diluent mass fractions. Even if the effect of X_{H2O} is considered to be linearly dependent, it was necessary to include also the water mass fraction as a variable for the definition of the proportionality coefficient, in order to capture the non-linearities described in Section 4, for lean mixtures and values of X_{H2O} greater than 4%.

By analysing the values of η_{ML} and k_{ML} for the full dataset available, reported in Figure 8 for their normalized values, it emerges how the distribution of the former looks widespread through the full range of possible values, while the latter is more normally distributed, with two distinct peaks. These distributions anticipate the fact that a regression algorithm should be capable of predicting more efficiently the values of k_{ML} than those of η_{ML}.

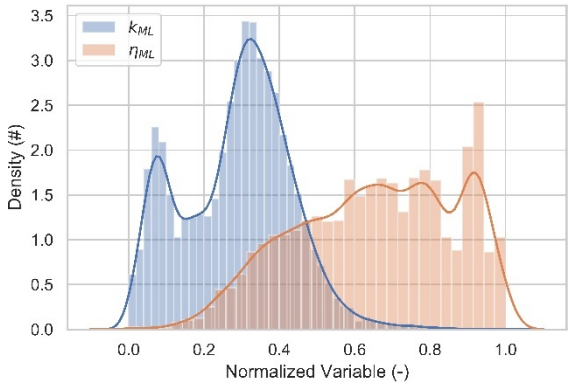

Figure 8 Distribution of the normalized values of k_{ML} and η_{ML} for the full dataset of TAE7000.

6. Definition of the best machine learning model

The results reported in the following sections refer to the best performance obtained from a machine learning algorithm chosen between a list of available models and tuned to enhance its capabilities. In particular, the algorithm should present the following characteristics:

- it must be capable of performing a supervised regression, since it will be trained on available data and it must provide a continuous value,
- it should not be prone to overfitting, since some data may present very different values from the main trends.

The list of the analysed algorithms contains standard regression models, such as Multivariate Linear Regression [20], Regression tree [20], Gaussian Processes for regression [21], Support Vector Regression with both linear and non-linear kernels; ensemble methods such as Random Forest [22], AdaBoost [23] and Gradient Boost [24]; and Neural Networks with different architectures and activation functions [25].

The regression models are applied on a previously processed version of each dataset. In particular, the method followed for the definition of the most suitable algorithm is the following:

- all variables are normalized with mean value set to 0 and standard deviation set to 1,
- all points of the dataset where the effect of the addition of water is considered unphysical (but possibly related to numerical reasons, i.e. those far from the mean value more than 3 times the standard deviation) are removed,

- each regression model, with several values for each tuning coefficient is cross-validated using the k-fold approach, which consists in randomly dividing the dataset into k groups of approximately equal size and perform validation on one set and training on the remaining data [20].
- after enough repeated cross validations, the model and relative tuning coefficients providing the best accuracy is chosen for the application.

After performing this procedure, the most suitable machine learning algorithm resulted to be AdaBoost, using the parameters reported in Table 4.

Table 4 Description of the main tuned features of AdaBoost.

Base Learner	Regression Tree
Feature of base learner	Maximum depth = 10
# of estimators	50
Learning rate	0.8

The AdaBoost (Adaptive Boost) algorithm is an ensemble method aimed at combining the outputs of a series of weak models (base learners) to perform a better prediction. This is achieved by consequently training the base learners on a weighted version of the training set, where more importance is given to the values that the previous base learners could not capture adequately.

The Regression Tree, used as base learner, is a supervised regression model, based on a tree-like graph where each node indicates a condition on one attribute, and the external nodes, called leaves, represent the output of the model. The choice of the regression tree as base learner for AdaBoost is quite common [23] because of its simplicity, while the other coefficients were defined on the ground of the best performance during the cross-validation.

7. Results with strategy #1

7.1 Results on TAE7000

The available dataset of LFS for TAE7000 was generated for a large sample of the total combinations, consisting of 17000 points, obtained by reducing the number of simulated points with EGR mass fraction higher than 6%.

