

## Article

# Comprehensive Method for Obtaining Multi-Fidelity Surrogate Models for Design Space Approximation: Application to Multi-Dimensional Simulations of Condensation Due to Mixing Streams

José Galindo <sup>1</sup>, Roberto Navarro <sup>1,\*</sup>, Francisco Moya <sup>1</sup> and Andrea Conchado <sup>2</sup>

<sup>1</sup> CMT-Motores Térmicos, Universitat Politècnica de València, 46022 Valencia, Spain; galindo@mot.upv.es (J.G.); framoto1@upv.es (F.M.)

<sup>2</sup> Center for Quality and Change Management, Universitat Politècnica de València, 46022 Valencia, Spain; anconpei@eio.upv.es

\* Correspondence: ronagar1@mot.upv.es

**Abstract:** In engineering problems, design space approximation using accurate computational models may require conducting a simulation for each explored working point, which is often not feasible in computational terms. For problems with numerous parameters and computationally demanding simulations, the possibility of resorting to multi-fidelity surrogates arises as a means to alleviate the effort by employing a reduced number of high-fidelity and expensive simulations and predicting a much cheaper low-fidelity model. A multi-fidelity approach for design space approximation is therefore proposed, requiring two different designs of experiments to assess the best combination of surrogate models and an intermediate meta-modeled variable. The strategy is applied to the prediction of condensation that occurs when two humid air streams are mixed in a three-way junction, which occurs when using low-pressure exhaust gas recirculation to reduce piston engine emissions. In this particular case, most of the assessed combinations of surrogate and intermediate variables provide a good agreement between observed and predicted values, resulting in the lowest normalized mean absolute error (3.4%) by constructing a polynomial response surface using a multi-fidelity additive scaling variable that calculates the difference between the low-fidelity and high-fidelity predictions of the condensation mass flow rate.

**Keywords:** surrogate modeling; multi-fidelity simulations; design of experiments; design space exploration; condensation



**Citation:** Galindo, J.; Navarro, R.; Moya, F.; Conchado, A. Comprehensive Method for Obtaining Multi-Fidelity Surrogate Models for Design Space Approximation: Application to Multi-Dimensional Simulations of Condensation Due to Mixing Streams. *Appl. Sci.* **2023**, *13*, 6361. <https://doi.org/10.3390/app13116361>

Academic Editors: Luís L. Ferrás and Alexandre M. Afonso

Received: 11 April 2023

Revised: 10 May 2023

Accepted: 15 May 2023

Published: 23 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Typical engineering problems are based on meaningful simplifications of the behavior of complex systems. Ideally, one would like to be able to predict the performance of the system based on a combination of inputs that define its characteristics and particular operating conditions. To do so, a set of (training) data is generated using experimental measurements or numerical simulations, and a so-called surrogate model of the system is built. Such experiments and simulations are usually expensive [1], so the number of evaluations should be kept as low as possible. Surrogate models (also known as black-box models or metamodels), therefore, aim at predicting the outcome of the system with adequate accuracy at different points from those employed to construct the model. In the framework of Industry 4.0 and the paradigm shift to the Digital Twin concept [2], surrogate models are believed to play a vital role [3–5].

Surrogate models can be used for many different purposes [1], including sensitivity analyses [6–8], optimization [9–11] and design space approximation. The latter term is applied to cases in which the surrogate is required to represent the global behavior of a complex system over the whole set of potential operating conditions. For instance, surrogates

can be used to create turbocompressor [12–14] and turbine [15] characteristic maps, which provide the performance of the devices in terms of pressure ratio and isentropic efficiency. Heat transfer in heat exchangers [16] and in other applications [17], aero thermoelastic performance of hypersonic vehicles [18], piston engine volumetric efficiency [19] and microstructure reconstruction of porous media [20] are problems of design space exploration that have been addressed using surrogate modeling.

There is a wide variety of surrogate models able to satisfy different objectives. Yondo et al. [21] conducted a comprehensive review of different surrogate-based methods regarding aircraft aerodynamic studies. The most common types of surrogates found in the literature are the Radial Basis Function (RBF) model, the Polynomial Response Surface (PRS) model, and the Kriging model. Reihani et al. [22] used PRS to assess the sensitivity of low-pressure exhaust gas recirculation characteristics on turbocharger compressor performance. Simpson et al. [23] found that kriging surrogates were more accurate than PRS ones in the scope of aerospike nozzle design. In contrast, Fang et al. [24] concluded that RBF models presented a higher quality to predict non-linear effects than PRS models in several test problems. Nevertheless, the accuracy of surrogate models is case-dependent [25], so methods that seek to obtain adequate surrogate models for a particular application should compare between different alternatives [26].

For problems with numerous parameters and computationally or these casdemanding simulations, engineers often cannot succeed in conducting an affordable numerical campaign to develop an accurate surrogate of a problem design space due to the required dense sampling. For these cases, the possibility of resorting to the so-called multi-fidelity surrogates arises to alleviate the computational effort. Multifidelity (MF) surrogate models integrate information from high-fidelity (HF) and low-fidelity (LF) simulations, obtaining an adequate trade-off between the high computational requirements of HF models and the lack of precision of LF-models [1]. The design space approximation allows building the MF-surrogate model by mapping the HF model's parameter space through the LF's space, despite having different dimensions.

Surrogate models involve some potential drawbacks that should be addressed. First, their use should be limited to high-fidelity simulations requiring high computational performance. The use of low-fidelity models usually leads to a reduction in the accuracy of the analyses, with the sole purpose of reducing computational time. On the other hand, using surrogate models for multi-objective problems requires specific algorithms and suitable sampling techniques capable of solving real-world engineering design problems.

Research in developing surrogate models has experienced a remarkable increase in the last decade, especially for exploring the design space and supporting optimal designs. Recent research has applied heuristic optimization algorithms to solve the optimization problems associated with surrogate models [27,28]. In addition, another way of building surrogate models using metaheuristic algorithms has been proposed [29,30]. In line with previous limitations of surrogate models, new algorithms focused on Surrogate-Based Optimization (SBO) have been proposed [21]. Other novel research approaches are oriented towards evolutionary optimization, including multi-objective approaches.

In this way, the surrogate employs not only the outcome from the expensive high-fidelity simulations but also involves the prediction of a much cheaper low-fidelity model, which could even be an analytical approximation [31]. The existing literature reviews [21,32,33] show that multi-fidelity surrogates have been applied essentially to optimization [10,34,35] and uncertainty quantification [36,37], with a few works in which there is some degree of application to design space exploration [38,39]. Unfortunately, useful surrogate methods for optimization may not be the right choices for design space approximation [25]. Furthermore, the performance of such surrogates is often assessed using synthetic and benchmark problems, so it is yet to be determined how to adequately employ multi-fidelity surrogates for original, authentic scientific, and engineering problems.

The main contribution of this paper is the proposal of a systematic method using multi-fidelity surrogate models for space design approximation. The proposed methodology

is applied in a problem of condensation mass flow in internal combustion engines, as an example of application in a complex engineering environment.

Therefore, an original procedure is developed to assess different multi-fidelity surrogate models systematically using well-established error metrics. The developed method is presented using a step-by-step procedure employing thorough mathematical notation, so that the interested researchers can obtain a multi-fidelity surrogate model useful for their problem. Different designs of experiments are employed to obtain multi-purpose numerical campaigns that only consider the relevant parameters according to the objective of each campaign. The computational effort of the multi-fidelity strategy proposed in this work has been reduced. Indeed, the novel introduction of the metamodeled variable in the scope of space design exploration allows increasing the accuracy of the surrogate at the same cost. In this way, the method can cope with situations in which it is not affordable to explore many parameters of a space design using expensive simulations, such as those conducted using 3D CFD.

The strategy is applied to the prediction of condensation when two humid air streams are mixed on a three-way junction with a twofold objective: to prove the method's validity in a real problem (in contrast with the synthetic examples employed in most of the works dealing with surrogates) and to set an example of how to apply the method. In this way, the results obtained from the numerical campaigns described in this work are available in the web version of this article.

Determining the condensation mass flow rate generated by mixing two streams of moist air is relevant for air conditioning [40] and internal combustion engines [41]. Such condensation can be roughly estimated using low-fidelity models (0D models, as described in Section 4.2 or more accurately predicted using high-fidelity models (3D CFD simulations, as described in Section 4.3) following [41,42]. In this work, the former will be used as a low-fidelity model of condensation mass flow rate and the latter as its high-fidelity counterpart, showing in this way how to apply a multi-fidelity surrogate strategy to predict condensation in a three-way junction.

The paper is organized as follows. The proposed method for obtaining multi-fidelity surrogate models for design space approximation is presented in Section 2. The statistical methodology is covered in Section 3, describing the selected designs of experiments, the characteristics of the screening stage, the surrogate models employed, and the error metrics considered in the framework of the developed strategy for creating accurate multi-fidelity surrogate models. Section 4 briefly overviews the low and high fidelity models employed to predict condensation due to mixing streams. The application of the proposed multi-fidelity method to this particular case is presented in Section 5. Finally, some concluding remarks are provided in Section 6.

## 2. Proposed Multi-Fidelity Method

Table 1 enumerates all the steps included in the comprehensive method designed in this work, which provides multi-fidelity surrogate models for design space approximation. The table distinguishes between the actions required in the HF and MF frameworks. The mathematical notation featured in Table 1 avoids particularities of the problem used as an example (condensation in mixing streams) and, thus, allows researchers from other fields to use the proposed method.

The first stage of Table 1 is the *Conceptual modeling*, in which one defines the complete dataset  $X_N = \{x_i^l\}$  by identifying the model parameters  $i$  and determining the particular values for each parameter at level  $l$ . Section 5.1.1 shows an example of such steps. The conceptual modeling ends by selecting a set of intermediate variables ( $z_m$ ) that will be assessed for their use as surrogates, together with the target variable, which is the one that is required to be mapped ( $z_t$ ). For the studied application, these variables are defined in Section 5.1.2.

The next section of Table 1 is the *Identification of significant parameters*, which aims to reduce the computational effort of the surrogate construction. In this way, a subset of  $X_{scr}$  points can be determined using a certain design of experiments (a fractional factorial

design is suggested in this work; see Section 3.1.1) and then run High-Fidelity models  $Y_{scr}^{HF}$  and Low-Fidelity models  $Y_{scr}^{LF}$  at these working points to obtain a multi-fidelity evaluation. After the screening analysis over all metamodeled variables  $z$  (using, for instance, ANOVA, as described in Section 3.2), the non-significant parameters can be removed, leaving only  $N'$  significant parameters for the subsequent steps.

The last stage of the proposed method (see Table 1) is the *Surrogate modeling (for space design exploration)* devoted to the surrogate modeling itself. Firstly,  $s$  surrogate candidates are selected (see Section 3.3), and a set of training points  $X_{tra}$  is determined using another design of experiments. Here, the Box-Behnken design is employed, as justified in Section 3.1.2. Next, High-Fidelity ( $g_{s,t}, g_{s,m}^{HF}$ ) and Multi-Fidelity ( $g_{s,m}^{MF}$ ) surrogate models are constructed using the evaluation of High-Fidelity  $Y_{tra}^{HF}$  and Low-Fidelity models  $Y_{tra}^{LF}$  at the training points. These surrogate models are then used to predict values of the target variable at the verification points  $\hat{Y}_{s,t,ver}, \hat{Y}_{s,m,ver}^{HF}$  and  $\hat{Y}_{s,m,ver}^{MF}$ . Notice that these works consider the screening subset  $X_{scr}$  to be used as verification subset  $X_{ver}$  to reduce the computational cost. Since the exact results are already available from stage 2.2 of Table 1, the error metrics (as defined in Section 3.4) are evaluated, allowing the different combinations of HF or MF surrogate models with the target or intermediate variables to be classified as adequate or not in terms of accuracy. Finally, the best couple of surrogate  $s$  and variable  $z$  is selected to be used for space design exploration at any working point within the assessed range  $X_{exp}$ .

