Machine Learning-Aided Design Optimisation of a Mechanical Micromixer

F.-J. Granados-Ortiz*, J. Ortega-Casanova**
School of Industrial Engineering, Department of Mechanical, Thermal and Fluid Engineering, University of Málaga
C/ Dr Ortiz Ramos s/n, 29071 Málaga, Spain
*Corresponding author: fjgranados@uma.es
**e-mail: jortega@uma.es

Abstract

In real-life mechanical engineering applications it is often complex to achieve an optimal multi-objective design, because of the costs related to fabrication and test of different prototypes. For this reason, the use of computational tools is a recommended practice. In this work, the design of an efficient mixing mechanical device composed of a rectangular pillar confined in a microchannel is aided by Machine Learning techniques (addressed as Machine Learning-Aided Design Optimisation, MLADO, proposed in this work). A Random Forest classifier is trained to predict which geometric configuration may lead to vortex shedding. Later, a multi-objective optimisation problem is investigated, which consists of minimising the required pumping power and maximising the mixing efficiency under some design constrains. If extra training data are desired for surrogates, the Random Forest classifier can be used to predict whether is worthy or not to simulate the new configuration, avoiding to run irrelevant computational intensive cases and accelerating the data-driven process. The resulting optimal designs from using the NSGA-II Genetic Algorithm on the surrogates are simulated and their performance is shown. The optimal geometric configurations, even for very unfavourable mixing conditions and a medium-low Reynolds number of 200, give a maximum mixing efficiency of around 50% at very low pumping power cost in a short channel, outperforming existing devices in the literature. The MLADO framework followed in this work can be easily extendable and automated for other similar design processes in mechanical engineering at any scale, by including shape parametrisation strategies.

Keywords: Vortex shedding, Machine Learning, Micromixing, CFD, Mechanical Engineering.

1. Introduction

To achieve a proper mixing is crucial for many medical, chemical, biological and industrial applications, such as medical devices to prevent thrombosis formation [1], reaction enhancement in chemical&pharmaceutical applications [2], gas turbine combustors [3] or heat transfer augmentation using vortex generators and nanofluids [4]. In addition, micromechanics&microfluidics is a field of research which has gained plenty of attention in recent years due to its multiple applications in the mentioned areas. Among some of these applications one could highlight LOC (Lab on a Chip) devices, which are small devices that integrate one or more laboratory functions on a single chip [5]. Within the field of mixing, this type of device has some advantages, such as low fluid consumption, fast response times due to its small dimensions and the possibility of mass production. Examples of LOC devices include those used in clinical analysis for the early detection of cancer or for the detection of diseases such as HIV, malaria or many other pathologies [6, 7]. An important field of application of mixing is the COVID-19 disease. The recent global pandemic highlights the use and need to develop inexpensive and rapid microdetection devices, since some parts of the current RT-PCR (Reverse Transcriptase-Polymerase Chain Reaction) detection protocol involve mixing and high cost centrifugation with body fluid samples and chemical reagents [8]. Furthermore, some vaccines for COVID-19, such as the Pfizer-BioNTech COVID-19 Vaccine, requires the mixing with a diluent in the preparation right before administration [9]. This mixing is performed by trained nurses manually by inverting a container several times and under the indication of no shaking [10].

This work aims to design an efficient, passive and cheap microdevice for heat and/or species mixing by means of vortex shedding mechanism generated when two fluids interact with a pillar structure confined in a channel. Opposite to the passive microdevices, where no additional energy needs are required (except the pressure difference for fluid motion), are the active microdevices. For active microdevices, extra energy consumption is necessary to enhance the efficiency of the devices. In this sense, previous works of the same research group [11, 12, 13] carried out studies about microdevices with a moving central pillar in a channel which undergoes heaving, pitching, and heaving&pitching motion, respectively. For the three different tested motions, optimal values of the input parameters (frequency and amplitude of each motion) are given for their best performance. Regarding the present microdevice, a sketch of its geometry is depicted in Figure 1. Despite the main purpose of this device is a microscale application within the UMA18-FEDERJA-184 project,

the performance of this device can be applied to larger scale mechanics. The parameters of interest for optimal design are pillar dimensions, in terms of the aspect ratio AR and blockage ratio BR, and the Reynolds number Re. The axial position L_u of the pillar with respect to the inlet is not relevant since the flow is assumed to be fully-developed across the 2D channel of width H. The downstream length is also assumed to be long enough as no impact on the simulation is noticed by the outlet. More details on the geometry and computation will be given in Section 2.

The fluid mechanism responsible for an adequate mixing in this device is the vortex shedding generated downstream the pillar. This well-known flow behaviour was studied by [14] more than a century ago, followed by [15, 16], and of course by [17], which finally led to the assignation of the "von Karman streets" name to this type of fluid wake. Since then, these flow structures have been vastly investigated by a large number of authors, either from a theoretical, numerical or experimental side. Some authors have focused for instance on the use of different pillar shapes [18, 19], on analysing the frequency of the shedding [20, 21], the noise emission [22, 23], the species/heat mixing effect [24, 25], etc. Despite vortex shedding has been studied in the literature extensively, there are still many recent and on-going investigations on the flow physics of this phenomenon. For instance, [26] studied experimentally vortex shedding past a circular cylinder in the subcritical regime. The early separation at this regime leads to large wakes, which are studied by means of Particle Image Velocimetry (PIV) and pressure surface measurements. Proper Orthogonal Decomposition (POD) is also applied to study the pressure fluctuations, observing primarily two modes (antisymmetric mode and symmetric mode), which determine the vortex shedding characteristics. Vortex shedding wakes of a circular cylinder with fins are analysed experimentally in [27]. They analysed the flow in the subcritical regime and using the same techniques as [26]. In their study they concluded that the addition of fins to a cylinder changes the flow distribution around the cylinder. They observed that as the diameter ratio between the cylinder and fins is increased, the recirculation region behind is reduced. In [28] a laminar flow through arrays of square cylinders is studied numerically to predict the transition behaviour depending on the density of pins (solid fraction). Several interesting relations are found to demonstrate how the critical Reynolds number, individual critical Reynolds number and the Roshko number are influenced by the solid fraction. In [25] laminar flow-induced vibration of a D-section cylinder in a channel is studied for heat transfer enhancement by means of numerical simulations. In this work, the channel walls are heated and the effect of the galloping motion of the D-section cylinder on the generation of

vortex streets is analysed. They concluded that the disruption of the thermal boundary layer due to the interaction with the D-section cylinder and the downstream periodic vortical structures is critical for the increase in heat transfer. Finally, some researchers have oriented recently their efforts to study the suppression of vortex shedding as, for instance, the investigation carried out in [29]. In their work, a complete compressible flow model including thermophysical properties is used to replicate large-scale heating effects of a laminar flow past a square cylinder. The effect of large-scale heating effects in the suppression of vortex shedding is discussed, and the effect of buoyancy effects due to variations in transport properties is found important in the suppression. For a more extensive review on the many different aspects of the present configuration studied in the literature, we refer the reader to our precedent paper [30], where the onset of the vortex shedding is characterized for a fixed value of BR. In [30] is also shown the good performance of the vortex shedding mechanism for mixing fluids in a specific configuration, concluding that vortex shedding can enhance the mixing efficiency more than a 50% and decrease the mixing cost two order of magnitude with respect to configurations without vortex shedding. However, despite the large research evidence in the literature, there are still applications of this flow problem under consideration for improvement, as the present work. Therefore, the current investigation could be considered a continuation of the univariate exploratory investigation developed in [30], but in the present work a multi-objective optimisation framework involving all parameters in a multivariate approach and Machine Learning classification algorithms is introduced and tested to design an optimal microdevice.

An initial objective in the present paper is to be able to predict whether vortex shedding may or may not take place. It is important to know beforehand if a design will lead to this phenomenon, because of the importance in the mixing process and the limitations in the AR, BR and Re. Since the potential applicability of the device is for micromechanics&microfluidics in medical, pharmaceutical or biological applications, the maximum Re, due to the very reduced dimensions of the device, will be restricted to $Re \leq 200$. On the other hand, AR and BR are constrained by the geometry of the pillar and channel itself. Therefore, under these constrains may be complex to achieve a good design. To predict the appearance of vortex shedding in this work, Machine Learning classification algorithms [31, 32, 33] are used. According to [34], Machine Learning "usually refers to changes in systems that perform tasks associated with artificial intelligence (AI)". It must be outlined that the present investigation is aimed at achieving vortex shedding, because of the

enhancement of mixing. However, the Machine Learning predictive models provided in this work can be used to avoid designs that lead to vortex shedding because of potential undesired effects e.g. in efficiency of downstream elements [35, 36], generate resonance [37, 38] or noise emission [39, 40, 41]. Within this design context, the use of this type of predictive models can be a matter of interest in the Digital Twin modelling [42], which consists of the integration of virtual and physical designs to share data between them constantly. This makes Digital Twin one of the most promising technologies to revolutionise smart manufacturing in industry 4.0.

Concretely, Logistic Regression and Random Forest are the classification techniques used in the present work. Logistic Regression (LR) [43, 44, 45] is used as a model to assign a probability, which is a number ranging between 0 and 1. The LR has the following form:

$$P^{LR}(\boldsymbol{\xi}) = \frac{1}{1+\exp(-z(\boldsymbol{\xi}))}, \tag{1}$$

with $\boldsymbol{\xi}$ the vector of design variables space and the regression function $z(\boldsymbol{\xi})$ modelled as:

$$z(\boldsymbol{\xi}) = \alpha_0 + \sum_{i=1}^{k} \alpha_i z_i(\boldsymbol{\xi}), \tag{2}$$

where α_i are the coefficients to estimate and $z_i(\boldsymbol{\xi})$ are the independent literal function terms of the monomials. These can be designed for linear or non-linear regression. $z_i(\boldsymbol{\xi})$ are the design space variables ξ_i in case of first order linear regression modelling. Higher-order polinomials or variable transformations can be also tried in linear logistic regression. Once the probability is calculated, one can define a threshold above which the cases can be classified as positive (1) and vice-versa to define negatives (0). This threshold is usually the intermediate value 0.5, but may change depending on the application. LR is a very popular predictive model in medical and biological sciences, due to the facility to interpret the results, as opposite to, e.g., Deep Learning methods [46]. However, no applications in industrial aerodynamics have been found in our literature survey. We rather suggest the reader to see medical applications in fluid dynamics [47, 48, 49] to comprehend the goodness of the method.

Another interesting but more complex predictor model is the namely Random Forest (RF), first presented by Breiman [32]. RF is an advanced decision tree method [50, 51], which consists of a combination of trees in an ensembled learning way, and by creating these predictor trees based on a randomised sampling (namely bootstrap with or without replacement, which means to re-select or not already selected samples) on the original data set. Ensemble learning means to use different

predictors trained for the same task to further combine them [52]. RF has been established in the Machine Learning literature as a very robust and accurate method, becoming one of the most popular predicting algorithms in the current *Big Data* era [53], as can be seen in the vast amount of works in the literature. For this reason, since in the present work there is no new development on the method and this is an application of RF, no theoretical descriptions are given here. We suggest the reader to see these popular books and publications on the RF algorithm formalities: [32, 54, 55, 56].