The performance of the methodology, reported in Table 5 was evaluated by splitting the database into a training set and a test set, with different proportions. The training set is used to fit the coefficients of the machine learning model, while the test set is used to assess the performance of the predictions with respect to the real values. The split is performed randomly, but in order to maintain a distribution in the target values (η_{ML}) similar to that of the full database for both sets. The accuracy of the methodology, reported in Table 5, is assessed with three metrics:

- MAE (Mean Absolute Error), which is the mean absolute difference between the target and the predicted value;
- RMSE (Root Mean Square Error) which is the square root of the mean squared difference between the target and predicted value;
- R^2 (determination coefficient) which is calculated in its general form as in Equation 8, where y_i is the target value, $\dot{y}$ is the mean observed data, and f_i is the predicted value:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \dot{y})^2} \quad (8)$$

All the metrics are calculated for both sets on the target value of the machine learning model (η_{ML}) and for the LFS obtained by inverting Equation 6.

The results demonstrate that a high level of accuracy in the determination of the LFS can be reached by using only 5% of the available database as training set, and the performance on a test set composed of more than 10000 points shows a value of R^2 higher than 0.99 and MAE lower than 2 mm/s, which is below 2% of the mean LFS of the database.

In Figure 9, the accuracy in the prediction of η_{ML} is reported for train and test set, further underlying that the test set, composed by 95% of the available dataset is fully captured by the model, and there are no particular regions of interest where the model is less predictive.

Table 5 Determination coefficients for train and test sets with different splits.

	TRAIN SET	TEST SET
Fraction	0.5	0.5
R^2 for η_{ML} (-)	9.99E-1	9.99E-1
R^2 LFS (-)	9.99E-1	1.00
MAE k_{ML} (-)	1.68E-3	2.6E-3
MAE LFS (m/s)	3.75E-4	3.85E-4
RMSE k_{ML} (-)	2.27E-3	3.62E-3
RMSE LFS (m/s)	3.20E-5	3.91E-5
Fraction	0.2	0.8
R^2 for η_{ML} (-)	9.99E-1	9.99E-1
R^2 for LFS (-)	1.00	9.99E-1
MAE k_{ML} (-)	1.34E-3	3.80E-3
MAE LFS (m/s)	4.93E-4	5.81E-4
RMSE k_{ML} (-)	1.96E-3	5.87E-3
RMSE LFS (m/s)	8.37E-4	1.09E-3
Fraction	0.05	0.95
R^2 for η_{ML} (-)	1.00	9.83E-1
R^2 for LFS (-)	1.00	9.99E-1
MAE k_{ML} (-)	5.83E-4	7.81E-3
MAE LFS (m/s)	2.78E-4	1.16E-3
RMSE k_{ML} (-)	1.29E-3	1.28E-2
RMSE LFS (m/s)	3.45E-4	1.97E-3

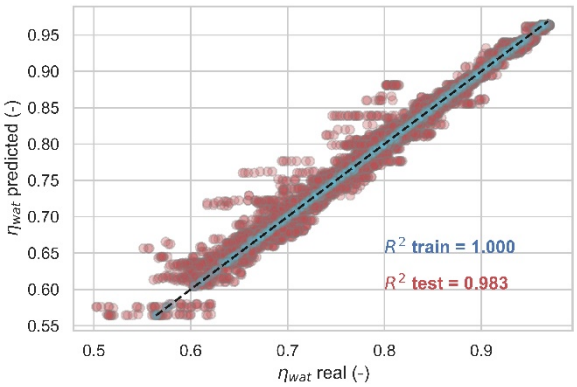

Figure 9 Scatter plot of the values of η_{ML} for TAE7000.

7.2 Results with other reference fuels

In order to further validate the presented methodology, the same process has been applied on reduced forms of the dataset generated for other reference fuels, described in Table 2. The distribution of the simulated points has been varied, focusing more attention on less and more spread values of EGR (0%, 5%, 10%, 30%), and X_{H2O} (0%, 1%, 3%, 6%) but more breakpoints of P, T_u and ϕ especially in the range usually reached in internal combustion engine applications. The total number of simulated points is 2100, and the results reported refer to a split of the database that lead to a test set composed by 75% of the full dataset (i.e. fitting of the model performed on about 500 points, which is the same size of the train set used for TAE7000).