**Table 1.** List of method steps and actions to be conducted depending on the fidelity level.

Stage and Included Steps	High-Fidelity Model <sup>1</sup>	Multi-Fidelity Framework
<b>1. Conceptual modeling</b>		
1.1 Definition of $N$ model parameters and $l$ number of levels	$X_N = \{x_i^l\}, i = 1, \dots, N; \ell = 1, \dots, L$	-
1.2 Definition of target and intermediate variables	$z_t = y(x)z_m(y, x) : y = f_m(z_m, x)$	$z_m^{MF}(y^{HF}, y^{LF}) : y^{HF} = f_m^{MF}(z_m^{MF}, y^{LF})$
<b>2. Identification of significant parameters</b>		
2.1 Determination of screening subset	$X_{scr} \subseteq X_N$	-
2.2 Running the models at screening points	$Y_{scr} = y(X_{scr}) Z_{m,scr} = z_m(Y_{scr}, X_{scr})$	$Y_{scr}^{LF} = y^{LF}(X_{scr}) Z_{m,scr}^{MF} = z_m^{MF}(Y_{scr}^{HF}, Y_{scr}^{LF})$
2.3 Identification of $N'$ significant parameters ( $x_j$ )	$\forall j, \exists z : F_N(x_j, z) > F_{(L-1),residual}(x_j, z)$	-
2.4 Removal of $N - N'$ non-significant parameters	$X_{N'} = \{x_j^l\}, j = 1, \dots, N'; \ell = 1, \dots, L N' \leq N$	-
<b>3. Surrogate modeling (for space design exploration)</b>		
3.1 Selection of surrogate candidates	$g_s : \hat{z}_s = g_s(x)$	-
3.2 Determination of training subset	$X_{tra} \subseteq X_{N'}$	-
3.3 Running the models at training points	$Y_{tra} = y(X_{tra}) Z_{m,tra} = z_m(Y_{tra}, X_{tra})$	$Y_{tra}^{LF} = y^{LF}(X_{tra}) Z_{m,tra}^{MF} = z_m^{MF}(Y_{tra}^{HF}, Y_{tra}^{LF})$
3.4 Construction of surrogate models	$g_{s,t} : \min \epsilon_{s,t}(\hat{Y}_{tra}, Y_{tra}) \quad g_{s,m} : \min \epsilon_{s,m}(\hat{Z}_{m,tra}, Z_{m,tra})$	$g_{s,m}^{MF} : \min \epsilon_{s,m}^{MF}(\hat{Z}_{m,tra}^{MF}, Z_{m,tra}^{MF})$
3.5 Determination of verification subset	$X_{ver} \equiv X_{scr}$	-
3.6 Prediction of target variable at verification subset	$\hat{Y}_{s,t,ver} = g_{s,t}(X_{ver}) \quad \hat{Y}_{s,m,ver} = f_m(g_{s,m}(X_{ver}), X_{ver})$	$\hat{Y}_{s,m,ver}^{MF} = f_m^{MF}(g_{s,m}^{MF}(X_{ver}), y^{LF}(X_{ver}))$
3.7 Calculation of error metrics and selection of best surrogate variable and model	$z, s : \min \epsilon(\hat{Y}_{s,z,ver}, Y_{ver})$	-
3.8 Usage of surrogate for space design exploration	$Y_{exp} \simeq g_{s,t}(X_{exp}) \quad Y_{exp} \simeq f_m^{HF}(g_{s,m}^{HF}(X_{exp}), X_{ver})$	$Y_{exp} \simeq f_m^{MF}(g_{s,m}^{MF}(X_{exp}), y^{LF}(X_{exp}))$

<sup>1</sup> For the sake of simplicity, HF superscript is omitted from the notation in this column, as it affects all variables and functions.

### 3. Statistical Methodology

#### 3.1. Design of Experiments

The proposed multi-fidelity method requires several sets of working points for screening, training, and verification purposes. Different designs of experiments are considered, depending on the required characteristics of each set of points. In this way, a fractional factorial design obtains the screening points, and a Box-Behnken design is used to provide the training points. For the sake of computational effort, the same set of points employed for screening will also be used to verify the accuracy of the surrogate models.

### 3.1.1. Fractional Factorial Design

The proposed method requires a subset of points  $X_{scr}$  to conduct the screening stage. The full factorial design would be a candidate for this purpose, as it allows exploration a wide variety of combinations between the parameters at different levels, entailing a high level of resolution in the estimation effects. However, it may produce a verification subset with a prohibitive computational cost, depending on the number of parameters in the model and the cost of each simulation. Furthermore, presenting great accuracy in estimating high-level interaction terms at the screening stage is unnecessary.

Thus, to save a little computational effort, it was decided to perform a fractional factorial design instead. The objective of the factorial fractions is to identify the characteristic parameters with the greatest effects on the high and multi-fidelity variables. One-half fraction designs reduce the number of simulations by half. Still, researchers can further reduce the number of simulations by selecting a quarter, an eighth, or even a sixteenth fraction. Increasing the fraction's denominator reduces the number of simulations required and the resolution needed to estimate higher-order interaction effects. Factor fractions of order  $p$  for factors with two levels and  $N$  characteristic parameters, with the possible inclusion of an intermediate level for center points, are represented by  $2^{N-p}$ .

The resolution of the fraction design influences the number of high-order interaction effects that are confounded to estimate main effects and low-order interactions based on a reduced number of simulations. Since the characteristics of this fraction design meet the requirements of the screening stage as well as the verification stage, the verification subset will coincide with the screening subset ( $X_{ver} \equiv X_{scr}$ ) so that the effort of this extensive numerical campaign is exploited twice.

### 3.1.2. Box-Behnken Design

When fitting a polynomial response surface (PRS) model, quadratic surfaces are estimated to represent the curved shape of continuous factors. Box-Benken designs [43] are incomplete second-order designs based on three levels, making them appropriate for obtaining training subsets  $X_{tra}$  to construct surrogates. The number of simulations this design requires is  $2N(N - 1)$ . Center points can be added to this set of simulations to improve the statistical power of the analysis. This design is more efficient than other response surface designs, such as composite centers or three-level full factorial designs [44]. In addition, this design does not include combinations of levels where all characteristic parameters are simultaneously at high or low levels. Therefore, it only forces the execution of simulations under extreme conditions that may provide meaningful results in design space exploration.

## 3.2. Screening Stage

The proposed method (see Table 1) conducts a screening phase to identify the characteristic parameters that generate significant differences in either high-fidelity or multi-fidelity variables. In this analysis stage, the focus is on direct and linear effects rather than possible quadratic or nonlinear effects. Analysis of variance (ANOVA) is a versatile method for this screening phase as it allows the exclusion of nonlinear terms from the model, thus increasing its power. Thus, non-significant terms that usually correspond to interaction effects can be specifically eliminated to improve the significance of the model by including only the main results. ANOVA proves to be a valuable technique to screen input parameters and thus reduce the extension of subsequent numerical [45] or experimental campaigns [46] for optimization.

This analysis of variance (ANOVA) is based on multiple comparisons of the means of the high and multi-fidelity variables at different levels of the characteristic input parameters. In the ANOVA model, the F-statistic and its corresponding  $p$ -value are critical in identifying the characteristic parameters that cause significant differences in the high-fidelity and multi-fidelity variables. This F-statistic is calculated for the main effect of each characteristic parameter as well as for the residual variability of the model. The F-statistic follows a Fisher-Snedecor distribution, assuming the null hypothesis ( $H_0$ ) that different levels of the characteristic parameters do not produce significant differences in the high

and multi-fidelity variables. The F statistic represents a ratio between two measures of variability weighted by degrees of freedom (mean squares): the variability explained by each characteristic parameter relative to the residual variability of the model. Thus, the F statistic represents the variability explained by each characteristic parameter relative to the variability not explained (residual) by the model.

The variability explained by different levels of characteristic parameters can be calculated as the Explained Sum of Squares (ESS):

$$ESS = \sum n_i \frac{(\bar{Y}_i - \bar{Y})^2}{l - 1} \quad (1)$$

where  $\bar{Y}_i$  represents the sample mean of the high or multi-fidelity variable at each level of the characteristic parameter,  $\bar{Y}$  represents the overall mean of this variable, and  $l$  denotes the number of levels of the parameter.

The residual (intra-level) variability (Residual Sum of Squares) is calculated as follows:

$$RSS = \sum_{ij} \frac{(Y_{ij} - \bar{Y}_i)^2}{N - l} \quad (2)$$

where  $Y_{ij}$  represents the  $j$ -th simulation at each of the  $i$ -th level of the high or multi-fidelity variable, and  $\bar{Y}_i$  represents the mean value of that variable at each level of the characteristic parameter,  $N$  represents the total sample size, and  $l$  represents the number of levels.

Using both terms, the F statistic of each parameter characteristic can be computed as the ratio between the variability explained by different levels and the (unexplained) residual variability, within levels of characteristic parameters.

$$F = \frac{ESS}{RSS} \quad (3)$$

Along with this F-statistic,  $p$ -values can be obtained for identifying significant effects ( $p$ -value  $< \alpha$ ). These  $p$ -values are based on the comparison between calculated F-statistics and critical values of the Fisher - Snedecor distribution. The degrees of freedom for the numerator and denominator of this distribution are calculated by combining the number of levels and the screening sample size. A 95% confidence level is usually considered for assessing the significance of  $p$ -values, although more restrictive approaches (99% confidence level). In this screening stage, some low-level interactions (double or even triple) may show significant effects. The retention of these interaction terms is at the discretion of the researcher, although in most cases it is not interesting for research.

### 3.3. Surrogate Models

A key stage in the developed method (see Table 1) is the construction of surrogate models. In this work, three widely employed surrogate models have been considered: Polynomial Response Surface (PRS), Radial Basis Function (RBF), and Ordinary Kriging. PRS models are based on classical regression models and analysis of variance to analyze the relationship between a response variable and a set of predictor variables. RBF models are based on a collection of functions modeling a measure of distance (usually Euclidean) between the point to be measured and a fixed point (usually the origin or an established center), as described in Section 3.3.2. Furthermore, ordinary kriging models are built from a kernel function considering the deviation between the metamodel variable of interest and its prediction, as shown in Section 3.3.3. Thus, RBF and Ordinary kriging models differ from the PRS model in the use of collection functions to approximate the variable of interest, whereas PRS models apply classical modelization techniques to estimate the parameters of the model. All of them will be used to obtain metamodels of every variable  $z$  assessed, based on the training subset  $X_{tra}$ . Notice that the MF surrogates developed in this work are based on scaling functions  $z_m^{MF}$  [1].

### 3.3.1. Polynomial Response Surface Models (PRS)

Polynomial response surface models (PRS) have been extensively used for data fitting and regression analysis to examine the relationship between variables and outputs [47–49]. A second-order polynomial surrogate can be written as:

$$g_s = \beta_0 + \sum_{i=1}^{N'} \beta_i x_i + \sum_{i=1}^{N'} \sum_{j=1}^{N'} \beta_{ij} x_i x_j \quad (4)$$

The calibration of the coefficients  $\beta$  of Equation (4) for the polynomial response surface has been set using the R software library [50].