To the knowledge of the authors, there are no works in the literature on the use of such classification techniques to predict the generation of vortex shedding, which is of importance in this work to achieve a desired mixing. The predictors also permit to decide whether or not simulate a new case (sample point) for the functions to evaluate in the optimisation process, as will be explained later. Despite this dearth of literature, other few works using also Artificial Intelligent (AI) algorithms for flow predictions can be found in previous works (not considering turbulence, of course, which is a field where the use of AI in fluid dynamics is becoming a hot-topic). For instance in the very recent work in [57], Deep Neural Networks are used to study the Vortex Induced Vibrations (VIV) for a given structure. In this work, the authors demonstrated that their predictive model (namely *Physics Informed Learning Machines* [58]) can infer the flow characteristic from scattered snapshot data. They did demonstrate that can infer lift and drag from limited velocity data, and also obtained other quantities of interest with the same procedure. The authors further demonstrated that from smoke or dying visualisations one can use these AI techniques and extract flow quantities of interest, namely "Hidden Fluid Dynamics", whose work has been published last year in Science [59]. In [60], Deep Neural Networks and Random Forest are used for tracking of chemical leaks using CFD data, which is an important issue in chemical plants due to the potential disastrous effects that may take place in the event of leak. In [61], Random Forests are used to predict the degree of erosion due to solid particles in pipe elbows, problem specially relevant in the oil&gas industry, since severe damage can lead to hydrocarbon leaks. Several Machine Learning techniques were tested, and Random Forest performed the best. In [62], cerebral aneurysm is studied by means of creating a Supported Vector Machines (SVM) classificator for rupture risk prediction, which is supported by CFD simulations. No more Machine Learning classification-based works were identified during our review. However, Machine Learning is a broad field, which includes more applications than classification. Many authors did apply

Machine Learning in surrogate-based optimisation [63, 64] or Uncertainty Quantification (UQ) [65, 66, 67]. Among these techniques used in optimisation or UQ one can find the Gaussian Processes or Kriging interpolation [34], which are used in this work to generate surrogates, also known as Response Surfaces or metamodels. These surrogates are the starting point for the optimisation process. Kriging aims at obtaining an approximation of an exact model with the form:

$$y(\boldsymbol{\xi}) = \hat{y}(\boldsymbol{\xi}) + \epsilon(\boldsymbol{\xi}), \tag{3}$$

where y is the exact model, $\hat{y}$ is the surrogate model and ϵ the error between the surrogate and the exact model, all defined in the parameter space $\boldsymbol{\xi}$. Kriging is deployed in two-steps: first a regression function $f(\boldsymbol{\xi})$ is generated based on the data set, and which is intended to capture the largest variance in data (the general "trend"). Second, a Gaussian process $Z(\boldsymbol{\xi})$ is built upon, obtaining the expression:

$$\hat{y}^{KG}(\boldsymbol{\xi}) = \hat{f}(\boldsymbol{\xi}) + Z(\boldsymbol{\xi}) = \sum_{i=1}^{k} \gamma_i f_i(\boldsymbol{\xi}) + Z(\boldsymbol{\xi}), \tag{4}$$

where $f(\boldsymbol{\xi})$ stands for the $k \times 1$ vector of basis regression functions $[f_1(\boldsymbol{\xi})\ f_2(\boldsymbol{\xi})...f_k(\boldsymbol{\xi})]$ and γ_i denotes the coefficients. Depending on the order of the regression, the method is called Universal Kriging ($f(\boldsymbol{\xi})$ is a multivariate polynomial of order $n \geq 1$), Ordinary Kriging ($f(\boldsymbol{\xi}) = \gamma_0$, with γ_0 an unknown coefficient to estimate) or Simple Kriging ($f(\boldsymbol{\xi}) =$ constant). There is a vast literature on the method on the theoretical and practical aspects. For further details see for instance [68, 69, 70, 71]. Examples of application of this popular method are aircraft wing optimisation under a large number of design parameters [72], optimisation of an opening ventilator shaft for the ventilation of an underground shelter [73], the aerodynamic optimisation in civil engineering [74], CFD-based optimisation of aeronautical combustion chambers [75] or optimal design of centrifugal pumps [76].

Once the surrogates are available as costless functions to evaluate, these have to be explored to achieve the optimal design. For this purpose, the popular Non-dominated Sorting Genetic Algorithm NSGA-II algorithm is used [77], which is an improved version of the original NSGA algorithm [78, 79]. NSGA-II is a powerful Genetic Algorithm (GA) with a formulation of non-elitism. The lack of elitism means that each solution is "ranked" according to the level of non-domination, for the sake of aiming at non-dominated solutions more directly. GAs are popular because of their ability to mimic the evolutionary ideas of natural selection and genetics according to the "survival

of the fittest" theory. This makes the method very helpful as adaptive search algorithm for optimisation problems, even for non-linear optimisation in complex engineering problems [80, 81, 82]. These algorithms are robust and have a good performance in complex optimisation scenarios. A formal description of the method can be found in [77]. Many examples on using the NGSA-II algorithm can be found in the CFD optimisation literature. For instance, in [83] an axial compressor is optimised to obtain the maximum efficiency at the design and choked working points. In [84], a wavy fin-and-elliptical tube heat exchanger has been optimised in order to maximise heat transfer and minimise pressure drop. In [85], an Autonomous Underwater Vehicle (AUV) is analysed. Their study consisted of integrating a Computer Aided Design (CAD) tool with CFD to obtain an optimal vehicle geometry in terms of a minimum drag and maximum nominal wake fraction and total volume. Radial Basis Functions (RBF) interpolation is also a popular method in the literature to define metamodels suitable for optimisation. In [86] and [87] the RBF metamodelling approach is described. An aeronautical advanced application successfully exploiting the RBF approach is [88], where a numerical shape optimisation framework is followed using mesh morphing with RBF to optimise the performance of a glider suffering important wing-fuselage joint separation. Another example in the use of RBF in optimisation is [89], where stochastic RBF are used to create multi-fidelity metamodels in an adaptive grid environment for optimisation purposes.

To conclude this section, the structure of the paper is outlined. First, in Section 2 a description on the geometry and computational considerations are given. In Section 3 the MLADO framework is introduced and explained. In Section 4 a vortex shedding predictor is developed. This predictor allows to know beforehand if a configuration will exhibit vortex shedding or not in the computations. Later, surrogates are built in Section 5. In the event of needing extra data points for the surrogates, the vortex shedding predictor can be used for evaluation and to avoid to compute vortex-free simulations (the computational URANS require computational efforts), which accelerates the design life-cycle. Then, the multi-objective optimisation NGSA-II is deployed on the surrogates. Finally, Section 6 is devoted to the final conclusions from this investigation.

2. Computational Geometry and Numerical Approach

The geometry of the microdevice under study is shown in Fig. 1: a two-dimensional (2D) rectangular pillar (which is h and l meters long in the x and y-direction, respectively) is located

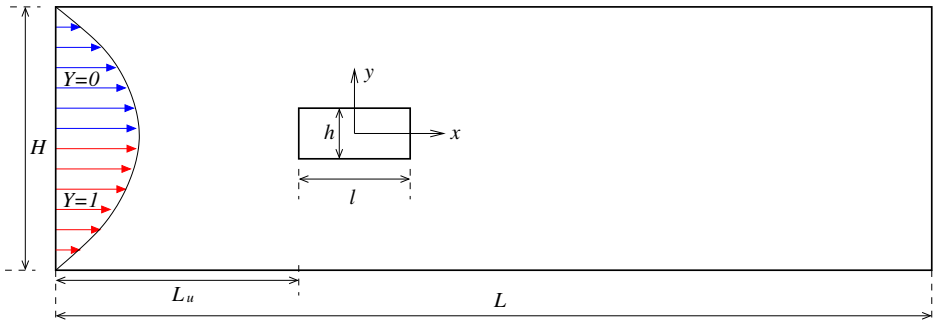


Figure 1: Sketch of the geometry.

at the centreline of a straight H-meter-wide microchannel. This means a pillar-to-channel width ratio, i.e. a blockage ratio BR, of $\frac{h}{H}$ and a length-to-width pillar ratio, i.e. the aspect ratio AR, of $\frac{l}{h}$; the pillar is centered at the origin of coordinates (x, y), being the channel inlet at a distance L_u meters upstream the front face of the pillar, while the channel is L meters length.

At the channel inlet, a fully developed laminar flow of two identical fluids with different values of a scalar magnitude Y for mixing purposes (for instance, temperature or mass fraction), scalar diffusivity coefficient D_c (for the temperature D_c would be thermal diffusivity α, while for the mass fraction this parameter would be the mass diffusivity D), and known density and viscosity, ρ and μ, respectively ($\nu = \mu/\rho$ is the kinematic viscosity), enters the channel inlet with a parabolic velocity profile of mean velocity value of U. A null value of the scalar magnitude Y in the fluid enters the channel through the upper half of the inlet section, i.e., $Y = 0$ at $x = -(L_u + l/2)$, $0 < y \leq H/2$, while the fluid enters through the lower half with unity value of the scalar magnitude, i.e., $Y = 1$ at $x = -(L_u + l/2)$, $-H/2 \leq y \leq 0$. With this kind of boundary conditions for Y, it is assumed that it has been properly re-scaled to range from 0 to 1. Downstream, the quality of the mixing of the fluids will depend on the Reynolds number, the scalar diffusivity coefficient, and the aspect and blockage ratio of the geometry. Therefore, under the assumptions of transient motion and incompressible fluid, the velocity $\mathbf{v} = (u,v)$ and pressure p fields of the 2D flow on the above described geometry are governed by the continuity and momentum equations which, in a dimensionless notation, can be written, respectively, as

$$\nabla \cdot \mathbf{v} = 0, \tag{5}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\nabla p + \frac{1}{Re}\nabla^2 \mathbf{v}, \tag{6}$$

while the mixing is governed by the scalar convection-diffusion equation, written as

$$\frac{\partial Y}{\partial t} + (\mathbf{v} \cdot \nabla)Y = \frac{1}{Pe}\nabla^2 Y. \tag{7}$$

The characteristic length, velocity, pressure and time used in the study are H, U, ρU^2 and H/U, respectively. Re in Eq. (6) is the Reynolds number defined, as usual in channel flow problems, as $Re = \rho UH/\mu$, while Pe in Eq. (7) is a Peclet-like number defined as $Pe = UH/D_c$, which is a ratio between two characteristic times: the diffusive one, H^2/D_c; and the convective one, H/U. High values of Pe indicate that the time needed for diffusion-based mixing will be very long and the mixing conditions are very unfavourable. The Peclet number can be also written as $Pe = Re\,\Gamma_c$, where Γ_c, defined as $\Gamma_c = \nu/D_c$, is a ratio between two molecular transport mechanisms (Γ_c would be the Prandtl number Pr or the Schmidt number Sc in case the scalar Y were the temperature or the mass fraction, respectively). One can see that, on the one hand, Pe is both fluid and geometry dependent, while on the other, Γ_c is only fluid dependent, being a ratio between two diffusion mechanisms at thermal/molecular level: the ones of momentum and temperature/mass. For a given and known fluid with constant physical properties, if the dimensions of the geometry change, Pe and Re can change, despite the use of the same fluid, while Γ_c will remain constant. In fact, a very unfavourable and high value of $\Gamma_c = 10^4$ for mixing has been kept constant throughout the study.