The results on PRF87 are displayed in Figure 10, reporting the accuracy of the machine learning predictions on the target variable η_{ML}, while in Figure 11 the same comparison is reported for the surrogate TRF95-2. In both cases the

performance is aligned with the results obtained with TAE7000 and the results in terms of determination coefficient of the prediction of the actual value of LFS are higher (R^2 test = 0.985 for LFS of PRF87 and R^2 test = 0.981 for LFS of TRF95-2)

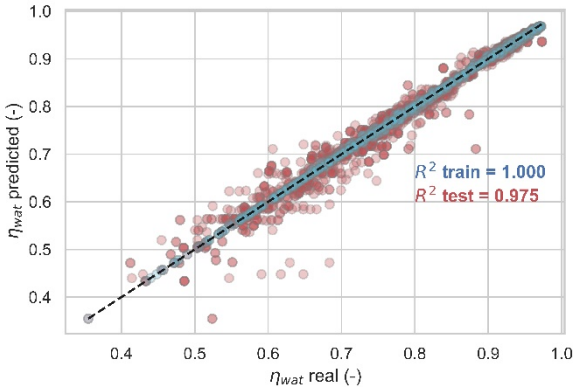

Figure 10 Scatter plot of the values of η_{ML} *for PRF87.*

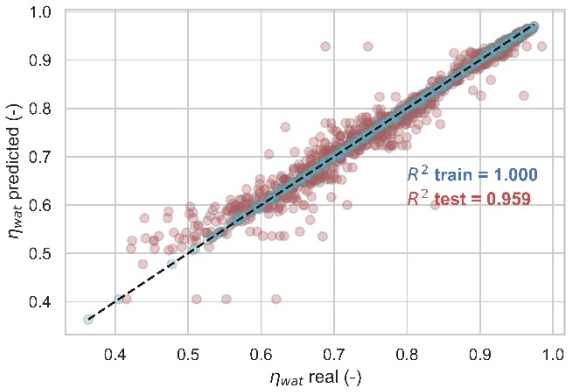

Figure 11 Scatter plot of the values of η_{ML} *for TPRF95-2.*

8. Results with strategy #2

8.1 Results on TAE7000

The same machine learning algorithm and tuning coefficients described in Section 7 resulted to be the best performing also for Strategy #2. The results of determination coefficient, MAE and RMSE for the target variable (k_{ML}) and for the LFS obtained by splitting the database into train and test sets with different proportions are reported in Table 6. The performance of the algorithm is similar to that reached for methodology #1 and provides extremely positive results in terms of accuracy in predicting the LFS values, with no particular region of error, as displayed in Figure 12.

Table 6 Determination coefficients for train and test sets with different splits.

	TRAIN SET	TEST SET
Fraction	0.5	0.5
R^2 k_{ML}(-)	9.98E-1	9.91E-1
R^2 LFS (-)	1.00	9.99E-1
MAE k_{ML}(-)	4.38E-2	8.22E-2
MAE LFS (m/s)	2.71E-4	4.08E-3
RMSE k_{ML}(-)	5.92E-2	1.31E-1
RMSE LFS (m/s)	5.37E-4	7.50E-4
Fraction	0.2	0.8
R^2 for k_{ML}(-)	9.99E-1	9.86E-1
R^2 for LFS (-)	1.00	9.99E-1

MAE k_{ML}(-)	2.96E-2	1.02E-1
MAE LFS (m/s)	2.41E-4	4.63E-4
RMSE k_{ML}(-)	4.48E-2	1.62E-1
RMSE LFS (m/s)	7.30E-4	8.09E-4
Fraction	0.05	0.95
R^2 for k_{ML} (-)	1.00	9.66E-1
R^2 for LFS (-)	1.00	9.98E-1
MAE k_{ML} (-)	1.32E-1	1.66E-3
MAE LFS (m/s)	1.96E-4	7.98E-4
RMSE k_{ML} (-)	2.79E-2	2.63E-1
RMSE LFS (m/s)	4.92E-4	1.43E-3