### 3.3.2. Radial Basis Function Models (RBF)

A so-called radial function employs the Euclidean distance between the input parameter and a fixed point, known as the center. A Radial Basis Function (RBF) model is constructed using a linear superposition of different radial basis functions, which is presented in Equation (5):

$$g_s = \sum_{t=1}^{N_{tra}} w_t \psi(\|x - x_t\|) \quad (5)$$

To use the combination of  $N_{tra}$  radial functions as a RBF surrogate, each point  $x_t$  corresponding to the training subset ( $X_{tra} = \{x_t\}$ ) is employed as the center,  $w_t$  are weights associated with each center  $x_t$  and  $\psi$  is a basis function that acts over the aforementioned euclidean distance. In this work, the library PySMO developed by the Institute for the Design of Advanced Energy Systems based on the work of Forrester and Rippa et al. [10,51] is employed on Python to construct RBF surrogates. Depending on the variable to be metamodeled, the best function  $\psi$  amongst those available in the library (linear, cubic, thin-plane spline, gaussian, and multi-quadratic) is selected to obtain the greatest fitting.

### 3.3.3. Ordinary Kriging Model

A surrogate model combining a known function  $f(x)$  (regression) with a function  $Z(x)$  as Equation (6) shows:

$$g_s = \sum_{t=1}^{N_{tra}} \beta_t f(x) + Z(x) \quad (6)$$

is referred to as a Kriging model due to its first use in the field of geostatistics [52]. The so-called kernel  $Z(x)$  considers the deviation between the metamodeled variable and the prediction of the deterministic function  $f(x)$  (here considered a constant) using a stochastic Gaussian process, so Kriging models are also known as Gaussian process regressions. In Equation (6),  $\beta_t$  are the regression coefficients,  $N_{tra}$  the number of training points, and  $Z(x)$  presents zero mean and a co-variance of  $\sigma^2 \psi$ , where the correlation function  $\psi$  employed in this work is the squared exponential (see [53,54]).

For this work, Kriging surrogates have been built using a Toolbox on Python environment developed by Bouhlef et al. [53]. According to Xiao et al. [55], Kriging models better describe the local features of the data to be metamodeled than PRS. Comparing Equation (6) with Equation (5), the functions  $\psi$  play a similar role as either radial basis functions in the RBF model or kernels in the Kriging model [56]. In this way, Chandrashekarappa and Duviigneau [57] noted that the normal distribution considered in the Kriging kernel functions is more adequate for experimental measurements than for computational simulations, which is the application example considered in this work (Sections 4 and 5).

### 3.4. Error Metrics

Several steps of the proposed strategy to develop multi-fidelity surrogates (see Section 2) require a quantitative assessment of the obtained models, to evaluate and compare their

performance to select the best combination of metamodeled variables and surrogate types. Generally, these metrics check the error committed between the observed values and the predicted values, but the roles of the observed and predicted data at the corresponding stage will determine which is the error indicator best suited for the task.

The first indicator employed in this work is the coefficient of determination  $R^2$ . It quantifies the proportion of the total variability of the data that the model can explain. To do so, this error metric employs the co-variance of the predicted and observed cases as well as the variance of each case, as can be seen in Equation (7):

$$R^2 = \frac{\sigma_{z\hat{z}}^2}{\sigma_z^2 \sigma_{\hat{z}}^2} = \frac{\left(\frac{\sum_{i=1}^n (z_i - \bar{z})(\hat{z}_i - \bar{\hat{z}})}{n}\right)^2}{\frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n} \frac{\sum_{i=1}^n (\hat{z}_i - \bar{\hat{z}})^2}{n}} \tag{7}$$

Notice that  $R^2$  in this work is only employed to assess the fitting capability of each model outcome ( $y_i$ ) to the training data ( $x_i$ ). Indeed, the coefficient is known to be inadequate when attempting to evaluate the predicting capability of the models in the verification stage [58]. For the latter purpose, the Mean Absolute Error (MAE) is preferred [58,59]. MAE directly calculates the average magnitude of the error between the values predicted by a surrogate ( $\hat{y}_i$ ) and the values observed at the verification stage ( $y_i$ ), as Equation (8) shows.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \tag{8}$$

Following the work performed by Roy et al. [60], MAE can be applied as a means of qualitative assessment to determine whether a surrogate model can be considered to provide a good prediction or not. To do so, the 5% of data with the highest residual values is first removed, and then MAE is normalized considering the training dataset range  $R_{tra}$  and the dispersion of the errors. This elimination of a small part (5%) of the values with the highest residuals in the test set is performed to attenuate the influence of possible outliers that increase prediction errors. This procedure aims to evaluate the quality of the prediction using 95% of the data present in the test set, thus increasing the robustness of the estimate. Equations (9) and (10) show the conditions required to grant a good evaluation to the surrogate. When the two conditions are fulfilled at the same time, a good agreement is obtained. If only one of these conditions is satisfied, a moderate agreement is set. If no condition is met, a bad agreement is therefore considered.

$$\widetilde{MAE} = \frac{MAE(95\%)}{R_{tra}} \leq 0.1 \tag{9}$$

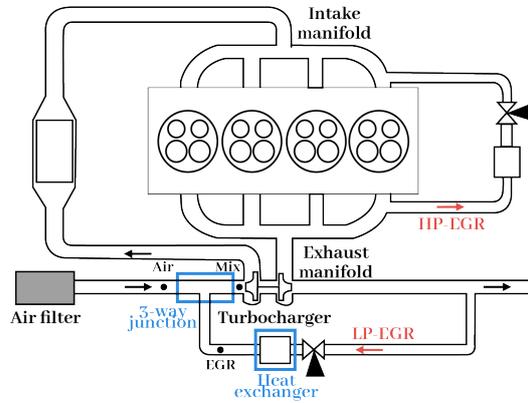
$$\widetilde{MAE}_\sigma = \frac{MAE(95\%) + 3 \cdot \sigma}{R_{tra}} \leq 0.2 \tag{10}$$

#### 4. Description of the Application Problem and Employed Multi-Fidelity Models

##### 4.1. Background of Volume Condensation Due to Mixing

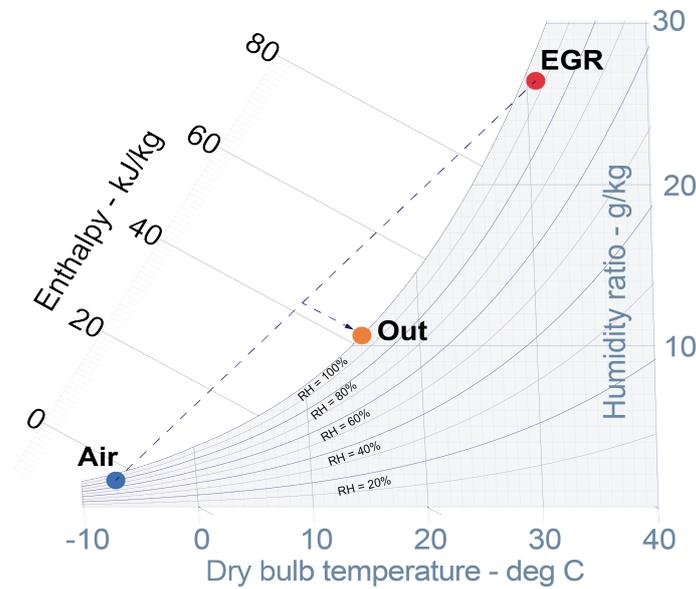
The application of this method is devoted to characterizing a problem that appears in the automotive sector when a vehicle features an internal combustion engine such as the one depicted in Figure 1. For such turbocharged engines, the use of the so-called Low-Pressure Exhaust Gas Recirculation (LP-EGR) is being widely researched to reduce  $CO_2$ ,  $NO_x$ , and particulate matter emissions [61,62]. EGR consists of mixing a fraction of exhaust gases (with high humidity content) with fresh air coming from the ambient. As Figure 1 shows, LP-EGR performs this mixture upstream of the compressor, so condensation can happen at certain cold conditions, either at the LP-EGR cooler [63,64] or at the three-way junction [65,66]. This possibility of generating condensation upstream of the compressor constitutes one of the main drawbacks of LP-EGR. Indeed, water droplets reaching the impeller will erode its blades and thus reduce its lifetime and performance [41,67]. In this way, the ability to predict the condensation mass flow rate for the whole engine map would

allow automotive engineers to maximize the LP-EGR rate for each working point (and thus minimize CO<sub>2</sub>, NO<sub>x</sub> and particulate matter emissions) without exceeding the damage threshold of the impeller, to keep its durability.



**Figure 1.** Engine layout, highlighting EGR systems and main condensation locations.

Focusing on the mixing process at three-way junctions, the explanation of how two sub saturated streams can produce condensates is given by the psychrometric diagram displayed in Figure 2. Following the LP-EGR junction highlighted in Figure 1, a humid, warm stream of EGR can be blended with a flow of cold ambient air. In these cases, Figure 2 confirms that their mixture can be located in the supersaturation region so that the excess of water until the point “mix” lies within the saturation line should be condensed into the form of liquid water.



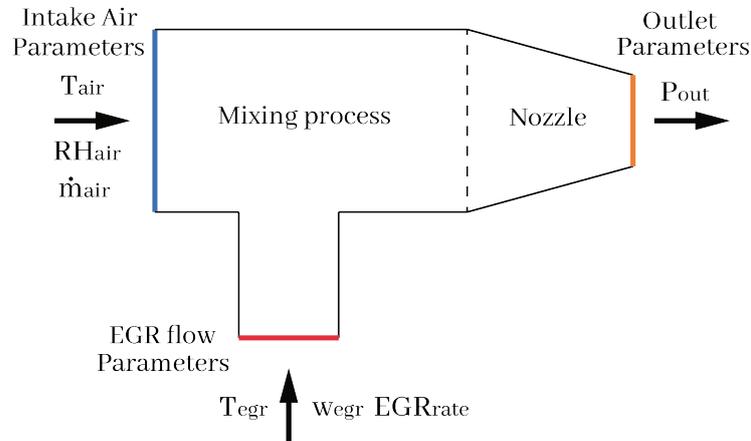
**Figure 2.** Psychrometric diagram showing the process of humid streams mixed in a three-way junction as considered by the 0D model.

#### 4.2. Low Fidelity: 0D (Perfect Mixing) Condensation Model

In this work, a 0D mixing condensation model is considered a low-fidelity model. Figure 3 shows the parameters of this model. It can be regarded as an evolution of the one developed by Serrano et al. [41], in which a total (perfect) mixing between the inlet streams (air and EGR) is considered in an isobaric and isenthalpic way, as depicted in Figure 2. After calculating the mixing condensation, the 0D model employed in this work also evaluates the temperature decrease due to the area reduction (labeled “nozzle” in Figure 3) usually existing in LP-EGR junctions as a potential source of additional condensation. More

details about the psychrometric calculations to obtain the condensation considering perfect mixing can be found in the work of Serrano et al. [41].

In the perfect mixing process, both streams are completely mixed in a homogeneous flow, obtaining the maximum amount of water condensed in perfect conditions.

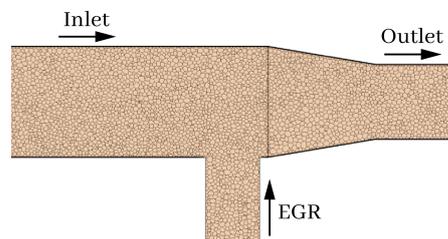


**Figure 3.** Scheme of the input parameters on the perfect mixing model.

#### 4.3. High Fidelity: 3D CFD Condensation Model

This work employs a 3D CFD model of the three-way junction embedded with a condensation subroutine as a high-fidelity model. The approach used is the one described by Galindo et al. [68] and has been validated in several ways using a series of different experimental techniques: secondary flows measured with laser particle image velocimetry [68], qualitative condensation distributions observed using planar laser-induced visualization [68], temperature fields [69] and condensation mass fluxes [69]. Indeed, this numerical tool has been able to predict which junction configuration provides the least amount of condensation mass flow rate, which in turn provides the lowest impeller wear in agreement with the level observed in experimental durability tests [70].