Regarding the geometry, the non-dimensional measures are: 1 unit wide ($H = 1$) crosswise, 5 units lengthwise ($L = 5$) and the pillar is located at 1 units from the channel entrance ($L_u = 1$). Additionally, the blockage ratio of the channel BR and the rectangular cylinder aspect ratio AR are, respectively, $BR = h/H$ and $AR = l/h$. Thus, for given values of Re, AR and BR, the vortex shedding from the pillar may, or may not, appear downstream. Consequently, the flow would be either unsteady or steady. Therefore, for each pair (AR, BR), there is a critical Reynolds number Re_{cr} above which the flow is unsteady and oscillatory: $Re_{cr} = Re_{cr}(AR, BR)$. For a given $Re < Re_{cr}$, the flow would be steady, and also the lift and drag forces of the fluid on the pillar. These forces, in a dimensionless notation, can be written as

$$Cl = \frac{F_y}{\frac{1}{2}\rho U^2 h},\ Cd = \frac{F_x}{\frac{1}{2}\rho U^2 h}, \tag{8}$$

where Cl and Cd are the lift and drag coefficients, respectively, and where h, as usual in studies of flows around pillars, has been used as characteristic length in order to have comparable coefficients. The oscillatory behaviour of Cl because of the vortex shedding from the pillar when $Re \geq Re_{cr}$, allows us to identify both the critical value Re_{cr} and the Cl frequency of oscillation f. This frequency is represented in dimensionless notation by means of the Strouhal number: $St = fh/U$. Besides this, if an arbitrary dimensionless magnitude g oscillates with time, its time-averaged value

for an oscillation period St^{-1} can be calculated as

$$\langle g \rangle = \frac{1}{St^{-1}} \int_{t_0}^{t_0+St^{-1}} g(t')\, dt', \tag{9}$$

being t_0 any reference for the periodic flow.

Pressure losses are an important aspect in flows through a channel. This defines the pumping power needed for the fluid to flow along the channel. Concretely, this is the power needed to maintain a given flow rate q in the channel for which the pressure and viscous losses throughout the whole channel must be overcome (which can be evaluated as the inlet-outlet pressure difference times the mass flow rate). Thence, the dimensionless pumping power, denoted by Π, can be evaluated as

$$\Pi = \Delta p\, q, \tag{10}$$

where Δp is the pressure drop between the channel inlet and outlet. Due to the chosen characteristic magnitudes, $q = 1$, and Π has been made dimensionless by using $\rho U^3 h$. Taking all the aforesaid specifications into account, finally the pumping power is defined by $\Pi = \Delta p$.

The most important parameter in the design of the device is the assessment of the mixing quality. This feature must be evaluated at the outlet section, being defined as the mixing efficiency η, in %, as done in e.g. [90, 13]:

$$\eta = \left(1 - \frac{\langle \sigma \rangle}{\sigma_{max}}\right) \times 100, \tag{11}$$

where σ is the standard deviation of the corresponding scalar magnitude at the channel exit and σ_{max} is the maximum standard deviation at the inlet: 0.5 in the present case. Given that $0 \leq Y \leq 1$ and $0 \leq \sigma \leq 0.5$; $\sigma = 0$ means uniform distribution of Y at the outlet, achieving a full mixing ($\eta = 100\%$), whereas $\sigma = 0.5$ would mean no mixing at all ($\eta = 0\%$).

The conducted numerical simulations to simulate the performance of the device as micromixer have been carried out by using the commercial software ANSYS-Fluent. To that end, the set of Eqs. (5)-(7) have been numerically solved by means of the pressure-based formulation, with second order spatial discretization methods. On the other hand, the SIMPLE (Semi-Implicit Method for Pressure-Linked Equations) algorithm [91] was used to handle the velocity-pressure coupling. Additionally, the numerical simulations were run with the following boundary conditions. The inlet was a velocity-inlet boundary condition with a imposed parabolic velocity profile and known temperature/mass fraction. For all walls (pillar and channel), a solid no-slip wall boundary

condition was imposed with null heat/mass flux through them. At the outlet, atmospheric constant pressure was imposed. This allows for reverse and non-developed flow, so it is considered more suitable than the 'outflow' boundary condition available in the software for the problem under study. Additionally, regarding the temporal integration of Eqs. (5)-(7), an implicit second order method has been used. The time step Δt was chosen in order to have the maximum Courant number below 1.

Numerical investigations must ideally conduct both a grid convergence and validation study. These provide an overview on the optimum mesh grid and allow to quantify the discretisation uncertainty (Grid Convergence Index, GCI). Both studies are summarized in Table 1 and Fig. 2, respectively. Different characteristic magnitudes such as St, $\langle Cd \rangle$, lift coefficient peak-to-peak value Cl_{pp} and its root mean square Cl_{rms}, have been obtained for three different meshes with uniform cell size ($ds_i = \{0.0125, 0.025, 0.05\}h$, i = 1,2,3). This has been tested for four values of the blockage ratio ($BR = \{1/8, 1/5, 1/4, 1/3\}$), a fixed Reynolds number ($Re = 100$) and a fixed aspect ratio $AR = 1$ (squared pillar). By means of the computation of the aforementioned output mechanical magnitudes, their GCI and their Richardson Extrapolation can be estimated, whose value corresponds to $ds \to 0$ (see Fig. 2). The $GCI_{i+1,i}$ allows us to quantify the discretisation uncertainty of each magnitude for the two finest grids, i.e., for i = 1,2. Taking into account that the grid refinement ratio is 2, and by using a safety factor F_s of 1.25 as recommended in [92], the uncertainty of the finest grid ranged between 0.5% and 2.3%, whereas for the medium grid, it ranged between 1.8% and 6.5%, as can be seen in Table 1, where all GCI values are summarized. As illustrated in the table, the medium mesh with ds_2 gives reasonable uncertainty discretisation values and could be chosen as the optimal one. However, in order to further reduce the discretisation uncertainty, a ds smaller than ds_2 has been finally used: $ds = 0.02h$. This means a 20% of increase in the number of grid nodes along the pillar of width h. In Fig. 3 it is depicted a detailed view of the optimal uniform mesh around the pillar. Regarding the goodness of the optimal mesh from a convective-diffusive mixing point of view, it is worth to mention that special attention must be paid on properly solve the smaller length scales in the scalar magnitude field Y, specially in liquids with Γ_c normally very large. In this case, the smallest size of the mass fraction field length scale can be estimated by using the Batchelor length scale, which in a dimensional way is denoted as λ_B and, by following [93, 94], can be written as

$$\lambda_B = \frac{\lambda_{vel}}{\sqrt{\Gamma_c}}, \tag{12}$$

with λ_{vel} the length scale of the smallest velocity structures. Assuming that for the viscous flow of this study $\lambda_{vel} \sim H$, with $\Gamma_c = 10^4$, $\lambda_B \sim 0.01H = 0.01h\,BR^{-1}$, which means that the most unfavourable situation would be the one with the highest BR. Since in our investigation the maximum value of BR would be 0.5, this means that $\lambda_B \sim 0.02h$, which is the order of magnitude of the final and optimal chosen mesh.

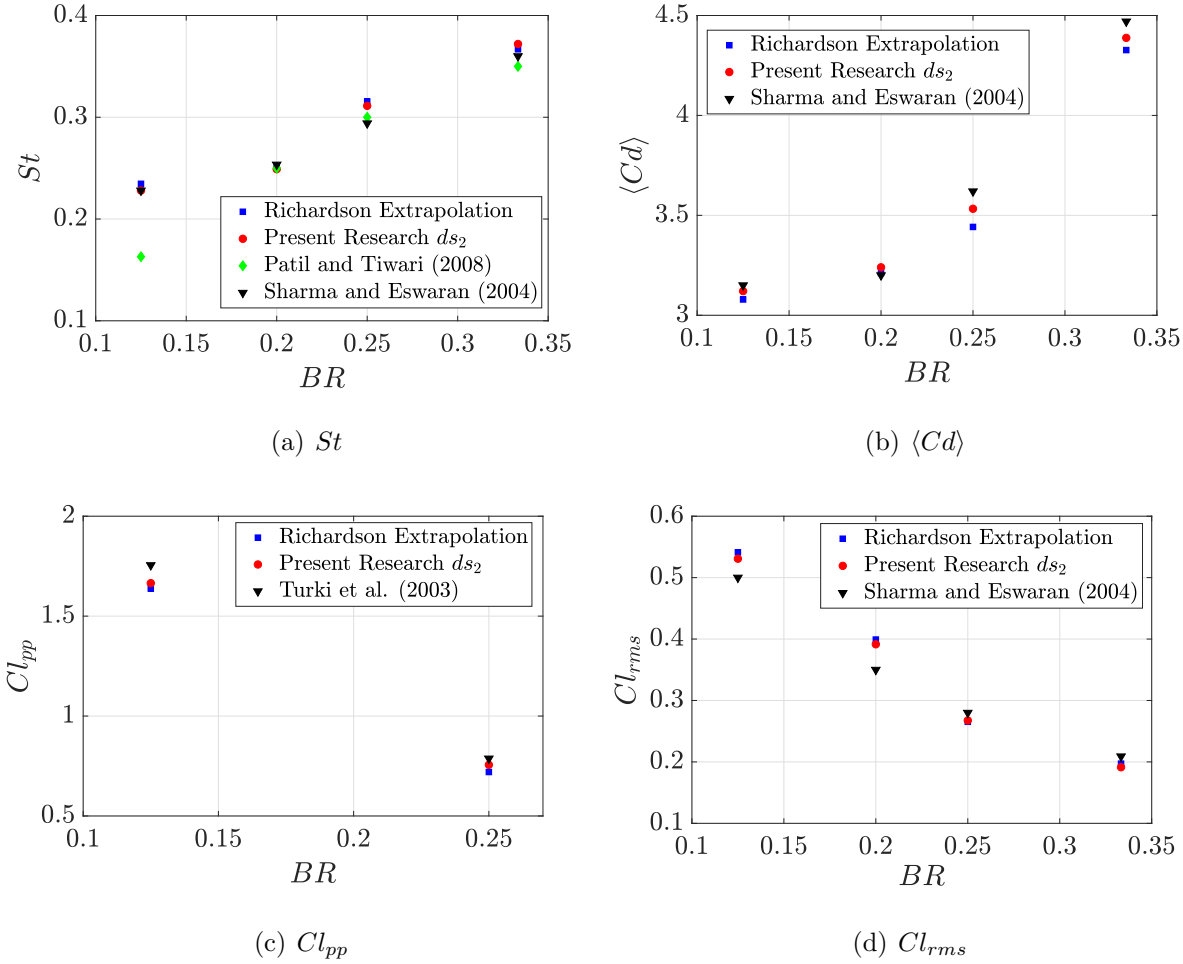
The computations in the present work have been also compared to other works in the literature for validation. Fig. 2 shows a comparison between the results obtained by the numerical model in the present research and those given by different previous authors. The good agreement between the present simulation and the previous results, in addition to the reduced GCI previously analysed, evince that the numerical considerations in the followed methodology are appropriate for the problem under study.

				St	$\langle Cd \rangle$	Cl_{pp}	Cl_{rms}
Re	AR	BR	Grid: i	$GCI_{i+1,i}$	$GCI_{i+1,i}$	$GCI_{i+1,i}$	$GCI_{i+1,i}$
100	1	1/8	Fine: 1	1.0%	2.1%	0.8%	0.6%
			Medium: 2	2.9%	4.6%	4.9%	5.9%
100	1	1/5	Fine: 1	0.5%	0.6%	–	0.9%
			Medium: 2	2.6%	1.8%	–	5.9%
100	1	1/4	Fine: 1	1.9%	0.5%	1.4%	0.2%
			Medium: 2	4.5%	4.1%	6.5%	2.5%
100	1	1/3	Fine: 1	1.0%	0.6%	–	2.3%
			Medium: 2	4.1%	4.2%	–	4.7%

Table 1: Grid convergence study results with the GCI values for the indicated grids.