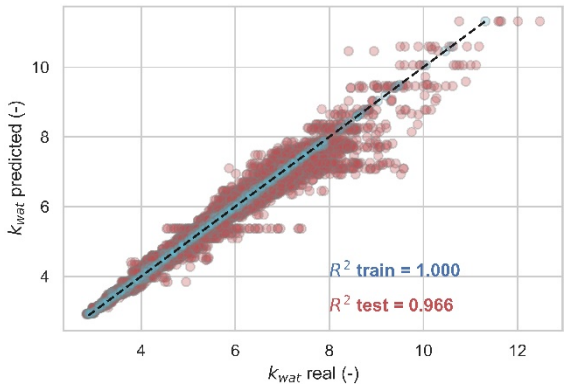

Figure 12 Scatter plot of the values of k_{ML} *for TAE7000.*

8.2 Results with other reference fuels

The application on the other reference fuels followed the same steps described in Section 7, and the results in the prediction of k_{ML} for each surrogate are reported in Figure 13 and 14. In both cases, the accuracy of the prediction is worse than for the prediction of η_{ML}, however, the fact that the process of inverting Equation 7 does not only rely on the prediction of the machine learning algorithm, but also on the knowledge of X_{H2O}, the overall performance in predicting the LFS of methodology #2 is slightly better (R^2 test = 0.989 for LFS of PRF87 and R^2 test = 0.985 for LFS of TRF95-2).

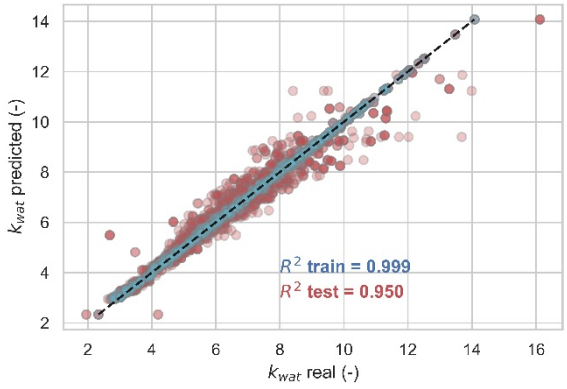

Figure 13 Scatter plot of the values of k_{ML} *for PRF87.*

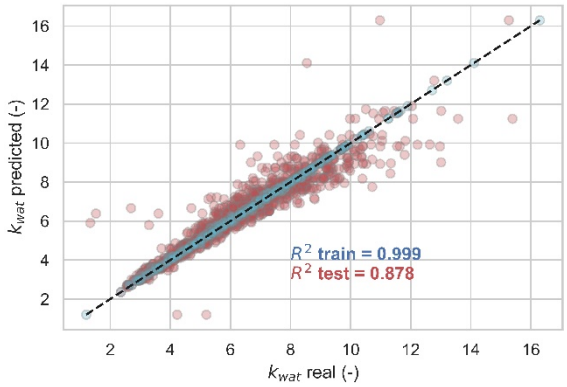

Figure 14 Scatter plot of the values of k_{ML} *for TPRF95-2.*

9. Output of the two strategies

As reported in Section 7 and 8, both strategies can generate values of LFS that account for the presence of water vapour in the reacting mixture with an absolute relative error below 4% for all surrogate fuels. As showed in Figure 15, the methodology that employs k_{wat} performs slightly better than the other, thanks to the fact that it relies also on the physical interpretation of the behaviour of the LFS, with an imposed linear correlation, and not only on the pure machine learning prediction.

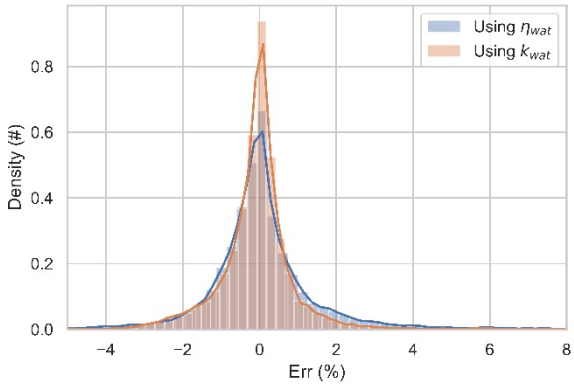

Figure 15 Distribution of the relative error committed by both methodologies on the full dataset for TAE7000 surrogate.