The geometry of the study is an LP-EGR three-way junction where fresh inlet air and EGR are mixed before the compressor. As shown in Figure 4, there are two inlet branches (a 50 mm diameter duct for air and a 22.4 mm diameter pipe for EGR) and one outlet duct, whose diameter is 34 mm. The nozzle section after the EGR insertion represents the tapered inlet of the compressor. To reduce the impact of boundary locations on the region of interest, both inlet ducts are extruded three additional diameters upstream, and the outlet pipe is extended two diameters downstream.



**Figure 4.** LP-EGR three-way junction with a cross-section of the mesh employed.

The mesh depicted by Figure 4 comprises polyhedral cells with a base size of 1.5 mm and eight prism layers next to the walls to improve the boundary layer resolution. A mesh independence study has selected this grid of 3M cells. Employing a coarser grid of 300 thousand cells results in a difference of 10.4% in condensation at the outlet section. Refining the mesh instead until it features 11M cells provides a difference in condensation of only 3%. Considering the computational effort of the numerical campaign (more than 180 simulations), the mesh of 3M cells was therefore selected, which achieves  $y^+$  to be below 1.5 for 99% of the wall cells.

The segregated solver of Simcenter STAR-CCM+ is selected for the simulations. Regarding the turbulence formulation, Reynolds-Average Navier-Stokes (RANS) was selected with the *SST k- $\omega$*  turbulence model [71]. The cases are initially run steadily, and their condensation mass flow rate is monitored. For those working points with an oscillation in this variable above 2%, the temporal treatment was switched to unsteady RANS (URANS), employing a second-order implicit transient solver with a time-step size of  $\Delta t = 10^{-4}$  s. As boundary condition types, the mass flow rate, specific humidity, stagnation temperature, and turbulence intensity are specified at both inlets, whereas the static pressure is set at the outlet. The particular values for the numerical campaign are indicated in Section 5.1.1, which correspond to representative conditions for a C-segment passenger car engine.

A condensation subroutine [42] is embedded into the 3D CFD model. This condensation submodel reads the psychrometric state of each cell and compares it to the corresponding saturated equilibrium state. When supersaturation is detected, the model instantaneously condenses the required amount of water, interacting with the transport equations solved by the 3D CFD code using the corresponding source terms. For more details, please refer to Serrano et al. [42].

#### 4.4. Comparison between Low-Fidelity (LF) and High-Fidelity (HF) Condensation Prediction

To assess the difference in performance between the 0D model (LF) and the 3D CFD model (HF), two cases are selected from the numerical campaign that will be described in Section 5.2.1: one delivering low condensation (Figure 5) and another providing more significant condensation (Figure 6). The results predicted by these models are included in Table 2.

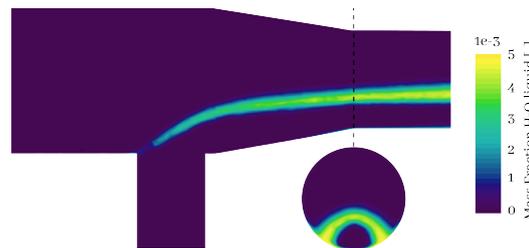


Figure 5. Low condensation case

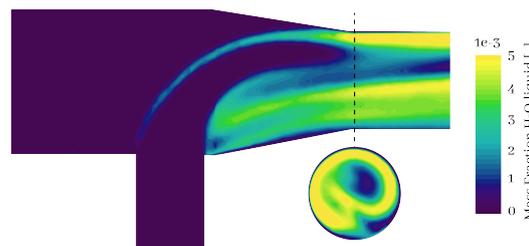


Figure 6. High condensation case

Table 2. Condensation mass flow rates predicted by the LF (0D) and HF (3D) model for two working points.

Working Point	$\dot{m}_{0D}$ [kg/h]	$\dot{m}_{3D}$ [kg/h]
A (Low condensation)	1.182	0.2406
B (High condensation)	2.673	1.651

The working points of Table 2 present the same air mass flow rate, but point B features a much greater EGR (humid stream) mass flow rate than point A, which is the main reason for the increase in the 0D condensation mass flow rate by 130%. The high-fidelity model also predicts an increase in condensation mass flow rate when shifting from case A to case B, but a much greater one (590%; see Table 2). Whereas the LF 0D model only considers the global psychrometric information, the HF 3D CFD model also considers the local flow features.

The different predictions of the LF and HF models can therefore be explained based on the flow patterns depicted in Figures 5 and 6. Figure 5 presents a low EGR mass flow rate, which results in a wall-jet flow configuration (see Kimura et al. [72]), while Figure 6, with a greater EGR rate, displays a flow pattern in the threshold between a deflecting-jet and an impinging-jet [72]. The 0D model provides an upper bound for the condensation mass flow rate, as it considers a perfect mixture between the streams. The different deviation between the 3D and the 0D models depends on the flow mixing, as the 3D CFD model predicted, as condensation is strongly correlated with mixing [66]. The mixing intensity (and thus condensation) is much greater for the high condensation case (Figure 6) due to the high penetration of the EGR stream than for the low condensation case (Figure 5). In the framework of multi-fidelity modeling [73], the 0D case could be useful to provide the psychrometric information of the operating conditions, being a LF model with little computational effort. However, a comprehensive set of expensive 3D CFD simulations will still be required. The examples analyzed in this section show how the deviation between the LF and HF models can differ when shifting the working point.

### 5. Application of the Proposed Method to the Problem of Volume Condensation Due to Mixing: Results and Discussion

#### 5.1. Conceptual Modeling

##### 5.1.1. Operating Conditions: Definition of Model Parameters and Their Range

The developed method starts the conceptual modeling stage by identifying the model parameters, as stated in Table 1. In this case, one only needs to identify the boundary conditions employed in the 3D CFD model ( $x_{HF}$ ), as described in Section 4.3, and the inputs of the 0D model ( $x_{LF}$ ), as depicted by Figure 3. By doing so, Table 3 is obtained, providing  $N = 9$  parameters. Notice that  $N_{LF} = 7$ , as the parameters for the LF model are the same as the HF one save for the turbulent intensities at both inlet and EGR ducts. Furthermore, three levels are selected for each parameter, with values covering a typical range of working points in a piston engine featuring LP-EGR that may produce condensation. The rightmost column of Table 3 includes the matrix of eligible parameter values  $x_i^\ell$  that will be employed for conducting the different designs of experiments. Notice that, in this case, the geometry of the junction has been kept constant so that the exploration is performed in terms of operating conditions.

**Table 3.** HF model parameters and levels considered for the different designs of experiments.

Parameters	Variables	Units	Low, Medium and High Values
Inlet air mass flow	$\dot{m}_{air}$	[kg/h]	50, 175, 300
Inlet temperature	$T_{air}$	[°C]	−10, 0, 10
Inlet relative humidity	$RH_{air}$	[%]	50, 75, 100
Inlet turbulent intensity	$I_{air}$	[%]	1, 5.5, 10
EGR temperature	$T_{egr}$	[°C]	40, 50, 60
EGR rate	$EGR_{rate}$	[%]	10, 25, 40
EGR specific humidity	$w_{EGR}$	[g/kg]	30, 50, 70
EGR turbulent intensity	$I_{EGR}$	[%]	1, 5.5, 10
Outlet pressure	$P_{out}$	[bar]	0.8, 0.9, 1

##### 5.1.2. Definition of Target and Intermediate Variables

The method developed in this work is applied to obtain a surrogate for the condensation mass flow rate that 3D CFD simulations (HF model) would predict. Therefore, this is the target variable  $z_t$  sought:

$$z_t = \dot{m}_{cond,3D} \tag{11}$$

Furthermore, the method suggests (see Table 1) that additional intermediate variables could be modeled through surrogates. The underlying idea is that the distribution of such variables across the design space could be more adequate to build surrogate models than that of the target variable. Table 4 presents the target and intermediate variables metamodeled in this work.

**Table 4.** Target and intermediate variables studied.

Name	Variable	Type
3D condensation	$\dot{m}_{cond.,3D}$	HF Target var.
Condensation humidity	$w_{cond.,3D}$	HF Intermediate var.
Condensation ratio	$\varphi$	MF Multiplicative scaling var.
Condensation difference	$\Delta$	MF Additive scaling var.
Junction efficiency	$\eta$	MF Hybrid scaling var.

An intermediate variable can be defined within the domain of the HF model ( $z_m$ ), as long as it involves the target variable and other input parameters. This is the case of the condensation humidity, which normalizes the target condensation mass flow rate by the inlet mass flow rate of ambient dry air:

$$(z_{m=w_{cond.,3D}}) : w_{cond.,3D} = \frac{\dot{m}_{cond.,3D} \cdot (1 + w_{air})}{\dot{m}_{air}} \cdot 1000 \tag{12}$$

$$(f_{m=w_{cond.,3D}}) : \dot{m}_{cond.,3D} = \frac{\dot{m}_{air}}{1 + w_{air}} \cdot \frac{w_{cond.,3D}}{1000} \tag{13}$$

Notice that the equation for the function  $f_m$  is also included before, as it will be required to obtain later  $\dot{m}_{cond.,3D}$  on the basis of the surrogate model for  $w_{cond.,3D}$  and the parameters  $w_{air}$  and  $\dot{m}_{air}$  corresponding to the working point studied.

But aside from the HF intermediate variables, this work aims to show how to build and assess multi-fidelity surrogate models for space design exploration. In this way, Jiang et al. [1] identified three types of variables for MF models based on scaling functions: multiplicative, additive, and hybrid variables. For instance, a multiplicative variable can be defined as just the ratio between the HF and LF predictions for condensation mass flow rate:

$$(z_{m=\varphi}^{MF}) : \varphi = \frac{\dot{m}_{cond.,3D}}{\dot{m}_{cond.,0D}} \tag{14}$$

$$(f_{m=\varphi}^{MF}) : \dot{m}_{cond.,3D} = \dot{m}_{cond.,0D} \cdot \varphi \tag{15}$$

An MF additive variable is also built by calculating the difference between the condensation mass flow rates obtained by the 0D (LF) and 3D CFD (HF) models:

$$(z_{m=\Delta}^{MF}) : \Delta = \dot{m}_{cond.,0D} - \dot{m}_{cond.,3D} \tag{16}$$

$$(f_{m=\Delta}^{MF}) : \dot{m}_{cond.,3D} = \dot{m}_{cond.,0D} - \Delta \tag{17}$$

Finally, a hybrid variable may combine multiplications and additions of the LF and HF variables in its definition. In this work, a so-called junction efficiency  $\eta$  is defined:

$$(z_{m=\eta}^{MF}) : \eta = \frac{\dot{m}_{cond.,0D} - \dot{m}_{cond.,3D}}{\dot{m}_{cond.,0D}} \cdot 100 \tag{18}$$

$$(f_{m=\eta}^{MF}) : \dot{m}_{cond.,3D} = \dot{m}_{cond.,0D} \cdot (1 - \eta/100), \tag{19}$$

So that it is useful not only as an MF intermediate variable for design space exploration but also as a figure of merit to assess the junction design [74,75]. Indeed, since  $\dot{m}_{cond.,0D}$  is an upper bound for  $\dot{m}_{cond.,3D}$ , the junction efficiency defined in 18 ranges from 0% (maximum condensation) to 100% (avoids all condensation).