3. Machine Learning-Aided Design Optimisation (MLADO) Framework

For the sake of achieving an efficient and optimal mixing device, an optimisation study is conducted in this paper. The optimisation process, named Machine Learning-Aided Design Optimisation (MLADO) in this work, is developed according to the diagram presented in Fig. 4. The MLADO framework consists of using CFD simulations, a predictive model, surrogate modelling and a optimisation algorithm. Upon the initial data considered to build the surrogates, one has to consider whether these are enough data points and surrogates have enough quality. One should



(a) St (b) $\langle Cd \rangle$

(c) Cl_{pp} (d) Cl_{rms}

Figure 2: Validation and comparison with other authors: Turki et al. (2003) [95], Patil and Tiwari (2008) [96] and Sharma and Eswaram (2004) [97].

in principle decide where to place new data points (e.g. new CFD simulations) based on the Mean Square Error (MSE), R^2 or any other statistical measure given by the considered surrogate method. However, since the optimal configuration is vortex shedding oriented to achieve a proper mixing, vortex-free cases are useless. Thus, by including a predictor, the process to achieve useful surrogates can be data-driven and save important computational resources. If the predictor output is that the new potential configuration yields vortex shedding, then is strongly considered as new point for surrogate enhancement and is simulated.

The first stage of the process is to acquire sufficient data to construct an initial surrogate. Any Design of Experiment (DoE) may be used here, depending on the nature of the problem, computational budget, etc. However, to explore initially at least three points per direction (variable in the design space) is recommended, as the response may be non-linear. Then, surrogates are

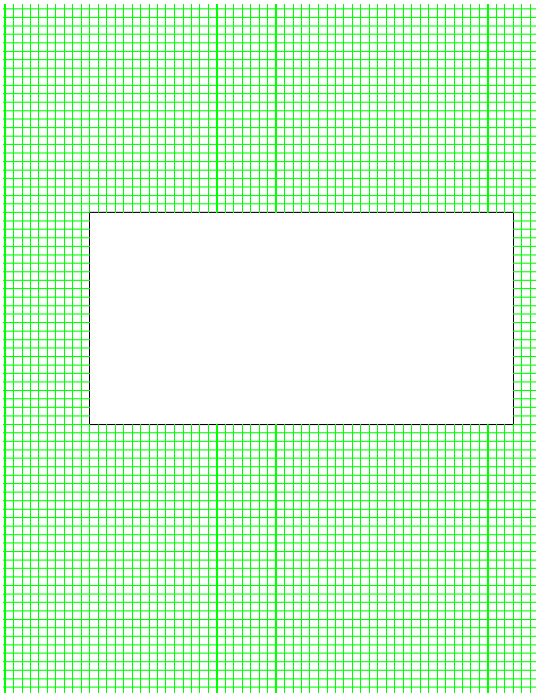


Figure 3: Detail of the optimal mesh around a pillar with $AR = 2$.

constructed. As the initial surrogates are likely to be coarse, some regions must be refined. For such task, the Mean Square Error (MSE) tells whether or not is necessary to refine. If refinement is necessary, then the vortex shedding predictor (classification model) filters if the configuration is a vortex shedder ($VS = 1$). If so, the point can be simulated and added to the surrogate data set composed of N_s samples.

Must be outlined that N_p (number of samples of the vortex shedding predictor) and N_s (number of samples for the surrogate model) does not need to coincide. This is so because the smoothness of the prediction of vortex shedding may be greater than the smoothness in e.g. the Kriging surrogates of the mixing efficiency. This will depend on the quantities of interest and the problem under consideration, as for instance the pressure drop is very smooth. Additionally, the vortex shedding predictor does not require full simulations to be trained. One just need to differentiate between $VS = 0$ and $VS = 1$ cases, and this can be done for instance from low fidelity simulations or external data.

The vortex shedding predictor is a pre-trained algorithm. This can be trained with data of any nature (steady simulations of unsteady problems, coarser meshes, low fidelity simulations, analytical solutions, databases, other literature, etc.). As vortex shedding is a problem where configurations with and without oscillatory behaviour can be easily clustered, then few simulations

have been observed to be enough to train well the predictor. Once in the MLADO loop, to update the predictive model for classification is a recommendable option, but one must be careful with this. The potential learning on imbalanced data is a problem to point out [98]. If the vortex shedding predictor is updated automatically when new samples not considered in the training of the classifier are fully simulated, then imbalance in the training may appear. Since new training data would be simulated based on the prediction of $VS = 1$, one would increase only the number of cases of $VS = 1$, thus increasing the imbalance in the training. Only the false positives from the predictor would improve balance in the data set, since the simulation would then include data with $VS = 0$. For this reason, we do not recommend to update automatically the classificator and rather use a well-trained predictor since the beginning if possible, which should be updated only according to expert criteria.

After execution of the loop, one could let the framework work automatically until certain level of refinement is achieved. The convergence criteria for this is open to the researcher, and in the present application it has been controlled manually. When a good enough surrogate model is achieved, the optimisation algorithm is used to evaluate them and find a Pareto front for the multi-objective problem. This MLADO framework could be extendable to any engineering design process subject to classification. The development of the predictive model is explained in Section 4 and the application of the framework to a mechanical micromixer design is developed in Section 5, which is the main purpose of the present investigation. It is of interest to remark that, in the present work, for each geometric variation, a new mesh must be generated. An interesting feature to add to the MLADO framework can be to include a mesh morphing option. In this scenario, a reduced order model of the computation could be defined and the complete quickly solution inspected in real-time. Having the flow represented on the same topology of the mesh could also help the definition of methods addressing flow feature detection. An interesting application of such mesh morphing technique for real-time inspection can be found for instance in [99]. In this work, similarly to [88], the parametric shape analysis is made by means of RBF-based mesh morphing approach. The strong difference in the methodology in [99] is that an inspection tool is developed within a reduced order model framework to visualise in an interactive manner the flow field in real-time.

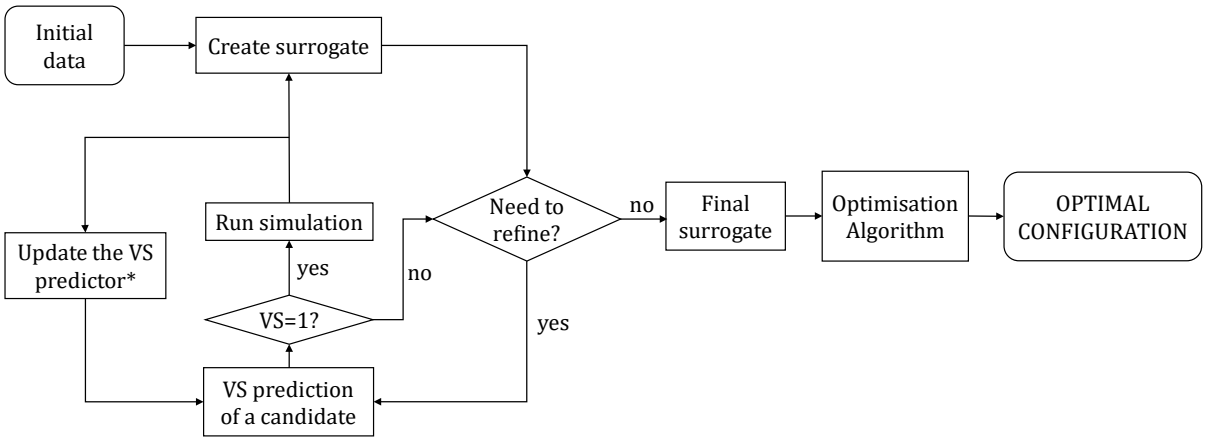


Figure 4: MLADO framework.

4. Classification Model for Vortex Shedding Prediction

As explained in Section 1, it is of interest to be able to predict the presence of vortex shedding in many applications because of the desirable/undesirable effects on a system performance [30, 22, 36, 37, 38, 29]. In the present investigation, vortex shedding is desired to achieve an efficient mixing of the corresponding scalar magnitude Y.

For the objective of classifying which configuration may ($VS = 1$) or may not ($VS = 0$) lead to vortex shedding, two predictive models have been developed and tested on a data set of 80 simulations of the pillar confined in the channel. The data set of 80 simulations is a very complete one used in this manuscript, but it is a bit expensive to create. However, as the traditional optimisation will be also deployed on 80 simulations for comparison with MLADO, the 80 full simulations are already available and it is decided to examine and test the construction of predictive models for this full data set available as reference. These consist of cases with their configuration parameters ranging in $120 \leq Re \leq 200$, $0.2 \leq BR \leq 0.5$ and $0.125 \leq AR \leq 1$, while, as previously said, $\Gamma_c = 10^4$. The reason of having such high value is to have a very unfavourable mixing case, where the diffusion of the scalar magnitude between the fluids is very poor and, unless additional mechanisms were used, the mixing efficiency would be very poor too. This helps to identify how the vortex shedding and the downstream wake oscillations enhances the fluids mixing. Among the 80 simulated configurations, 27 were non-vortex shedders (33.75%), so the data set was not worrisome imbalanced [98]. The full sample has been split into two parts (in a stratified way, ensuring similar % of $VS = 0$ and $VS = 1$ cases in both samples): training data (80% of the

sample, 64 samples) and test data (20% of the sample, 16 samples). Must be outlined that the determination of whether each case corresponds to $VS = 1$ or $VS = 0$ is based on the lift coefficient oscillation. Although this could be detected automatically, this assignment has been done at a post-processing stage for the sake of simplicity, as the analysis is not carried out in real-time.

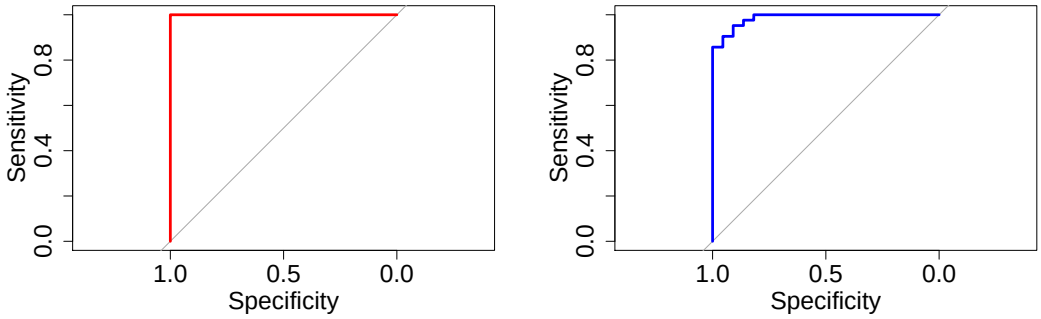
The first model tested was a Logistic Regression (LR). This model is simple and usually have a good performance. The predictive model has been trained with the training data, obtaining the model shown in Table 2. It can be seen that all the variables included in the model are significant (i.e. they are statistically relevant to model). If any variable had a p-value above 0.05 value, should be considered whether to discard such variable from the model.

Variable (ξ_i)	Coefficient (α_i)	Std Error	p-value
Intercept term	-26.80569	9.20788	0.00360
AR	-13.98463	4.92284	0.00450
BR	41.79316	14.84238	0.00487
Re	0.13692	0.04784	0.00421

Table 2: Logistic Regression model for vortex shedding prediction.

The LR model must be validated against the test data, to ensure independence in the construction and evaluation. This is done through the use of the analysis of the so-called Receive-Operation Curve (ROC) [100]. The area delimited by this plot is referred to as Area Under Curve (AUC) in the literature, and it is used as accuracy measure in classification models. The value $AUC = 1$ means that all cases are classified correctly. A value of $AUC = 0.5$ means that the model performs as a random decision, and then the model is not reliable. From the analysis on the test data, the ROC curve is shown in Fig. 5(a). The LR model performed with an $AUC = 1$, being able to accurately predict all vortex shedding configurations in the test data. Despite it is not recommendable, the performance was also tested on the same data used to train the model (training data), with a $AUC = 0.9859$ (as seen in Fig.5(b). An additional test to demonstrate split sample independence and potential overfitting was a 5-fold cross-validation. This consisted of splitting the data into 5 parts. Then, the model is trained with 4/5 parts of the data, and validated with the remaining 1/5. Then process is repeated through the 5 partitions, obtaining 5 different models evaluated on 5 different test data. The 5-fold has been split with stratification (i.e. trying to preserve a balanced $VS = 0$ and $VS = 1$ percentage of cases). Results are shown in Table 3, where it can be

seen consistency in terms of high accuracy and that the model fails to predict very few situations in three partitions.