A further test has been performed, to investigate the applicability of the machine learning algorithm trained on the dataset generated for TAE7000 on the points simulated with the other two fuel surrogates. This strategy, referred to as transfer learning [24] might lead to a further reduction in the computational cost to generate a full look-up table, since it would remove the requirements to perform any simulation with $X_{H2O}>0$.

The results, however, as reported in Figure 16 for the PRF87 and Figure 17 for the TRF95-2, show that the absolute relative error committed in predicting the values of LFS with a given water mass fraction can reach values up to 20%. These results confirm the need to perform the fitting phase of the model for each surrogate on enough training points, in order to obtain a better performance with both methodologies.

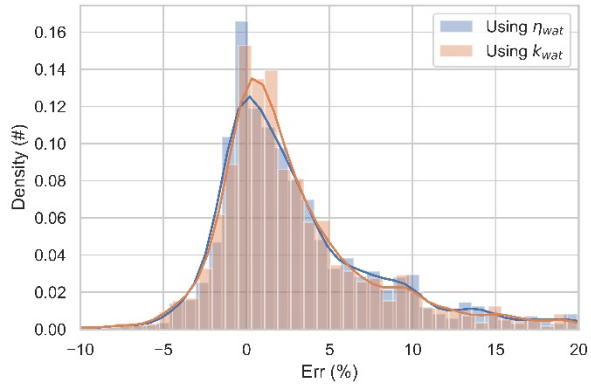

Figure 16 Distribution of the relative error committed by applying transfer learning for both methodologies on PRF87 surrogate.

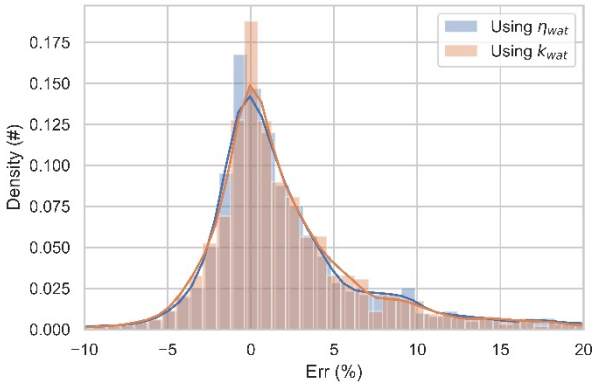

Figure 17 Distribution of the relative error committed by applying transfer learning for both methodologies on TPRF95-2 surrogate.

10. Focus on laminar flame thickness

9.1 Definition of laminar flame thickness

The application of a flamelet combustion model based on the flame surface density transport requires not only the knowledge of the LFS, but also of the LFT (laminar flame thickness), in order to account for the efficiency of the turbulent vortices to wrinkle the flame [5]. The definition of LFT in literature has several meanings, but many authors agree that the most useful in combustion modelling is the thermal thickness [5], which is derived from the temperature profile inside the reaction zone (represented in Figure 18) as defined in Equation 9, where T_2 represents the temperature of the burnt gases and T_1 is the temperature of the fresh mixture.

$$\delta_L{}^0 = \frac{T_2 - T_1}{max\left(\left|\frac{dT}{dx}\right|\right)} \qquad (9)$$

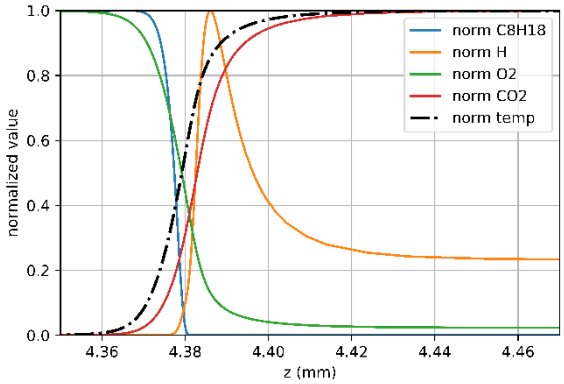

Figure 18 Normalized profile of temperature and mass fractions of the main components of the reacting mixture near the flame front in a simulated flat flame.