To conclude this section, the working points already discussed in Section 4.4 are again applied to obtain the values of the intermediate variables stated in Table 4, based on the LF and HF results featured in Table 2. In this way, Table 5 shows an example of the values that these different variables may take. For instance, the lower condensation case compared with the high condensation case (see Figures 5 and 6) results in a greater junction efficiency in terms of avoiding condensation (79.65% vs. 38.24%), following Table 5.

**Table 5.** Values of target and intermediate variables for two working points.

Working Point	$\dot{m}_{cond.,3D}$ [kg/h]	$w_{cond.,3D}$ [g/kg]	$\eta$ [%]	$\Delta$ [kg/h]	$\varphi$ [-]
A (Low condensation)	0.2406	0.8023	79.65	0.9417	0.2035
B (High condensation)	1.651	5.505	38.24	1.022	0.6176

### 5.2. Determination of Significant Parameters

#### 5.2.1. Selection of Screening Points and Results of the HF and LF

The objective of the screening stage is to identify the input parameters whose variation generates significant effects on the variable of interest. To achieve this goal, it is advisable to work with  $2^{N-p}$  factorial fractions, where  $k$  represents the number of factors ( $k = 9$  in this paper) and  $p$  the number of partitions ( $p = 2$  in this paper) to the total set of simulations using the full factorial design ( $2^N = 2^9 = 512$ ). The results obtained from a set of simulations designed by fractional factorial design always have a lower resolution than those obtained from a full factorial design. When the number of factors is high, the decrease in resolution is usually compensated by a remarkable reduction in the number of simulations. Using this method, 128 simulations obtained as a quarter of the full factorial design ( $2^{9-2}$ ) were generated. Simulations were executed for the 3D Model and the 0D Model. Target and intermediate variables were computed on the basis of these simulated values. See the Supplementary Materials.

High and low levels of characteristic parameters were complemented with several center points to improve the accuracy of the standard error of estimation. When condensation phenomena were not physically feasible due to the influence of local irregular flows or negligible amounts of condensed water, values of condensation mass flow were considered missing. Some other experiments were slightly modified to ensure condensation was produced for a more profitable data analysis. Based on this screening design, the variability in the five metamodeled variables was assessed for different combinations in the initial set of characteristic parameters.

#### 5.2.2. Anova and Removal of Non-Significant Parameters

The analysis of variance described in Section 3.2 compares the variability of each metamodeled variable  $z$  regarding the main effects produced by each model parameter. Table 6 presents these values in terms of the F-statistic defined in Equation (3).

Comparing the results shown in Table 6 with the corresponding critical value, three non-significant parameters (inlet turbulent intensity, EGR turbulent intensity, and outlet pressure) were identified and excluded from subsequent analysis. It is worth noting the model parameters that were deemed significant (usually with the highest confidence level of 99%) for every variable  $z$ , namely, inlet mass flow, EGR temperature, EGR humidity, and EGR rate. The remaining parameters (inlet temperature and relative humidity) are significant or not depending on the variable considered, so they will not be dismissed, as the best variable to construct a surrogate is not known yet.

**Table 6.** Identification of significant main effects according to ANOVA F-statistics for different HF model parameters and metamodeled variables.

Parameter	$F(\dot{m}_{cond.,3D})$	$F(w_{cond.,3D})$	$F(\varphi)$	$F(\Delta)$	$F(\eta)$
Inlet air mass flow	59.6 **	31.1 **	38.8 **	87.9 **	38.7 **
Inlet temperature	4.4 *	16.1 **	35.8 **	0.2	52.5 **
Inlet relative humidity	0.5	0.9	27.2 **	0.5	15.4 **
Inlet turbulent intensity	0.1	0.0	0.0	0.1	0.0
EGR temperature	11.8 **	27.1 **	24.7 **	18.7 **	25.2 **
EGR rate	53.4 **	103.1 **	143.1 **	22.0 **	128.9 **
EGR specific humidity	38.8 **	85.8 **	61.3 **	53.0 **	77.7 **
EGR turbulent intensity	0.2	0.4	0.0	0.3	0.1
Outlet pressure	0.03	0.3	1.00	0.6	1.8

Significant effects with 99.0% (\*\*) or 95.0% (\*) confidence level.

### 5.3. Surrogate Modeling for Space Design Exploration

#### 5.3.1. Selection of Surrogate Candidates

All surrogate models defined in Section 3.3 are assessed for this problem, so  $s = PRS, RBF, Krig$ .

#### 5.3.2. Selection of Training Points and Subsequent Obtention of HF and LF Results

To calibrate the surrogate models corresponding to each of the five metamodeled variables, a numerical campaign of 54 simulations was produced using a Box–Behnken design (see Section 3.1.2), considering the remaining six significant parameters. Two replicates of this design were added with center points to deal with some missing data (certain combinations of operating conditions do not produce condensation, as Figure 2 shows) and also to improve the accuracy of the standard error of estimation. These replicates were obtained by including random modifications to previously discarded (non-significant) factors. Thanks to these enhancements, the percentage of missing data was reduced to 13% due to the sample size increase. The collection of previous HF and LF simulations (54 working points) was subsequently used as the training data set to obtain the surrogate models.

#### 5.3.3. Construction of Surrogate Models and Prediction of Target Variable at Verification Points

This work has selected three surrogate models: Polynomial Response Surface (PRS), Radial Basis Function (RBF), and Kriging model. These have already been described in Section 3.3. These three different surrogate models  $s$  were constructed for each of the variables defined in Section 5.1.2, seeking to minimize the error between the metamodeled variable  $\hat{Z}_{m,tra} = g_{s,m}(X_{tra})$  and the actual values  $Z_{m,tra}$  at the training dataset  $X_{tra}$ .

An example is presented applying the surrogate Polynomial Response Surface model ( $s = PRS$ ) for the multi-fidelity intermediate variable “condensation difference” ( $m = \Delta$ ). This model determines the coefficients  $\beta$  of Equation (4) that minimize the error of comparing  $\hat{\Delta}$  (the metamodel of  $\Delta$ ) against  $\Delta$  itself at the set of training points. A  $\hat{\Delta}$  value of 0.89 kg/h is predicted for case 24 among the training data when applying the surrogate model. Comparing this value with the actual  $\Delta$  for such a working point (0.96 kg/h), an error of 7% appears between the observed and predicted values.

#### 5.3.4. Calculation of Error Metrics and Assessment of Surrogate Models

The error metrics described in Section 3.4 were finally calculated for the 15 combinations of surrogated models and variables above, as shown in Table 7. Coefficients of determination  $R^2$  (see Equation (7)) were obtained to assess the fitting of each surrogate model to the corresponding variable in the training dataset. This adjustment is also evaluated for intermediate variables when the corresponding surrogate represents the target variable  $z_t$ .

Generally speaking, good values in terms of  $R^2$  were obtained for most combinations, as Table 7 displays. The analysis showed perfect goodness of fit for the estimations employing radial basis function models and kriging models. However, when such surrogates employ the ratio of condensations  $\varphi$  between HF and LF models, their fitting to the training dataset deteriorates when the results are translated into values of HF condensation mass flow rate  $\dot{m}_{cond.,3D}$ , which is the target variable. Polynomial Response Surface models also provide excellent (but not perfect) calibrations, except for the ratio of condensations  $\varphi$ .

The ultimate test for the surrogate models is their ability to predict results at points not used for their training, i.e., the verification stage. To do so, several metrics based on the Mean Square Error (MAE) of Equation (8) were identified in Section 3.4, such as the MAE retaining 95% of data with the smallest residuals and the normalized  $\widetilde{MAE}$  (Equation (9)) and  $\widetilde{MAE}_\sigma$  (Equation (10)). Finally, Table 7 includes a qualitative assessment of the predictive capability of each metamodel based on the last two indicators, in accordance with the work of Roy et al. [60].

Table 7 shows that 12 combinations of surrogate models and variables provide good predictive capability; 2 obtain a moderate only one presents a lousy performance. The

kriging model is the only surrogate that provides evaluation deemed as “good” for all metamodeled variables. Interestingly, at least one MF variable performs better than the target variable for each surrogate model. Still, not all MF variables necessarily provide better agreement than the HF target variable (such as  $\varphi$  for PRS and  $\Delta$  and  $\eta$  for Kriging model). This idea was already noted by Fernández-Godino et al. [76], who identify that MF models “may be less accurate than a surrogate built using only HF training data points”.

**Table 7.** Evaluation of the surrogate models to predict water condensation for different metamodeled variables.

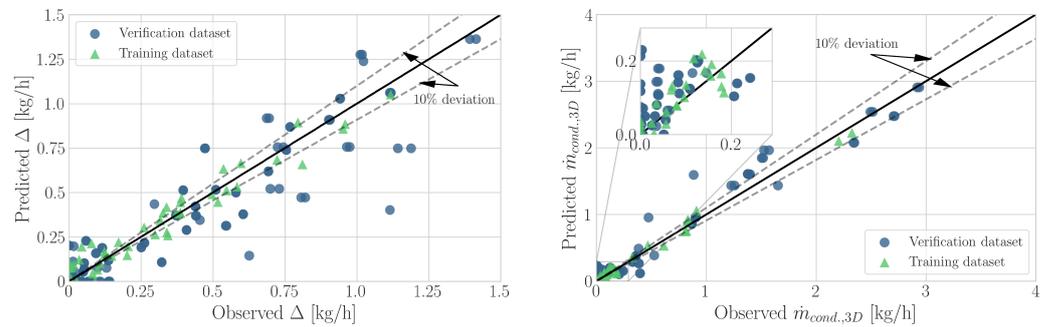
Surrogate Models		Training		MAE	Verification		$\widetilde{MAE}_\sigma$	Evaluation
		$R^2(z_m)$	$R^2(z_t)$		MAE 95%	$\widetilde{MAE}$		
PRS	$\dot{m}_{cond.,3D}$		0.94	0.1596	0.139	0.0601	0.1851	Good
	$w_{cond.,3D}$	0.95	0.98	0.1376	0.096	0.0415	0.1425	Good
	$\varphi$	0.56	0.39	0.4587	0.2428	0.1044	0.4605	Bad
	$\Delta$	0.95	0.98	0.0969	0.0793	0.0341	0.1001	Good
	$\eta$	0.97	1.0	0.1274	0.0829	0.0356	0.1726	Good
RBF	$\dot{m}_{cond.,3D}$ (Gauss.) *		1.0	0.1824	0.1553	0.0688	0.1988	Good
	$w_{cond.,3D}$ (Gauss.) *	1.0	1.0	0.1078	0.0826	0.0355	0.1252	Good
	$\varphi$ (Multi-quad.) *	1.0	0.74	0.1413	0.1061	0.0457	0.1867	Good
	$\Delta$ (Cubic) *	1.0	1.0	0.2348	0.2052	0.0883	0.2566	Moderate
	$\eta$ (Multi-quad.) *	1.0	1.0	0.1601	0.0925	0.0398	0.2313	Moderate
Kriging	$\dot{m}_{cond.,3D}$		1.0	0.2391	0.2176	0.0936	0.1964	Good
	$w_{cond.,3D}$	1.0	1.0	0.1409	0.1128	0.0473	0.1603	Good
	$\varphi$	1.0	0.85	0.1467	0.1066	0.0459	0.1916	Good
	$\Delta$	1.0	1.0	0.1616	0.1409	0.0606	0.1486	Good
	$\eta$	1.0	1.0	0.1171	0.0681	0.0293	0.1767	Good

\* On each case is selected the best transformation type to obtain the greatest  $R^2$  evaluating the study variable.