(a) (b)

Figure 5: ROC curve of the Logistic Regression performance on (a) test data and (b) training data.

AUC1	AUC2	AUC3	AUC4	AUC5
1	1	0.98	0.945	0.969

Table 3: 5-fold cross-validation of the Logistic Regression model.

Although the logistic regression model developed is very accurate and the interpretability of the method is an advantage, more promising classification methods do exist. Examples of high accuracy and popular classificators are Random Forest (RF), Neural Networks in Deep Learning, or boosting methods such as Adapting Boosting (AdaBoost), Gradient Boosting Machine (GBM) or XGBoost. Amongst these, a Random Forest algorithm has been selected to generate a predictive modelling for the vortex shedding. This method is very popular in overall scientific literature and it is well-known that provides accurate predictions and robustness. The algorithm [101] did perform with very high accuracy as will be shown next. Thus, there is no need to attempt other algorithms of higher complexity in this work. However practitioners may consider them for other applications.

For theoretical definition, let define as Θ the vector of hyper-parameters Θ_k that represent the RF algorithm. Each tree J_k is grown by using each random vector Θ_k and the training data set D_t, what leads to $J_k(\mathbf{X}) = J(\mathbf{X}, \Theta_k)$. As said in [32], the definition of a RF classifier is "a classifier consisting of a collection of tree-structured classifiers $\{J(\mathbf{X}, \Theta_k),\ k = 1, ..., ntrees\}$, where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input X_i", where $ntrees$ is the number of trees that construct the RF predictor. In classification problems RF is thus an ensemble of several models and the predicted

class C^{RF} upon the *ntrees* decision trees is selected as the majority vote of a class $c \in \mathbf{Y}$, that is [102]:

$$C^{RF}_{D_t,\Theta_1,\Theta_2,\ldots,\Theta_{ntrees}} = arg\,max \sum_{m=1}^{ntrees} \left(J(\mathbf{X},\, \Theta_m) = c\right). \tag{13}$$

Since Random Forest is an ensemble of several decision trees, the algorithm requires to define the number of trees *ntrees*. There is no theoretical limitation on this, but with a large *ntree* value computational resources would be unnecessarily increased. A hint of the optimal number of trees can be guessed from the Out-Of-Bag error (OOB). This parameter is an error measure that results from the evaluation by the trees of training data samples which were not included in the bootstrapped sample of the trees [103], so this is done for each iteration and consequent tree. Above a certain number of trees, the OOB is unperturbed. Some authors suggest that a good number of trees is usually around few hundred [104, 103]. Another relevant tuning parameter in the Random Forest is the number of variables to be randomly sampled at each split, *mtry*. This is particularly relevant for datasets with large number of features (variables). Due to in the present investigation there are only 3 variables, *mtry* has been fixed to 3. The OOB error with respect to the number of trees is shown in Fig. 6 for $ntrees = 500$. The error is stabilised from 400 approximately. Hence, a number of 500 trees is kept, since the difference in the computational time was found to be almost unperturbed. Replacement sampling has been permitted to the bootstrapping algorithm of Random Forest, because the training data had a very limited size to train all the underlying decision trees. The predictor obtained by Random Forest has a fully accurate performance, with no false positives nor false negatives for the test data validation. These results are shown in the confusion matrices shown in Table 4. The confusion matrix shows the predictor output (predicted value) versus the actual value (reference). This corresponds to a $AUC = 1$, since all cases are correctly predicted. Must be noted that the dataset is not split into train and test data, since the RF algorithm is already splitting and testing tree models. Thus makes no sense to further split the data. Nevertheless, althought not very relevant, as done for the LR model a 5-fold cross-validation is developed for the RF predictor, whose results are presented in Table 5. During the cross-validation can be observed that the accuracy is not $AUC = 1$ for some splits, but still very high. The RF predictor performed similarly to the LR predictor, so one can conclude that both are recommended in this particular application, although in other problems their performance may differ. In our opinion, RF is preferred because one can find applications in the literature where outperforms LR. Nevertheless, despite their good performance, interpretability is a well-known

key problem for these type of predictors yet [105], commonly interpreted as black-box because of its "impenetrability" to know the interaction of variables [32]. For this reason, the RF predictive model is provided as additional data to this manuscript.

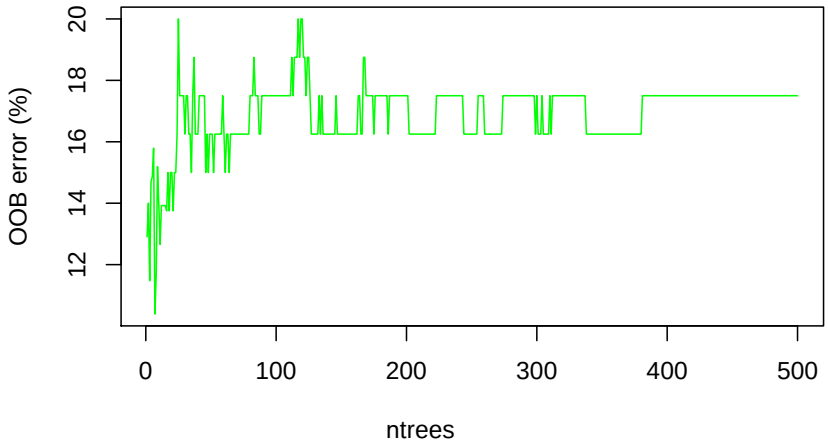


Figure 6: OOB error of Random Forest algorithm.

		Reference data	
		0	**1**
Predicted value	**0**	27	0
	1	0	53

Table 4: Confusion matrix for the Random Forest predictor.

AUC1	AUC2	AUC3	AUC4	AUC5
0.96	0.98	1	0.96	1

Table 5: 5-fold cross-validation of the Random Forest predictor.

Both the LR and RF models aforementioned are based on a total of 80 simulations, but to select this number of simulations as reference, a study on the number of simulations to train the models was developed. Since in the present work we have chosen the RF model, the analysis (similar to a convergence analysis) is shown only for this model. For the study, an initial "complete enough" set of data is required to train the predictor, which we have set in a Design of Experiment (DoE) of $N_p = 36$ points of almost equally spaced points as the combinatorial of $Re = \{120, 160, 200\}$, $BR = \{0.2,\ 0.3,\ 0.4, 0.5\}$ and $AR = \{0.125, 0.5, 1\}$. The RF is trained with this set of data and

used to predict the next level of refinement. In this study, four levels of refinement have been developed onto the $N_p = 36$ samples: a) Addition of $Re = 140$ to the combinatorial (now $N_p = 36$ to classify 48 samples), b) Addition of $AR = 0.25$ to the combinatorial of the previous refinement (now $N_p = 48$ to classify 64 samples), c) Addition of $Re = 180$ to the combinatorial of the previous refinement (now $N_p = 64$ to classify 80 samples), and finally d) Addition of $AR = 0.8$ to the combinatorial of the previous refinement (now $N_p = 80$ to classify 100 samples). The evolution of the accuracy of the prediction is shown in Table 6, where can be observed that when the predictor is trained with $N_p = 80$ simulations, the reference data of 100 simulations is predicted without misclassifications.

By observation of the evolution of the predictions, one can decide to stop at a certain number of simulations N_p, chosen as definitive number of simulations to train the predictor. From that value on, N_p can be increased (that is to say, the RF predictor can be updated within MLADO) depending on the Mean Square Error (MSE) or any other metric of the Kriging surrogates if desired. Must be recalled that N_p (number of samples of the RF predictor) and N_s (number of samples for the surrogate model) does not need to coincide, as will be demonstrated in Section 5. For instance, the predictor can be trained with basic simulations (steady state, coarser mesh, lower fidelity simulations, external data, etc.) as only to determine VS is necessary. On the other hand, the Kriging surrogates (or any surrogate/interpolation method used) require the full accurate unsteady simulation, as the accurate calculation of η and Π is a must. Fortunately, this MLADO method helps in the decision to simulate or not these costly simulations.

From the results given in Table 6 one can extract more useful conclusions. The nested univariate procedure (nested means that previous samples are kept) followed in the analysis of the convergence is just an example on how to proceed, and other options can be considered, such as random nested sampling [106] or sparse grids [107]. However, the problem under analysis did not require a more elaborated convergence analysis, since even with few samples, the RF algorithm provided an outstanding performance. It has been even tested that the RF trained with $N_p = 48$ predicted the 100 samples with just 5 misclassifications (two false positives and three false negatives). This outstanding performance is so because the activation of vortex shedding does not expect sudden changes as it is a smooth response problem. However, in other applications the situation may not be that favourable. Our analysis suggests that for other applications in vortex shedding prediction

		Reference data	
		0	**1**
Predicted value	**0**	21	2
	1	0	25

a)

		Reference data	
		0	**1**
Predicted value	**0**	22	0
	1	2	40

b)

		Reference data	
		0	**1**
Predicted value	**0**	27	2
	1	0	51

c)

		Reference data	
		0	**1**
Predicted value	**0**	41	0
	1	0	59

d)

Table 6: Evolution of the accuracy of the RF algorithm in the prediction of the next level of refinement as test samples. a) Confusion matrix in the classification of 48 samples (RF trained with $N_p = 36$, named RF1). b) Confusion matrix in the classification of 64 samples (RF trained with $N_p = 48$, named RF2). c) Confusion matrix in the classification of 80 samples (RF trained with $N_p = 64$, named RF3). d) Confusion matrix in the classification of 100 samples (RF trained with $N_p = 80$, named RF4).

one can predict the performance with a relatively low number of training data samples. It is relevant to outline that the misclassifications observed in Table 6 actually correspond to values in the separation limit of $VS = 0$ and $VS = 1$. For this application, this is not a concern. First, because the line that separates and groups $VS = 0$ and $VS = 1$ is vague (should I consider a very weak oscillation as $VS = 0$ or $VS = 1$?). In order to be conservative, in this work even a weak oscillation test case is marked with $VS = 1$. And second, because points with high mixing efficiency should be far from that region (stronger vortex shedding), so the misclassification in the vicinity of the separation limit between $VS = 0$ and $VS = 1$ is not a strong concern. The misclassifications are shown in Table 7 and in Figure 7 can be observed that these points are in the vicinity of the separation limit.

RF predictor, test samples	AR	BR	Re	VS (actual value)
RF1, 48 samples	0.125	0.2	140	1
	1	0.5	140	1
RF2, 64 samples	0.25	0.3	120	0
	0.25	0.2	140	0
RF3, 80 samples	0.5	0.3	180	1
	1	0.4	180	1
RF2, 100 samples	0.25	0.3	120	0
	0.25	0.2	140	0
	0.5	0.3	180	1
	1	0.4	180	1
	0.8	0.4	180	1

Table 7: Misclassified samples from Table 6 and prediction of RF2 of the 100 samples.