A correlation based on scaling laws [27] was introduced to overcome the lack of experimental data regarding the LFT, as proposed in Equation 10, calculated with the properties of the fresh mixture, where λ is the thermal conductivity of the gas, ρ is its density, C_P the specific heat at constant pressure and s_L is the laminar flame speed in that condition.

$$\delta_L{}^0 = \frac{\lambda}{\rho C_P s_L} \tag{10}$$

Blint [27] corrected the previous correlation by introducing a correction factor based on the burnt gas temperature, leading to Equation 11, where subscript 1 indicates that the property refers to the fresh mixture, and subscript 2 is the condition of the burnt zone.

$$\delta_L{}^{Blint} = \delta_L{}^0 \frac{\left(\lambda/C_P\right)_2}{\left(\lambda/C_P\right)_1} \tag{11}$$

9.2 Applicability of Blint's correlation

Since the value of the LFT in turbulent combustion modelling is extremely important, the accuracy of the Blint's correlation reported in Equation 11 was evaluated under engine relevant conditions, with the presence of EGR and water vapour. It was chosen to focus only on the data available for the TAE7000 surrogate fuel, and the results from a correlation analysis are reported in Table 7, as a function of the water mass fraction for values in the range $4 < P < 140$, $300 < T_u < 700$, $0.5 < \phi < 1.6$ and $0\% < X_{EGR} < 30\%$.

Table 7 Slope and determination coefficient of the linear correlation between the two possible values of LFT.

X_{H2O} (%)	Slope	R^2	MAE (m)	RMSE (m)
0	0.626	9.91E-1	2.11E-5	5.07E-5
1	0.624	9.93E-1	1.53E-5	3.27E-5
2	0.624	9.92E-1	1.65E-5	3.53E-5
3	0.623	9.92E-1	1.78E-5	3.79E-5
4	0.623	9.91E-1	1.93E-5	4.09E-5
5	0.622	9.90E-1	2.07E-5	4.38E-5
6	0.623	9.90E-1	2.23E-5	4.69E-5

The slope is the proportionality coefficient between the calculated thermal thickness and that obtained from the temperature profile, which can be used as a constant of proportionality. Besides the requirement for a scaling factor, the addition of water does not introduce a source of error in the Blint's correlation, which can be employed with the relative error distribution reported in Figure 19.

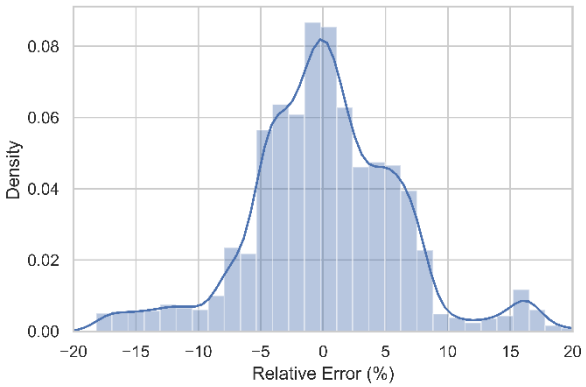

Figure 19 Relative error distribution between scaled values of LFT calculated with Blint's correlation and from temperature profile.

11. Conclusion

With the present work, the effect of the water vapour on the laminar flame speed of reacting mixtures has been evaluated from the results provided by detailed chemical kinetics simulations. This effect can be well captured with a linear correlation, within engine relevant conditions and for a given range of ϕ and X_{H2O}, but it has been showed that the linear coefficient must be a function of the operating condition (P, T_u, ϕ, X_{EGR}) and not a constant value.

A regression machine learning algorithm can be fruitfully applied to predict the effect of the water addition, or the value of k_{wat} after a proper training phase, performed on a limited number of simulated points. An analysis on the best algorithms and strategies has been performed, with the aim of defining a new methodology for the more rapid generation of a full LFS database accounting also for the effect of the water vapour. The final proposal can reduce by 81.5% the time required to produce a full look-up table, maintaining the relative error committed below 4%, as reported in Figure 15. When compared with the strategies presented in section 5, the proposed methodology results more time reducing and accurate than limiting the size of the dataset or using a reduced chemical kinetics scheme. However, a combination of these methods would lead to a further reduction in the time required to generate the full database.