The results included in Table 7 support the current work of assessing not only MF but also HF intermediate variables. Indeed, the HF intermediate variable  $w_{cond.,3D}$  performs better than the HF target variable  $\dot{m}_{cond.,3D}$  for all surrogate models when applied to the problem of volume condensation due to mixing. Furthermore,  $w_{cond.,3D}$  is the best intermediate variable to be surrogated using a Radial Basis Function since it presents the lowest  $\widetilde{MAE}$  and  $\widetilde{MAE}_\sigma$  amongst all assessed variables, including the MF ones. For the Kriging model, there is not an MF variable that presents lower values of both  $\widetilde{MAE}$  and  $\widetilde{MAE}_\sigma$  than those of  $w_{cond.,3D}$ , so the best intermediate variable to be metamodeled using a Kriging model will depend on the weighting of such error metrics. Finally, the best overall combination of surrogate and variable is the Polynomial Response Surface of the MF variable  $\Delta$ , by the values of  $\widetilde{MAE}$  and  $\widetilde{MAE}_\sigma$  shown in Table 7.

To visualize the results underlying Table 7 and exemplify the implications of the multi-fidelity proposal, an analysis of the surrogate model based on a polynomial response surface acting over condensation difference (Equation (16)) is conducted since this combination is the one providing the lowest error, as aforementioned. Figure 7 shows the calibration plots for the observed and predicted (metamodeled) values of the intermediate variable  $\Delta$ , and the equivalent plot in terms of the target variable  $\dot{m}_{cond.,3D}$ . This latter calibration plot has been obtained on the basis of the surrogate of  $\Delta$  and the LF model  $\dot{m}_{cond.,0D}$  by means of Equation (17). Considering the training dataset (green triangles), the surrogate is constructed for the intermediate variable (see Figure 7) with a  $R^2(z_m)=0.95$ , increasing the correlation up to  $R^2(z_t)=0.98$  when representing the target variable. For the verification dataset (blue circles), the error for the difference is much greater than that of the target variable, in which a few points lie outside the depicted 10% dispersion band. A close-up of Figure 7 shows that the performance of the surrogate is less adequate close to the origin. However, this region corresponds to low mass flow rates, so their contribution to the absolute error is low (notice that the normalized  $\widetilde{MAE}$  for this combination is below 3.5%).

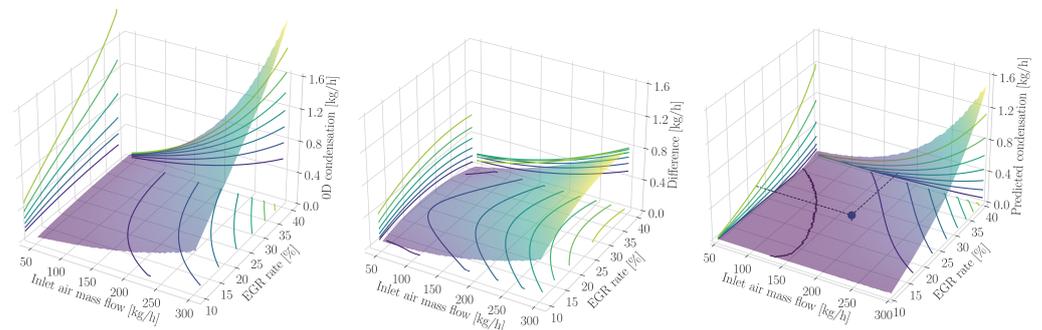
Indeed, when applying this MF surrogate model, it is worth reminding that greater impeller damage is produced with greater condensation mass flow rates [70].



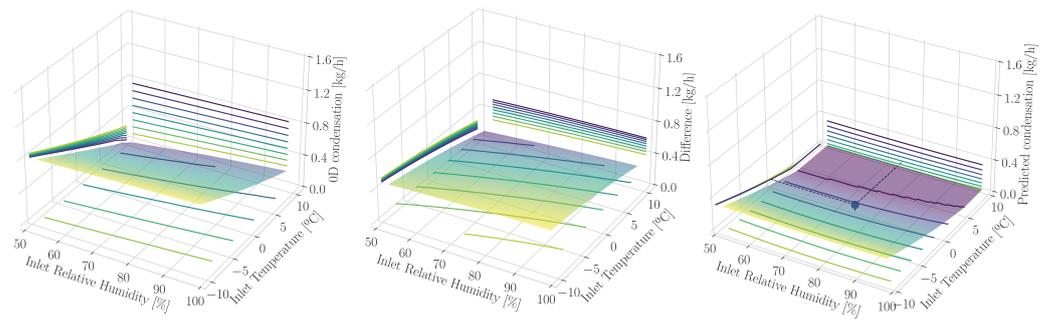
**Figure 7.** Assessment of the metamodeled variable  $\Delta$  and the target variable obtained by means of  $\Delta$ .

Finally, a graphical representation of the relation between the MF variables and the six significant input parameters regarded by ANOVA (see Section 5.2.2) is provided. Figures 8–10 show such 3D plots by considering pairs of parameters and leaving the omitted ones at their medium level (see Table 3). These figures represent low-fidelity, intermediate-variable, and high-fidelity features. Due to the negligible computational effort of the LF model employed in this work (see Section 4.2), the surface representation of 0D representations is obtained using an ad-hoc  $100 \times 100$  simulation campaign followed by linear interpolation. Intermediate figures showing differences are the graphical depictions of the corresponding polynomial response surface. Furthermore, Equation (17) allows us to obtain the metamodel of  $\dot{m}_{cond,3D}$  using the proposed MF strategy, represented in the predicted condensation. To further assess the method, the actual condensation mass flow rate calculated by the HF 3D CFD simulation is depicted as a small blue sphere in the rightmost subplots of Figures 8–10. Comparing this value with the one predicted using the MF PRS surrogate based on  $\Delta$ , a relative error of 0.23% is obtained for this specific point.

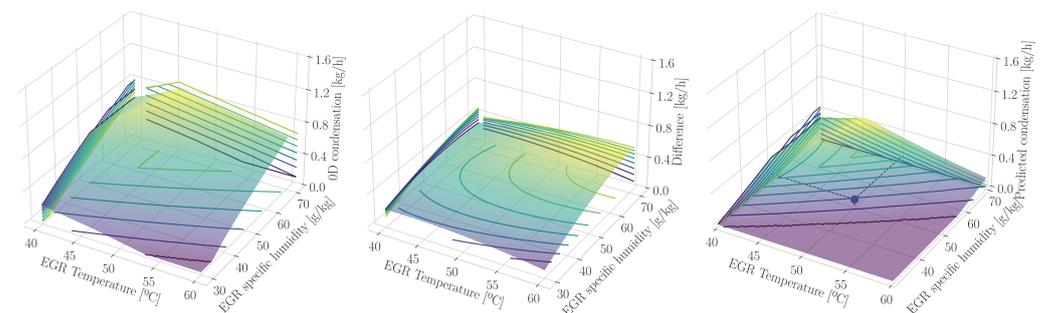
Considering these figures, the similarities between the LF prediction and the prediction of the target variable are likely to be one of the root causes of this combination providing the overall best goodness-of-fit. Indeed, Fernández-Godino et al. [76] considered that a strong correlation between LF and HF data indicates a good performance of the subsequent MF surrogate model. In this case, the metamodeled MF intermediate variable  $\Delta$  moderates this relationship by correcting the differences in magnitude.



**Figure 8.** Surface plot of  $\dot{m}_{cond,0D}$  calculations (left), PRS surrogate of  $\Delta$  (center) and subsequent  $\dot{m}_{cond,3D}$  MF surrogate (right) with varying inlet air mass flow and EGR rate, considering the remaining parameters at their medium values (see Table 3).



**Figure 9.** Surface plot of  $\dot{m}_{cond.,0D}$  calculations (left), PRS surrogate of  $\Delta$  (center) and subsequent  $\dot{m}_{cond.,3D}$  MF surrogate (right) with varying inlet relative humidity and inlet temperature, considering the remaining parameters at their medium values (see Table 3).



**Figure 10.** Surface plot of  $\dot{m}_{cond.,0D}$  calculations (left), PRS surrogate of  $\Delta$  (center) and subsequent  $\dot{m}_{cond.,3D}$  MF surrogate (right) with varying EGR temperature and EGR specific humidity, considering the remaining parameters at their medium values (see Table 3).

## 6. Concluding Remarks

The present work provides a methodological proposal for obtaining and assessing surrogate models for design space approximation. Apart from specific metamodelling of the target variable, the usage of intermediate HF variables and MF scaling functions to provide surrogates is addressed systematically and quantitatively for the first time, to the author's knowledge. Considering the steadily increasing production of new metamodelling in digital twins and machine learning frameworks, the proposed method could be useful to discern which combinations of surrogate models and intermediate variables should provide the best results for design space exploration. The developed method is founded upon a screening, training, and verification strategy, reusing the screening dataset for the verification stage to reduce the computational effort. After describing the comprehensive method, it is applied to the complex engineering problem of condensation prediction at three-way junctions. This original example validates the proposed methodology for the case of "real-life" problems compared with the synthetic and benchmark problems often found in the literature.

In this case study, the best reduction of the normalized MAEs has been found by constructing a polynomial response surface of a multi-fidelity additive scaling variable that computes the difference between the low-fidelity and high-fidelity predictions of condensation mass flow rate. These results confirm the utility of low-fidelity models in combination with selected campaigns of HF simulations to evaluate the performance of systems with otherwise highly computationally demanding models. Despite the overall performance of a PRS combination, the radial-basis function, and Kriging models obtained the best agreement for the training datasets, with the latter model providing "good" surrogates for all variables assessed in terms of normalized MAEs. Furthermore, HF intermediate variables (derived from the combination of target variables and model parameters) improve the accuracy of the surrogates built directly from the target variable, which is a benefit without resorting to MF models.

However, these findings are case-specific and cannot be generalized for application to other problems. Indeed, this work proposes a strategy to explore all combinations of surrogate models and intermediate variables (both HF and MF), evaluating them regarding the metamodeling error after the verification stage. For instance, in this work, there are two explored variables ( $\frac{\eta}{100}$  and  $\varphi$ ) whose summation yields one, yet they behave in a very different way. Therefore, we recommend assessing this double-entry matrix for each studied problem, as the computational effort of conducting simulation campaigns for complex engineering problems is much greater than that of constructing and evaluating surrogates. In future research, the specific algorithms to build the surrogates for intermediate variables in the training datasets could be modified to minimize not only the error of this intermediate variable but also the error in terms of the target variable, which may further improve their results in the verification stage.

Regarding the application of the developed method to the characterization of condensation produced in an LP-EGR junction, the proposed strategy has provided different findings. Both inlet and EGR turbulence intensities as well as outlet pressure were found to have a non-significant effect on condensation for all metamodeled variables assessed. Considering the best combination of surrogate (PRS) and multi-fidelity variable (condensation difference) found in this work, there are even less significant parameters (effect of inlet temperature and relative humidity can be neglected). Therefore, the campaign of 54 HF simulations required to build the surrogate and thus characterize the condensation produced by a certain junction geometry could be further reduced. With such affordable computational effort, several challenges could now be addressed in the framework of internal combustion engine cold start, such as determining the adequate moment to switch between HP and LP EGR modes [62,77] or introducing devices to limit the production or the impact of condensation, which could be a cooler bypass [78] and a water separator [79].