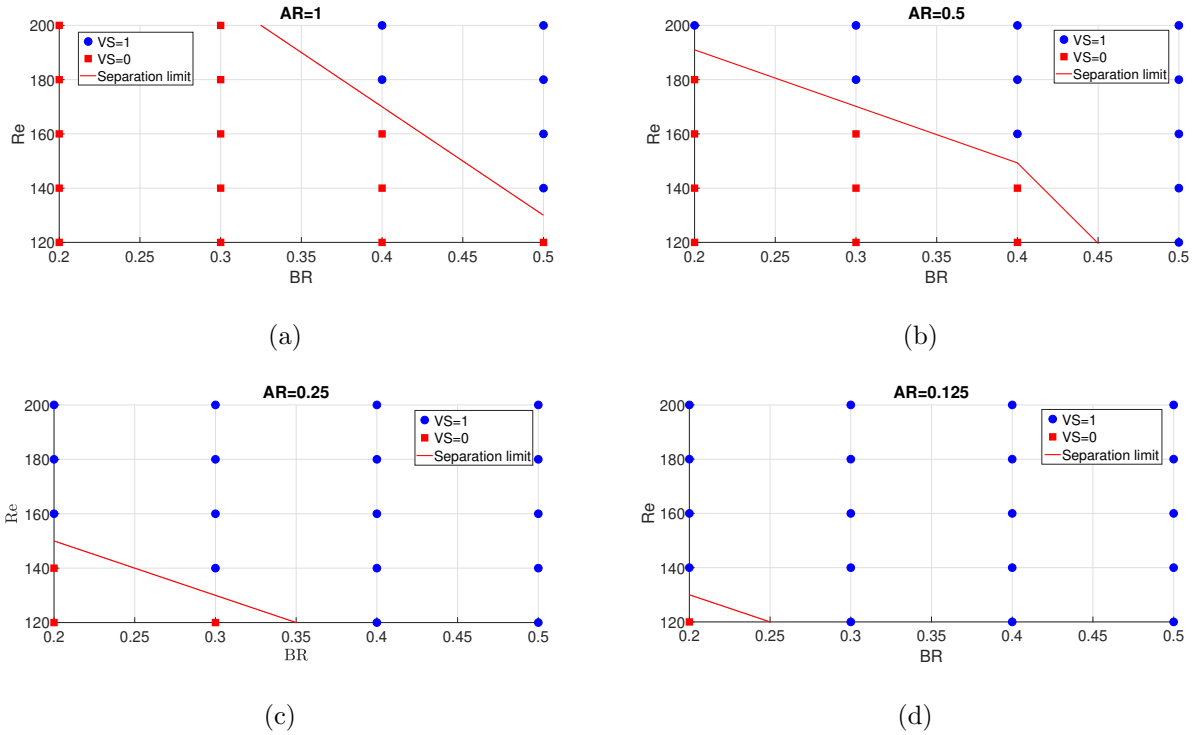


Figure 7: Representation of simulated cases (actual values) with and without vortex shedding and the separation limit for a) $AR = 1$, b) $AR = 0.5$, c) $AR = 0.25$, and d) $AR = 0.125$.

5. Application of a Machine Learning-Aided Design Optimisation Framework to the Design of a Mechanical Micromixer

The MLADO process is applied to the design of an optimal mechanical micromixer in this section. Starting the process as explained in Section 3, the initial step is to construct the Kriging surrogates upon data. As training data we have considered initially both cases with and without vortex shedding, because it helps to understand visually the performance. The change in the surrogates is negligible in the important areas (where mixing is important), as will be explained later. The visualisation of both cases in the same figure using different markers allows to understand visually regions differentiated between configurations with and without vortex shedding. However, the Mean Square Error (MSE) and the Random Forest prediction determine the real need of refining the data point distribution.

In this work (and for the MLADO method in general), the number of samples used in the surrogate model and the vortex shedding predictor does not need to coincide. This is strategic and one of the benefits of the method: the classification algorithm (RF) can be trained upon other data (steady simulations, coarse mesh, empirical correlations, low fidelity simulations, laminar, external data, etc.). From that data, one can classify and decide what to simulate (full simulation) next to create the surrogate models for optimisation. For instance, in the present investigation, the aforementioned set of simulations to train the RF predictor consisted of a DoE of equidistant points run in FLUENT as steady state. This was enough to observe when the oscillation in the lift forces do appear and then flag each case as $VS = 0$ or $VS = 1$. The RF predictor trained with 80 simulations has been preferred in our optimisation of the micromixer, as such data is available from the traditional optimisation study, so there is no point on using lower quality predictors. However, as explained in Section 4, a predictor with less than 80 simulations is still accurate and useful in predicting the oscillatory behaviour. The MLADO framework can be very useful when the optimisation procedure starts with a small data set or when a good amount of data is available to construct surrogates but additional refinement is expensive. The first option, which begins with few samples, is where one can take greater advantage of the MLADO, since the problem is transformed into an efficient data-driven solution to guide the refinement process. To demonstrate the capability of the method in the design of the micromixer, in this section, first, an optimisation process without MLADO but using nearly equidistant DoE for Kriging surrogates is presented. This will be used as reference. Second, the MLADO will be applied from an early design stage of

$N_s = 36$ simulations. These results will be compared to the reference ones to quantify the saved resources.

5.1. Traditional Approach to the Optimisation of a Micromixer.

It is frequent to see in the literature the use of Kriging surrogates for CFD optimisation where the DoE consists of a decent number of costly simulations in terms of computational resources. Then, the decision of whether to simulate a new additional expensive sample relies only on the MSE.

An optimisation study has been developed from an almost equidistant DoE set of individuals, which consists of the combinatorial of $Re = \{120,\ 140\,160, 180, 200\}$, $BR = \{0.2,\ 0.3,\ 0.4, 0.5\}$ and $AR = \{0.125, 0.25, 0.5,\ 0.8, 1\}$. This corresponds to a total population of 100 simulations. However, due to 100 simulations is too large for our computational resources, some will be discarded. By experience we know that $AR = 0.8$ is not a useful candidate, so that only 80 simulations are selected in the DoE, by discarding the $AR = 0.8$ combinations.

In the present work, the DACE toolbox [108] is used to build the Kriging surrogates. These are built upon the input samples. For such interpolation, it is advisable to normalise input data, but the toolbox performs normalisation automatically. For the Gaussian Process (Kriging) method, the correlation function for two points at locations x_i and x_i', $C(\theta_i; x_i, x_i')$, must be specified. This function is important to build a good Kriging estimator. The function depends only on the distance between two points: the smaller the distance, the higher the correlation. This helps to deal with clusters of data. On the other hand, if the distance between the two data points is increased, the correlation drops to zero. In the DACE toolbox, several options are available as correlation function (although one can construct custom functions as well). The correlations tested in this work are shown in Table 8.

In Figs. 8 and 9 can be seen the surrogates, which are shown for Re and AR at constant values, respectively. Analogously, in Figs. 10 and 11, the Mean Square Errors (MSE) are also given. In these figures, red squares are the simulated data cases without vortex shedding, whilst the blue points are the cases with vortex shedding. These surrogate models correspond to a Universal Kriging model of 2nd order regression with generalised exponential correlation (GEXP). In Figs. 12 and 13 the surrogates for η by means of other correlations are also depicted, just for $AR = 1$ and just for $Re = 120$, respectively, since these are the most challenging surrogates. As it can be seen in Fig. 12, the wavy shape of the function can be corrected by a linear correlation. However, the

Type	Function $C(\theta_i; x_i, x_i')$
Exponential	$e^{-\theta_i \lvert x_i - x_i' \rvert}$
General Exponential	$e^{-\theta_i \lvert x_i - x_i' \rvert^m}$; with $0 < m < 2$
Gaussian	$e^{-\theta_i \lvert x_i - x_i' \rvert^2}$
Linear	$\max(0, 1 - \theta_i \lvert x_i - x_i' \rvert)$
Spherical	$1 - 1.5\gamma_i + 0.5\gamma_i^3$; with $\gamma_i = \min(1, \theta_i \lvert x_i - x_i' \rvert)$
Spline	$1 - 15\gamma_i^2 + 30\gamma_i^3$; if $0 \leq \gamma_i \leq 0.2$; $1.25(1 - \gamma_i)^3$; if $0.2 < \gamma_i < 1$; 0; if $\gamma_i \geq 1$; with $\gamma_i = \theta_i \lvert x_i - x_i' \rvert$

Table 8: Correlation functions tested for the surrogates, as defined in [108].

sudden slopes at certain points produce a surrogate with unrealistic sharp areas. The use of splines with Ordinary Kriging provides good surrogates if observed for $AR = 1$ in the same figure, but the reality is that has not better performance than the GEXP, as observed with Reynolds number fixed to $Re = 120$ in Fig. 13. We also found that MSE is considerably greater than the values shown in Figs. 10 and 11. Spherical correlations performed the best with Universal Kriging with 2nd order regression, but produced sharp surrogates similar to those by linear correlation. Gaussian correlation performed similar to the GEXP since this is actually a generalised exponential of 2nd order, but the waviness is more noticed than for GEXP, as well as exhibits larger MSE. Therefore, the most appropriate correlation was found to be the GEXP, even though some waviness is noticed in the vicinity of $\eta = 0$. But since the optimal mixing efficiency values should be far from $\eta = 0$, this is not relevant.

Similarly, in Figs. 14-17 the surrogates and MSE for the dimensionless pumping power $\langle \Pi \rangle$ are shown. Since these training data points have an underlying smooth response, the modelling was simplified to Ordinary Kriging surrogates with simple exponential correlation.

By means of the generated surrogates it is possible to find an optimal configuration for the geometry of the pillar in the channel. For this optimisation study, there are two objective functions: the mixing efficiency η and the pumping power $\langle \Pi \rangle$. The aim is to maximise η and to minimise $\langle \Pi \rangle$. When dealing with these objectives separately, constrained by the given ranges of AR, BR and Re in the present work, the best configurations are given in Table 9.

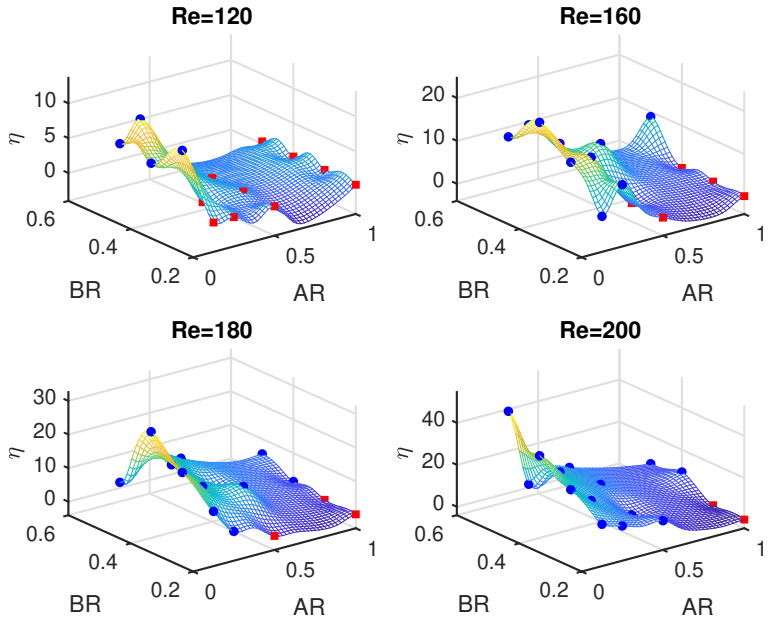


Figure 8: Surrogate models for η shown at different Re regimes. Red squares are configurations without vortex shedding and blue dots are configurations with vortex shedding.