This methodology has been tested on a reduced version of the LFS database for different surrogate fuels obtaining similar results in terms of accuracy, which demonstrates the applicability of this approach to other fuel surrogates.

An additional assessment has been performed on the definition of laminar flame thickness, which is another essential property of the reacting mixture in turbulent combustion models based on the flamelet approach. The value of the LFT can be derived from the temperature profile of the detailed chemical simulation, or it can be calculated using Blint's correlation, which requires only the knowledge of the properties of the mixture. Even though the correlation was developed for reference conditions [5], its validity has been verified also for engine relevant pressure and temperature values and in cases where the mass fractions of EGR and water vapour are relevant.

12. Future developments

The presented methodology can allow researchers and engineers to efficiently generate new databases of laminar flame speed under water injection conditions for a variety of fuel surrogates, from which just a few have been analysed. The effect of the application of the look-up tables generated with this methodology, in place of classic correlations for engine combustion simulations, must be tested and future work should focus on the effect of the choice of the fuel surrogate on engine performance.

BIBLIOGRAPHY

[1] Falfari, S., Bianchi, G. M., Cazzoli, G., Forte, C., Negro, S., "Basics on Water Injection Process for Gasoline Engines", Energy Procedia, Volume 148, 2018, Pages 50-57, DOI:10.1016/j.egypro.2018.08.018.

[2] Cavina, N., Rojo, N., Businaro, A., Brusa, A., Corti, E., De Cesare, M., “Investigation of water injection effects on combustion characteristics of a gdi tc engine.” SAE International Journal of Engines 2017;10(4). DOI:10.4271/2017-24-0052.

[3] Colin O., Benkenida A. “The 3-Zone Extended Coherent Flame Model (ECFM3Z) for computing premixed/diffusion combustion”, Oil Gas Sci. Technol. -Rev. IFP 59, 6, 593-609.

[4] Dekena, M., Peters, N., “Combustion modeling with the G-equation”, Oil Gas Sci. Technol. 54 (1999) 265–270

[5] Poinsot, T., Veynante, D., “Theoretical and Numerical Combustion”, 2001, Edwards, 9781930217058

[6] Cazzoli, G., Forte, C., Bianchi, G., Falfari, S., Negro, S., "A Chemical-Kinetic Approach to the Definition of the Laminar Flame Speed for the Simulation of the Combustion of Spark-Ignition Engines", SAE Technical Paper 2017-24-0035, 2017, DOI: 10.4271/2017-24-0035.

[7] Del Pecchia, M., Breda, S., D'Adamo, A., Fontanesi, S. et al., "Development of Chemistry-Based Laminar Flame Speed Correlation for Part-Load SI Conditions and Validation in a GDI Research Engine", SAE Int. J. Engines 11(6):715-741, 2018, DOI: 10.4271/2018-01-0174.

[8] Heywood, J. B., 1988. “Internal Combustion Engine Fundamentals”, McGraw-Hill, New York

[9] Metghalchi, M., Keck, J., 1982. “Burning velocities of mixtures of air with methanol, isooctane, and indolene at high pressure and temperature”. Combustion and Flame, 48(C), pp. 191-210.

[10] Gülder, Ö., 1982. “Laminar burning velocities of methanol, ethanol and isooctane-air mixtures”. Symposium (International) on Combustion, 19(1), pp. 275-281.

[11] Cazzoli, G., Falfari, S., Bianchi, G. M., Forte, C., “Development of a chemical-kinetic database for the laminar flame speed under GDI and water injection engine conditions”, Energy Procedia, Volume 148, 2018, Pages 154-161, DOI:10.1016/j.egypro.2018.08.043.