**Author Contributions:** Funding acquisition, J.G.; Project administration, J.G.; Supervision, J.G. and R.N.; Conceptualization, R.N.; Methodology, R.N., F.M. and A.C.; Software, F.M. and A.C.; Investigation, R.N. and F.M.; Formal analysis, F.M. and A.C.; Data curation, R.N., F.M. and A.C.; Visualization, F.M.; Writing-Original draft, F.M. and A.C.; Writing-Review and editing, R.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** Francisco Moya is partially supported through a FPI-GVA-ACIF-2019 grant of the Government of Generalitat Valenciana and the European Social Fund. This work has been partially supported by “Vicerrectorado de Investigación de la Universitat Politècnica de València” through grant number PAID-11-21.

**Data Availability Statement:** The data presented in this study are available in the article Supplementary Materials.

**Acknowledgments:** The authors of this paper wish to thank Guillermo García Olivas for his invaluable support to manage the CFD simulation campaign.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

### List of Symbols

$A$	Area	$m^2$
$EGR$	Exhaust gas re-circulation	%
$f_m$	Equation for target variable involving an intermediate variable $z_m$	–
$g$	surrogate model equation	–
$I$	turbulent intensity	%
$L$	number of levels	–
$\dot{m}$	mass flow rate	kg/h
$N, N'$	number of parameters before and after the screening	–

$P$	Pressure	barF
$RH$	Relative humidity	%
$R$	Training set range	–
$T$	Temperature	K
$w$	Specific humidity	g/kg
$x$	Parameter value for each points	–
$y$	Target variable (3D CFD condensation mass flow rate)	kg/h
$z$	Metamodeled variable	–
$\sigma$	standard deviation of the absolute error	kg/h
$\varepsilon$	Generic error	–
$\eta$	T-junction efficiency	%
$\Delta$	Difference between high and low fidelity condensation	kg/h
$\varphi$	ratio between high and low fidelity condensation	–

### Sub- and Superscripts

<i>air</i>	Intake air branch conditions
<i>cond.</i>	Condensation conditions
<i>egr</i>	EGR flow branch conditions
<i>exp</i>	Exploration dataset
<i>i, j</i>	Index for model parameters before and after the screening
<i>m</i>	Intermediate variable
<i>out</i>	Outlet branch conditions
<i>rate</i>	Ratio between air mass flow and EGR mass flow
<i>s</i>	Surrogate model
<i>scr</i>	Screening dataset
<i>t</i>	Target variable
<i>tra</i>	Training dataset
<i>ver</i>	Verification dataset
<i>0D</i>	Low fidelity simulation
<i>3D</i>	High fidelity simulation

### List of Abbreviations

ANOVA	Analysis of variance
CFD	Computer fluid dynamics
CO <sub>2</sub>	Carbon dioxide
Cubic	Cubic function
EGR	Exhaust gas recirculation
Gauss	Gaussian function
HF	High fidelity
Krig	Kriging
LF	Low fidelity
LP-EGR	Low pressure exhaust gas recirculation
MAE	Mean Square Error
$\widehat{MAE}$	Normalized Mean Square Error
$\widehat{MAE}_\sigma$	Standard deviation of normalized MAE
MF	Multi-fidelity
multi-quad	Multi-quadratic function
PRS	Polynomial Response surface
$R^2$	Coefficient of determination
RANS	Reynolds-Average Navier stokes
RBF	Radial basis function

### References

- Jiang, P.; Zhou, Q.; Shao, X. *Surrogate Model-Based Engineering Design and Optimization*; Springer Tracts in Mechanical Engineering; Springer: Singapore, 2020. [[CrossRef](#)]
- Rodic, B. Industry 4.0 and the New Simulation Modelling Paradigm. *Organizacija* **2017**, *50*, 193–207. [[CrossRef](#)]
- Barkanyi, A.; Chovan, T.; Nemeth, S.; Abonyi, J. Modelling for Digital Twins-Potential Role of Surrogate Models. *Processes* **2021**, *9*, 476. [[CrossRef](#)]

4. Wright, L.; Davidson, S. How to tell the difference between a model and a digital twin. *Adv. Model. Simul. Eng. Sci.* **2020**, *7*, 13. [[CrossRef](#)]
5. Chakraborty, S.; Adhikari, S.; Ganguli, R. The role of surrogate models in the development of digital twins of dynamic systems. *Appl. Math. Model.* **2021**, *90*, 662–681. [[CrossRef](#)]
6. Cheng, K.; Ling, C. Surrogate-assisted global sensitivity analysis: An overview. *Struct. Multidiscip. Optim.* **2020**, *61*, 1187–1213. [[CrossRef](#)]
7. Galindo, J.; Climent, H.; Navarro, R.; Garcia-Olivas, G. Assessment of the numerical and experimental methodology to predict EGR cylinder-to-cylinder dispersion and pollutant emissions. *Int. J. Engine Res.* **2021**, *22*, 3128–3146. [[CrossRef](#)]
8. Piqueras, P.; Ruiz, M.J.; Martin, J.; Tsolakis, A. Sensitivity of pollutants abatement in oxidation catalysts to the use of alternative fuels. *Fuel* **2021**, *297*, 120686. [[CrossRef](#)]
9. Li, Z.; Zheng, X. Review of design optimization methods for turbomachinery aerodynamics. *Prog. Aerosp. Sci.* **2017**, *93*, 1–23. [[CrossRef](#)]
10. Forrester, A.; Sobester, A.; Keane, A. Engineering Design via Surrogate Modelling. *Proc. R. Soc. A* **2007**, *463*, 3251–3269. [[CrossRef](#)]
11. Shan, S.; Wang, G. Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. *Struct. Multidiscip. Optim.* **2010**, *41*, 219–241. [[CrossRef](#)]
12. Galindo, J.; Navarro, R.; García-Cuevas, L.M.; Tari, D.; Tartoussi, H.; Guilain, S. A zonal approach for estimating pressure ratio at compressor extreme off-design conditions. *Int. J. Engine Res.* **2019**, *20*, 393–404. [[CrossRef](#)]
13. Galindo, J.; Tiseira, A.; Navarro, R.; Tari, D.; Tartoussi, H.; Guilain, S. *Compressor Efficiency Extrapolation for 0D–1D Engine Simulations*; SAE Technical Paper; SAE International: Warrendale, USA, 2016. [[CrossRef](#)]
14. Tirnovan, R.; Giurgea, S.; Miraoui, A.; Cirrincione, M. Surrogate modelling of compressor characteristics for fuel-cell applications. *Appl. Energy* **2008**, *85*, 394–403. [[CrossRef](#)]
15. Leylek, Z.; Neely, A.J. Global Three-Dimensional Surrogate Modeling of Gas Turbine Aerodynamic Performance. In Proceedings of the ASME Turbo Expo 2017: Power for Land, Sea and Air, Charlotte, NC, USA, 26–30 June 2017; ASME: New York, NY, USA, 2017; pp. 1–12. [[CrossRef](#)]
16. Qian, Z.; Seepersad, C.; Roshan, V.; Allen, J. Building Surrogate Models Based on Detailed and Approximate Simulations. *J. Mech. Des.* **2006**, *128*, 668–677. [[CrossRef](#)]
17. Sanchez, F.; Budinger, M.; Hazyuk, I. Dimensional analysis and surrogate models for the thermal modeling of Multiphysics systems. *Appl. Therm. Eng.* **2017**, *110*, 758–771. [[CrossRef](#)]
18. Xin, C.; Li, L.; Teng, L.; Zhenjiang, Y. A reduced order aerothermodynamic modeling framework for hypersonic vehicles based on surrogate and POD. *Chin. J. Aeronaut.* **2015**, *28*, 1328–1342. [[CrossRef](#)]
19. Luján, J.M.; Climent, H.; García-Cuevas, L.M.; Moratal, A. Volumetric efficiency modelling of internal combustion engines based on a novel adaptive learning algorithm of artificial neural networks. *Appl. Therm. Eng.* **2017**, *123*, 625–634. [[CrossRef](#)]
20. Li, K.Q.; Liu, Y.; Yin, Z.Y. An improved 3D microstructure reconstruction approach for porous media. *Acta Mater.* **2023**, *242*, 118472. [[CrossRef](#)]
21. Yondo, R.; Andres, E.; Valero, E. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. *Prog. Aerosp. Sci.* **2018**, *96*, 23–61. [[CrossRef](#)]
22. Reihani, A.; Hoard, J.; Klinkert, S.; Kuan, C.K.; Styles, D.; McConville, G. Experimental response surface study of the effects of low-pressure exhaust gas recirculation mixing on turbocharger compressor performance. *Appl. Energy* **2020**, *261*, 114349. [[CrossRef](#)]
23. Simpson, T.; Mauery, T.; Norte, J.; Mistree, F. Kriging models for global approximation in simulation-based multidisciplinary design optimization. *AIAA J.* **2001**, *39*, 2233–2241. [[CrossRef](#)]
24. Fang, H.; Horstemeyer, M. Global response approximation with radial basis functions. *Eng. Optim.* **2006**, *38*, 407–424. [[CrossRef](#)]
25. Williams, B.; Cremaschi, S. Selection of surrogate modeling techniques for surface approximation and surrogate-based optimization. *Chem. Eng. Res. Des.* **2021**, *170*, 76–89. [[CrossRef](#)]
26. Jia, L.; Alizadeh, R.; Hao, J.; Wang, G.; Allen, J.; Mistree, F. A rule-based method for automated surrogate model selection. *Adv. Eng. Inform.* **2020**, *45*, 101123. [[CrossRef](#)]
27. Jin, Y. Surrogate-assisted evolutionary computation: Recent advances and future challenges. *Swarm Evol. Comput.* **2011**, *1*, 61–70. [[CrossRef](#)]
28. Tang, Y.; Chen, J.; Wei, J. A surrogate-based particle swarm optimization algorithm for solving optimization problems with expensive black box functions. *Eng. Optim.* **2013**, *45*, 557–576. [[CrossRef](#)]
29. Jin, Y.; Olhofer, M.; Sendhoff, B. A framework for evolutionary optimization with approximate fitness functions. *IEEE Trans. Evol. Comput.* **2002**, *6*, 481–494. [[CrossRef](#)]
30. Akhtar, T.; Shoemaker, C.A. Multi objective optimization of computationally expensive multi-modal functions with RBF surrogates and multi-rule selection. *J. Glob. Optim.* **2016**, *64*, 17–32. [[CrossRef](#)]
31. Xu, H.; You, X.; Pop, I. Analytical approximation for laminar film condensation of saturated stream on an isothermal vertical plate. *Appl. Math. Model.* **2008**, *32*, 738–748. [[CrossRef](#)]
32. Fernandez, G.; Park, C.; Kim, N.; Haftka, R. Issues in Deciding Whether to Use Multifidelity Surrogates. *AIAA J.* **2019**, *57*, 2039–2054. [[CrossRef](#)]