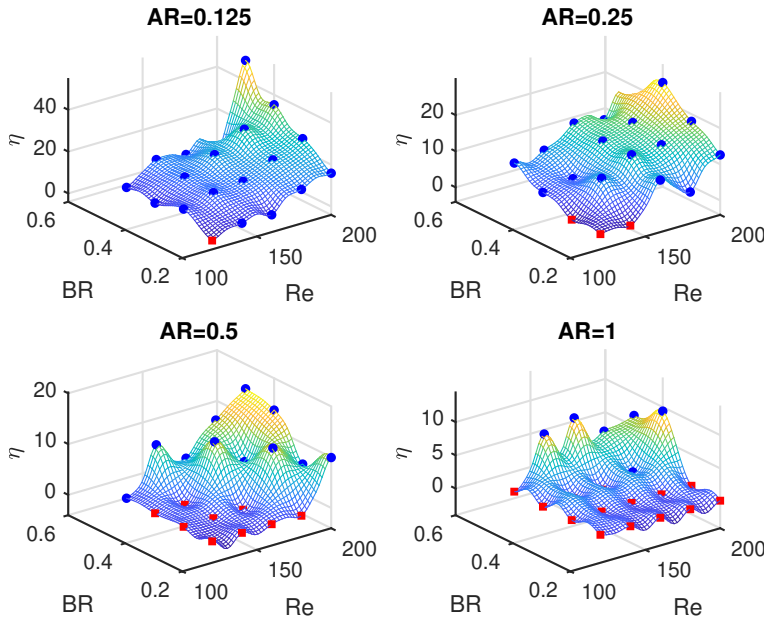


Figure 9: Surrogate models for η shown at different AR values. Red squares are configurations without vortex shedding and blue dots are configurations with vortex shedding.

These configurations are determined from the simulated CFD configurations. As seen in the previous table, when η is high, $\langle \Pi \rangle$ is also high and vice-versa. Thus, trade-off optimal values must be calculated by a multi-objective approach. For this purpose, the NSGA-II algorithm introduced

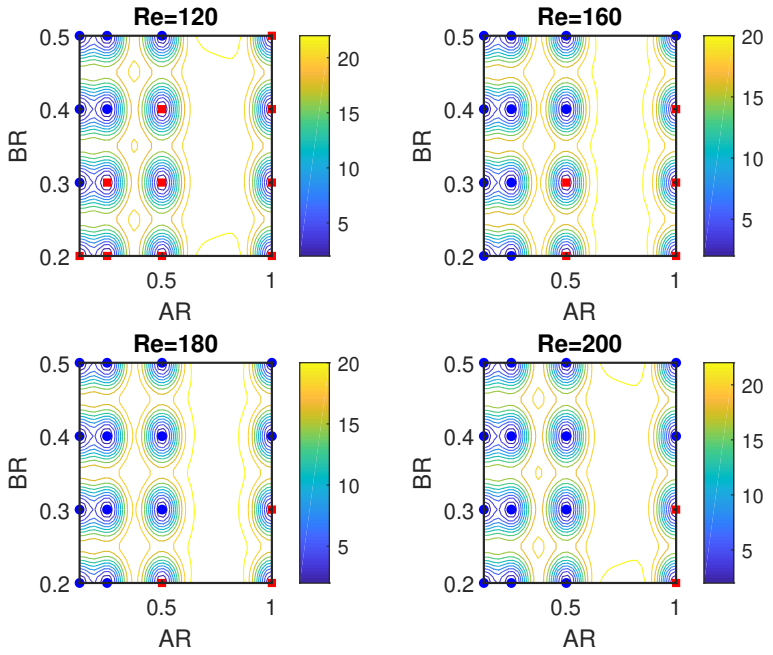


Figure 10: Mean Square Error (MSE) of surrogate models η shown at different Re regimes. Red squares are configurations without vortex shedding and blue dots are configurations with vortex shedding.

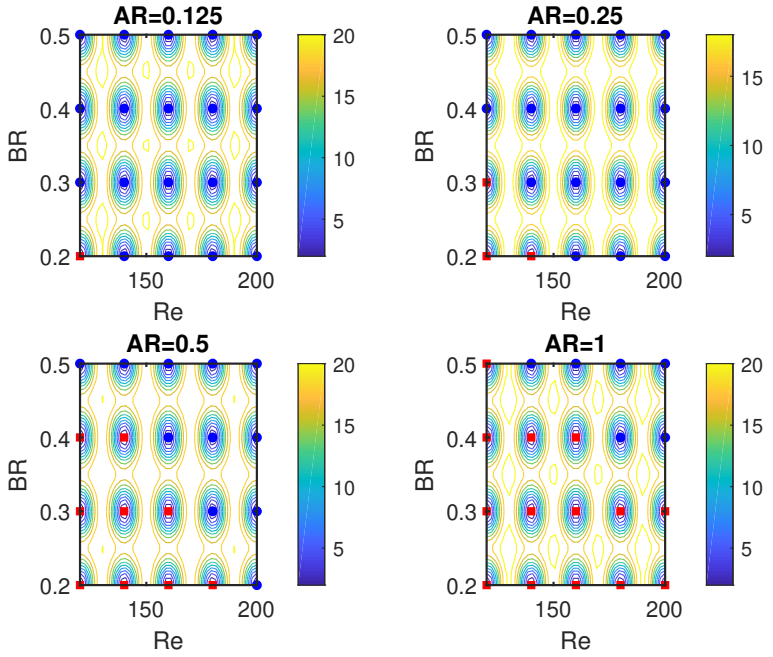


Figure 11: Mean Square Error (MSE) of surrogate models η shown at different AR values. Red squares are configurations without vortex shedding and blue dots are configurations with vortex shedding.

in Section 1 is used. The optimal candidate values of the Pareto front are given in Fig. 18. It can be seen that Re values remain clustered around $Re = 200$, so in practice that parameter could be

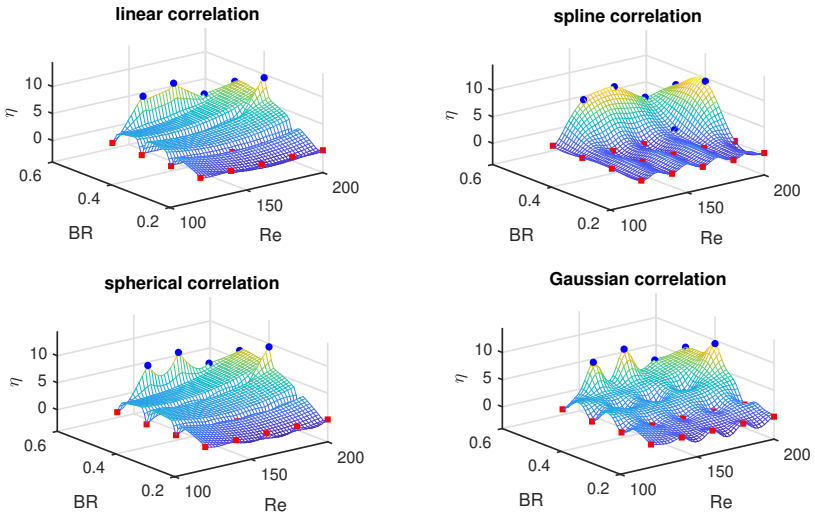


Figure 12: Surrogates for η by using different correlations for $AR = 1$.

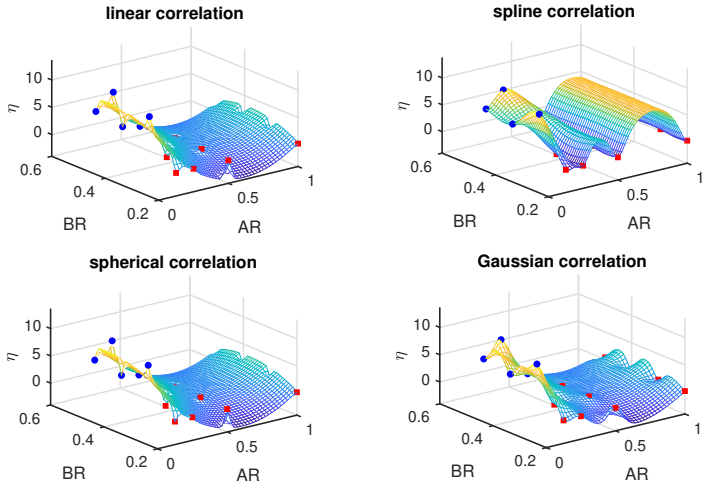


Figure 13: Surrogates for η by using different correlations for $Re = 120$.

Objective	AR	BR	Re	η	$\langle\Pi\rangle$
Max (η)	0.125	0.5	200	49.6869 %	3.30872
Min ($\langle\Pi\rangle$)	0.5	0.2	200	9.9221 %	0.8372

Table 9: Optimal values for each objective function.

fixed to 200 and thus the design space would be reduced to BR and AR. From the figure, any of the shown points in green with circular marker is a potential optimal solution to the problem. Among these, points A, B and C are finally selected as candidates. It can be noticed that if *Point*

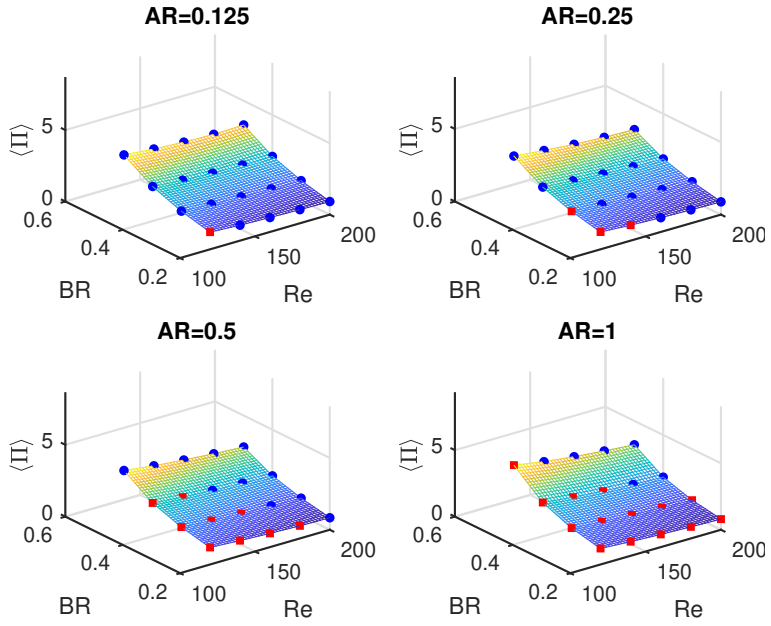


Figure 14: Surrogate models for $\langle \Pi \rangle$ shown at different AR values. Red squares are configurations without vortex shedding and blue dots are configurations with vortex shedding.

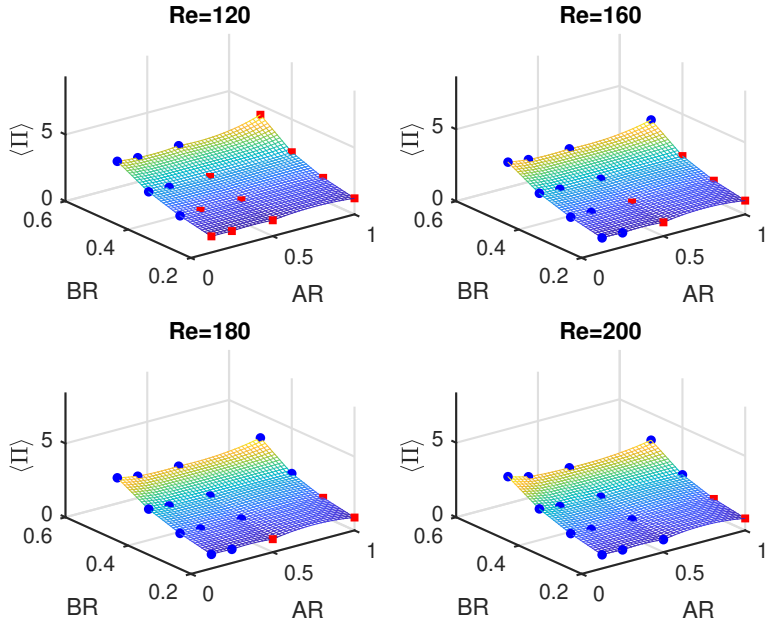


Figure 15: Surrogate models for $\langle \Pi \rangle$ shown at different Re regimes. Red squares are configurations without vortex shedding and blue dots are configurations with vortex shedding.