[12] Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. http://www.cantera.org

[13] Ranzi, E., Frassoldati, A., Grana, R., Cuoci, A., Faravelli, T., Kelley, A.P., Law, C.K., “Hierarchical and comparative kinetic modelling of laminar flame speeds of hydrocarbon and oxygenated fuels”, Progress in Energy and Combustion Science, 38 (4), pp. 468-501 (2012), DOI:10.1016/j.pecs.2012.03.004

[14] Ranzi, E., Frassoldati, A., Stagni, A., Pelucchi, M., Cuoci, A., Faravelli, T., “Reduced kinetic schemes of complex reaction systems: Fossil and biomass-derived transportation fuels”, International Journal of Chemical Kinetics, 46 (9), pp. 512-542 (2014), DOI:10.1002/kin.20867

[15] Mehl M., Pitz, W.J., Westbrook, C.K., Curran, H.J., "Kinetic modeling of gasoline surrogate components and mixtures under engine conditions", Proceedings of the Combustion Institute 33:193-200 (2011).

[16] Dirrenberger, P., Glaude, P.A., Bounaceur, R., Le Gall, H., Pires da Cruz, A., et al., “Laminar burning velocity of gasolines with addition of ethanol”, Fuel, Elsevier, 2014, 115, pp.162-169. DOI: 10.1016/j.fuel.2013.07.015

[17] S. Jerzembeck, N. Peters, P. Pepiot-Desjardins, H. Pitsch, “Laminar burning velocities at high pressure for primary reference fuels and gasoline: Experimental and numerical investigation”, Combustion and Flame, Volume 156, Issue 2, 2009, Pages 292-301, ISSN 0010-2180, DOI:10.1016/j.combustflame.2008.11.009

[18] Mannaa, O., Mansour, M. S., Roberts, W. L., Chung, S. H., “Laminar burning velocities at elevated pressures for gasoline and gasoline surrogates associated with RON”, Combustion and Flame, Volume 162, Issue 6, 2015, Pages 2311-2321, ISSN 0010-2180, DOI:10.1016/j.combustflame.2015.01.004

[19] Mazas, A., Fiorina, B., Lacoste, D., Schuller, T., “Effects of water vapor addition on the laminar burning velocity of oxygen-enriched methane flames”. Combustion and Flame 2011;158(12):2428–2440. DOI:10.1016/j.combustflame.2011.05.014

[20] Hastie, T., Tibshirani, R., Friedman, J., “The elements of statistical learning: data mining, inference and prediction”, Springer, 2009.

[21] Rasmussen, C. E., Williams, C. K. I., “Gaussian Processes for Machine Learning”, MIT Press 2006

[22] Geurts, P., Ernst., D., Wehenkel, L., “Extremely randomized trees”, Machine Learning, 63(1), 3-42, 2006

[23] Zhu, J., Zou, H., Rosset, S., Hastie, T., “Multi-class AdaBoost”, 2009.

[24] Freund, Y., Schapire, R., “A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting”, 1995.

[25] Kingma, D. P., BA, J. L., “Adam: A Method for Stochastic Optimization”, International Conference on Learning Representations, 2014

[26] Yosinski, J., Clune, J., Bengio, Y., Lipson, H., “How transferable are features in deep neural networks?”, Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14)

[27] Blint, R., J., “The Relationship of the Laminar Flame Width to Flame Speed”, Combustion Science and Technology, 49:1-2, 79-92, DOI:10.1080/00102208608923903

Contact Information

For further information/details please contact:

DEFINITIONS / ABBREVIATIONS

CFD – Computational Fluid Dynamics

RANS – Reynolds Averaged Navier-Stokes

TIT – Temperature at Inlet of Turbine

LFS – Laminar Flame Speed

LFT – Laminar Flame Thickness

EGR – Exhaust Gas Recirculation

MAE – Mean Absolute Error

RMSE – Root Mean Square Error

P – Pressure

T_u – Temperature

ϕ - Equivalence ratio

X_{EGR} - Mass fraction of the diluent

Sl^0 – Laminar flame speed at reference conditions

P_0 – Reference pressure

T_0 – Reference temperature

Sl_0 – Laminar flame speed without water addition

K_{egr} – Proportionality coefficient of EGR fraction

K_{wat} – Proportionality coefficient of water fraction

X_{H2O} – Water vapour mass fraction inside mixture

$\widetilde{s_L}$ – Laminar flame speed after correctio

$\delta_L^{\,0}$ – Laminar flame thickness

δ_L^{Blint} – Laminar flame thickness using Blint's relation

λ – Thermal conductivity

C_P – Specific heat at constant pressure

ρ – Density