33. Peherstorfer, B.; Willcox, K.; Gunzburger, M. Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization. *Soc. Ind. Appl. Math.* **2018**, *60*, 550–591. [[CrossRef](#)]
34. Razaaly, N.; Persico, G.; Congedo, P. Multi-fidelity surrogate-based optimization of transonic and supersonic axial turbine profiles. In *Turbo Expo: Power for Land, Sea, and Air*; ASME: New York, NY, USA, 2020.
35. Chakraborty, S.; Chatterjee, T.; Chowdhury, R.; Adhikari, S. A surrogate model based multi-fidelity approach for robust esign optimization. *Appl. Math. Model.* **2017**, *47*, 726–744. [[CrossRef](#)]
36. Bastidas, C.; DeFilippo, M.; Chrysosostomidis, C.; Karniadakis, G.E. A Multifidelity Framework and Uncertainty Quantification for Sea Surface Temperature in the Massachusetts and Cape Cod Bays. *Earth Space Sci.* **2020**, *7*, e2019EA000954. [[CrossRef](#)]
37. Vishal, K.; Ganguli, R. Multi-fidelity analysis and uncertainty quantification of beam vibration using co-kriging interpolation method. *Appl. Math. Comput.* **2021**, *398*, 125987. [[CrossRef](#)]
38. Hebbal, A.; Brevault, L.; Balesdent, M.; Talbi, E.; Melab, N. Multi-fidelity modeling with different input domain definitions using Deep Gaussian processes. *Struct. Multidiscip. Optim.* **2021**, *63*, 2267–2288. [[CrossRef](#)]
39. Xiao, M.; Zhang, G.; Breitkopf, P.; Villon, P. Extended Co-Kriging interpolation method based on multi-fidelity data. *Appl. Math. Comput.* **2018**, *323*, 120–131. [[CrossRef](#)]
40. Vojkuvkova, P.; Sikula, O.; Weyr, J. Assessment of condensation of water vapor in the mixing chamber by CFD method. In *Proceedings of the EPJ Web of Conferences*; EDP Sciences: Les Ulis, France, 2015; Volume 92. [[CrossRef](#)]
41. Serrano, J.R.; Piqueras, P.; Angiolini, E.; Meano, C.; De La Morena, J. *On Cooler and Mixing Condensation Phenomena in the Long-Route Exhaust Gas Recirculation Line*; SAE Technical Paper; SAE International: Warrendale, PA, USA, 2015. [[CrossRef](#)]
42. Serrano, J.; Piqueras, P.; Navarro, R.; Tarí, D.; Meano, C. Development and verification of an in-flow water condensation model for 3D-CFD simulations of humid air streams mixing. *Comput. Fluids* **2018**, *167*, 158–165. [[CrossRef](#)]
43. Box, G.; Behnken, D. Some New Three Level Designs for the Study of Quantitative Variables. *Technometrics* **1960**, *2*, 455–475. [[CrossRef](#)]
44. Ferreira, S.; Bruns, R.; Ferreira, H.; Matos, G. Box-Behnken design: An alternative for the optimization of analytical methods. *Anal. Chim. Acta* **2007**, *597*, 179–186. [[CrossRef](#)] [[PubMed](#)]
45. Song, W.; Keane, A. Parameter screening using impact factors and surrogate-based ANOVA techniques. In Proceedings of the 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization, Portsmouth, NH, USA, 6–8 September 2006; AIAA: Reston, VA, USA, 2006; pp. 188–196. [[CrossRef](#)]
46. Silva, V.B.; Rouboa, A. Optimizing the DMFC operating conditions using a response surface method. *Appl. Math. Comput.* **2012**, *218*, 6733–6743. [[CrossRef](#)]
47. Díaz, J.D.C.; Nieto, P.G.; Castro-Fresno, D.; Rodríguez, P.M. Steady state numerical simulation of the particle collection efficiency of a new urban sustainable gravity settler using design of experiments by FVM. *Appl. Math. Comput.* **2011**, *217*, 8166–8178. [[CrossRef](#)]
48. Chen, L.; Ma, Y.; Guo, Y.; Zhang, C.; Liang, Z.; Zhang, X. Quantifying the effects of operational parameters on the counting efficiency of a condensation particle counter using response surface Design of Experiments (DOE). *J. Aerosol Sci.* **2017**, *106*, 11–23. [[CrossRef](#)]
49. Subramani, S.; Govindasamy, R.; Lakshmi, G.; Rao, N. Predictive correlations for NOx and smoke emission of DI CI engine fuelled with diesel-biodiesel-higher alcohol blends-response surface methodology approach. *Fuell* **2020**, *269*, 117304. [[CrossRef](#)]
50. Lenth, R.V. Response-Surface Methods in R, Using rsm. *J. Stat. Softw.* **2009**, *32*, 1–17. [[CrossRef](#)]
51. Rippa, S. An algorithm for selecting a good value for the parameter  $c$  in radial basis function interpolation. *Adv. Comput. Math.* **1999**, *11*, 193–210. [[CrossRef](#)]
52. Krige, D.G. A Statistical Approach to Some Mine Valuation and Allied Problems on the Witwatersrand: By DG Krige. Ph.D. Thesis, University of the Witwatersrand, Johannesburg, South Africa, 1951.
53. Bouhrel, M.A.; Hwang, J.T.; Bartoli, N.; Lafage, R.; Morlier, J.; Martins, J.R. A Python surrogate modeling framework with derivatives. *Adv. Eng. Softw.* **2019**, *135*, 102662 [[CrossRef](#)]
54. Hu, Y.; Lu, Z.; Wei, N.; Jiang, X.; Zhou, C. Advanced single-loop Kriging surrogate model method by combining the adaptive reduction of candidate sample pool for safety lifetime analysis. *Appl. Math. Model.* **2021**, *100*, 80–95. [[CrossRef](#)]
55. Xiao, M.; Breitkopf, P.; Filomeno, R.; Knopf-Lenoir, C. Model reduction by CPOD and Kriging-application to the shape optimization of an intake port. *Struct. Multidiscip. Optim.* **2010**, *41*, 555–574. [[CrossRef](#)]
56. Bagheri, S.; Konen, W.; Bäck, T. Comparing kriging and radial basis function surrogates. In Proceedings of the 27 Workshop Computational Intelligence, Workshop Computational Intelligence, Dortmund, Germany, 23–24 November 2017; KIT Scientific Publishing: Karlsruhe, Germany, 2017; pp. 243–259.
57. Chandrashekarappa, P.; DuVigneau, R. *Radial Basis Functions and Kriging Metamodels for Aerodynamic Optimization*; INRIA Document; Unité de Recherche INRIA Sophia Antipolis: Nice, France, 2007; p. 44.
58. Legates, D.; McCabe, G. Evaluating the Use Of “Goodness-of-Fit” Measures in Hydrologic and Hydroclimatic Model Validation. *Water Resour. Res.* **1999**, *35*, 233–241. [[CrossRef](#)]
59. Davydenko, A.; Fildes, R. Measuring Forecasting Accuracy: Problems and Recommendations (by the example of SKU-level judgmental adjustments). *Intell. Fash. Forecast. Syst. Model. Appl.* **2013**, *43*, 43–70. [[CrossRef](#)]
60. Roy, K.; Nayanan, R.; Ambure, P.; Aher, R. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemom. Intell. Lab. Syst.* **2016**, *152*, 18–33. [[CrossRef](#)]

61. Park, Y.; Bae, C. Experimental study on the effects of high/low pressure EGR proportion in a passenger car diesel engine. *Appl. Energy* **2014**, *133*, 308–316. [[CrossRef](#)]
62. Luján, J.M.; Guardiola, C.; Pla, B.; Reig, A. Switching strategy between HP (high pressure)-and LPEGR (low pressure exhaust gas recirculation) systems for reduced fuel consumption and emissions. *Energy* **2015**, *90*, 1790–1798. [[CrossRef](#)]
63. Galindo, J.; Navarro, R.; Tari, D.; Moya, F. Development of an experimental test bench and a psychrometric model for assessing condensation on a Low Pressure EGR cooler. *Int. J. Engine Res.* **2020**, *22*, 1540–1550. [[CrossRef](#)]
64. Song, H.; Song, S. Numerical investigation on a dual loop EGR optimization of a light duty diesel engine based on water condensation analysis. *Appl. Therm. Eng.* **2021**, *182*, 116064. [[CrossRef](#)]
65. Galindo, J.; Navarro, R.; Tari, D.; García-Olivas, G. Centrifugal compressor influence on condensation due to Long Route-Exhaust Gas Recirculation mixing. *Appl. Therm. Eng.* **2018**, *144*, 901–909. [[CrossRef](#)]
66. Galindo, J.; Gil, A.; Navarro, R.; García-Olivas, G. Numerical assessment of mixing of humid air streams in three-way junctions and impact on volume condensation. *Appl. Therm. Eng.* **2022**, *201*, 14. [[CrossRef](#)]
67. Karstadt, S.; Werner, J.; Münz, S.; Aymanns, R. Effect of water droplets caused by low pressure EGR on spinning compressor wheels. In Proceedings of the 19th Supercharging Conference Dresden, Aachen, Germany, 17–18 September 2014.
68. Galindo, J.; Navarro, R.; Tari, D.; Moya, F. Analysis of condensation and secondary flows at T-junctions using optical visualization techniques and Computational Fluid Dynamics. *Int. J. Multiph. Flow* **2021**, *141*, 103674. [[CrossRef](#)]
69. Galindo, J.; Navarro, R.; Tari, D.; Moya, F. Quantitative validation of an in-flow water condensation model for 3D-CFD simulations of three-way junctions using indirect condensation measurements. *Int. J. Therm. Sci.* **2022**, *172*, 107303. [[CrossRef](#)]
70. Galindo, J.; Piqueras, P.; Navarro, R.; Tari, D.; Meano, C. Validation and sensitivity analysis of an in-flow water condensation model for 3D-CFD simulations of humid air streams mixing. *Int. J. Therm. Sci.* **2019**, *136*, 410–419. [[CrossRef](#)]
71. Menter, F.R. Two-equation eddy-viscosity turbulence models for engineering applications. *AIAA J.* **1994**, *32*, 1598–1605. [[CrossRef](#)]
72. Kimura, N.; Ogawa, H.; Kamide, H. Experimental study on fluid mixing phenomena in T-pipe junction with upstream elbow. *Nucl. Eng. Des.* **2010**, *240*, 3055–3066. [[CrossRef](#)]
73. Rezaeiravesh, S.; Vinuesa, R.; Schlatter, P. Towards multifidelity models with calibration for turbulent flows. In Proceedings of the 14th WCCM-ECCOMAS Congress, Online, 11–15 January 2021. .
74. Guilain, S.; Boubennec, R.; Doublet, M.; Clement, C.; Navarro, R.; Tari, D.; Moya, F. Condensation before compressor: A key issue of Low Pressure EGR in Eu7 context. In Proceedings of the 24th Supercharging Conference 2019, Aufladetechnische Konferenz, Dresden, 26–27 September 2019.
75. Galindo, J.; Climent, H.; Navarro, R. Modeling of EGR Systems. In *1D and Multi-D Modeling Techniques for IC Engine Simulation*; Onorati, A.; Montenegro, G., Eds.; SAE International: Warrendale, PA, USA, 2020; Chapter 7, pp. 257–278. [[CrossRef](#)]
76. Fernandez, M.; Dubreuil, S.; Bartoli, N.; Gogu, C. Linear regression-based multifidelity surrogate for disturbance amplification in multiphase explosion. *Struct. Multidiscip. Optim.* **2019**, *60*, 2205–2220. [[CrossRef](#)]
77. Cornolti, L.; Onorati, A.; Cerri, T.; Montenegro, G.; Piscaglia, F. 1D simulation of a turbocharged Diesel engine with comparison of short and long EGR route solutions. *Appl. Energy* **2013**, *111*, 1–15. [[CrossRef](#)]
78. Galindo, J.; Dolz, V.; Monsalve-Serrano, J.; Maldonado, M.; Odillard, L. Advantages of using a cooler bypass in the low-pressure exhaust gas recirculation line of a compression ignition diesel engine operating at cold conditions. *Int. J. Engine Res.* **2020**, *22*, 1624–1635. [[CrossRef](#)]
79. Choi, J.; Satpathy, S.; Hoard, J.; Styles, D.; Kuan, C.K. An experimental and computational analysis of water condensation separator within a charge air cooler. In Proceedings of the ASME 2017 Internal Combustion Engine Division Fall Technical Conference, Seattle, WA, USA, 15–18 October 2017; p. 11. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.