A is chosen, priority would be given to achieve a good mixing efficiency at high pumping power cost. On the contrary, if *Point C* is the definitive choice, one would be more interested in low power pumping to the detriment of mixing efficiency. *Point B* would be an intermediate solution.

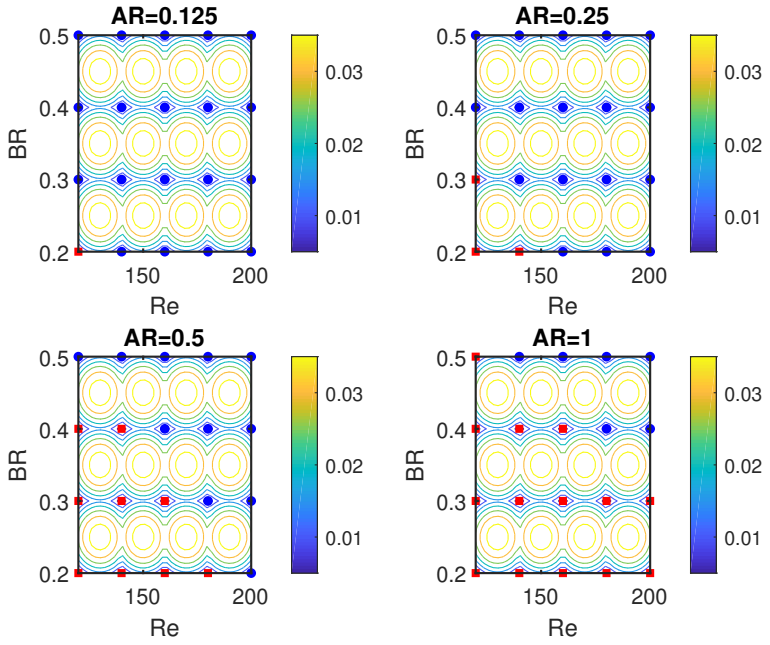


Figure 16: Surrogate models for $\langle \Pi \rangle$ shown at different AR values. Red squares are configurations without vortex shedding and blue dots are configurations with vortex shedding.

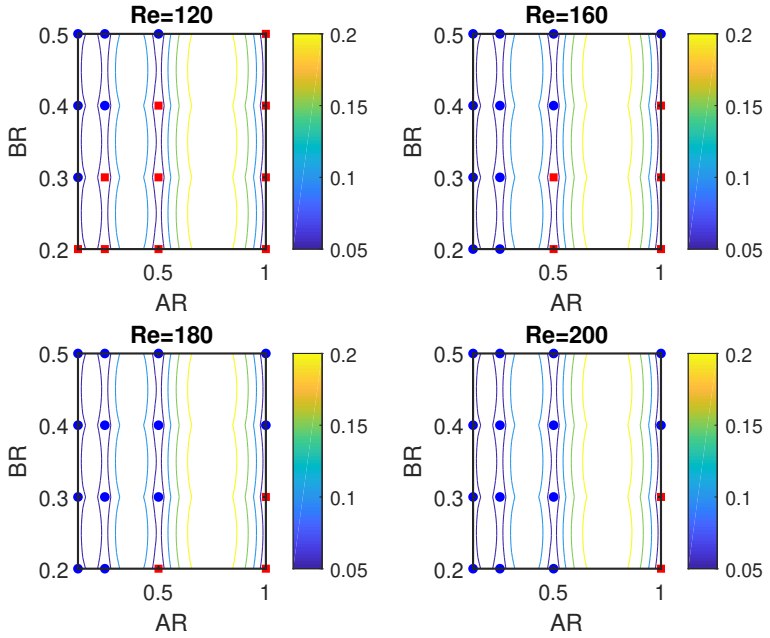
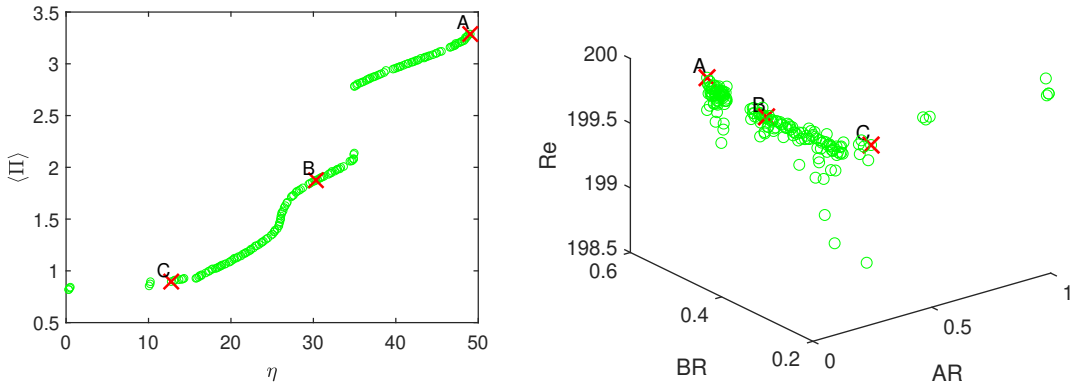


Figure 17: Surrogate models for $\langle \Pi \rangle$ shown at different Re regimes.

The cluster of points near $\eta = 0$, as it will be shown later, corresponds to cases without vortex shedding. Since the configurations of interest are, in principle, those related to vortex shedding, the surrogates could have been developed using only the training data from simulations with vortex

shedding ($VS = 1$). To demonstrate how would affect this to the final results, we have tested the optimisation process with $VS = 1$ only. The resulting Pareto front is shown in Fig. 19, which is almost identical to the Pareto front in Fig. 18 when $\eta \geq 10$, so the the cluster of optimal points near $\eta = 0$ seems to be attributed to vortex-free configurations.



(a) (b)

Figure 18: (a) Pareto front obtained by the NSGA-II algorithm on the η and $\langle \Pi \rangle$ objective functions with a population size of 500 samples. (b) Values of the optimal parameters. A, B and C are selected optimal points for reference study.

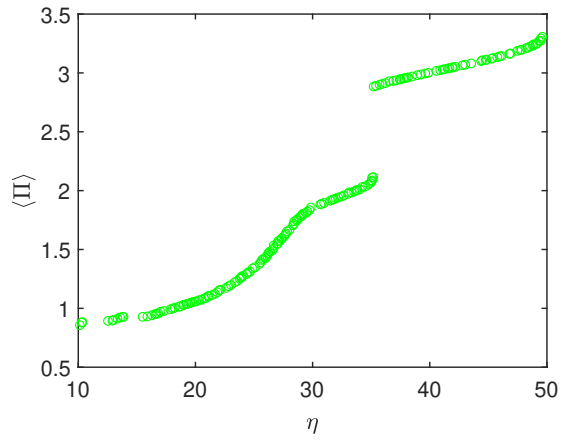


Figure 19: Pareto front obtained by the NSGA-II algorithm on the η and $\langle \Pi \rangle$ objective functions constructing the surrogates only with $VS = 1$ data.

5.2. Machine Learning-Aided Design Optimisation of a Micromixer.

The actual interest of the use of the MLADO method is to save computational resources. For this objective, the framework has been tested upon an initial set of few data samples consisting

of three points in the AR and Re design space coordinates, and four in the BR coordinate. This is a total of $N_s = 36$ simulations from the combinatorial of $Re = \{120, 160, 200\}$, $BR = \{0.2, 0.3, 0.4, 0.5\}$ and $AR = \{0.125, 0.5, 1\}$. Then, the surrogates are refined according to the MSE and the RF vortex shedding predictor trained with $N_p = 80$ (RF4), but only those cases with $VS = 1$ will be added to the original population data set of $N_s = 36$ individuals. Must be recalled that any of RF1, RF2 and RF3 vortex shedding predictors could be used, as it was demonstrated in Section 4 their capabilities (these misclassify only few configurations near the separation limit), and also that these do not need full computations for the training. Thus, we prefer to use the RF4, as this is the best trained one and there is no point on using a less accurate predictor, although the Pareto fronts to be obtained are essentially the same.

Thus, following the steps synthesized in Figure 4, and the configurations to construct surrogates explained in Section 5.1, the first step is to construct the surrogate from the initial $N_s = 36$ data and decide whether to refine. Due to the DoE was (nearly) equidistant data [1], the areas of highest MSE are those where the design space is less discretised. These areas are, for instance, those in the areas of $Re = 140$, $AR = 0.25$, $Re = 160$, and $AR = 0.8$, as both Re and AR had only 3 points in each direction and the response is highly non-linear for η. These directions can be refined in an univariate or multivariate manner. Since only $VS = 1$ will be considered to refine the design space, an univariate approach can be sufficient to demonstrate the benefit of MLADO. However, especially for more sophisticated DoE, a nested sampling may be more efficient.

According to Figure 4, there is a loop within certain conditions, which are opened to the practitioner. The first execution of the loop has been set to increase the initial $N_s = 36$ with combinatorial simulations of $Re = 140$, but only those that lead to $VS = 1$ will be included. Thus, the new number of simulation population is $N_s = 42$. Although the loop in Figure 4 suggests that the expert is who decides when to stop performing the refinement of the design space under some criteria, one can e.g. deploy the optimisation algorithms at each refinement stage and observe some sort of convergence. This is the process followed in this work.

As second execution of the loop, the MSE suggests to refine for instance the $AR = 0.25$ positions, due to the large separation between $AR = 0.125$ and $AR = 0.5$. The RF predictor is

[1] Other more sophisticated DoE can be tried with MLADO, but in this optimisation problem a distribution of equidistant individuals has been found a good option. The problem is multi-objective, so the design space must be explored sufficiently for η and Π.

used to predict which are the combinatorial $VS = 1$ cases to append to the previous surrogate samples, resulting to a total number of $N_s = 55$ samples. A third execution of the loop leads to a refinement in $Re = 180$ direction, again considering only the predicted $VS = 1$. This leads to a total number of $N_s = 68$ individuals. The problem can be further refined, but as observed in Figure 20, the results in the optimisation procedure reveal some sort of convergence. Multi-objective optimisation problems are complex to control (Pareto fronts are usually constructed to illustrate the optimals), but in single-objective optimisation, the optimal candidate is a single individual, so the procedure can be adapted e.g. to check the convergence of an optimal. Gradient-based methods could be also used in MLADO to better guide the collocation of new samples in potential regions of optimal candidates.

In Figure 20 can be observed that even with few N_s samples, the Pareto front is very close to the reference for $\eta < 35\%$ and $\langle \Pi \rangle < 2.25$. However, above this region, it is observed that some refinement for the $N_s = 55$ surrogate design space is necessary compared to the reference of $N_s = 80$ points. Thus, refinement in $Re = 180$ was necessary (the surrogates improve as the discretisation between $Re = 160$ and $Re = 200$ was too broad and optimals are located at high Reynolds). Further refinement between $Re = 180$ and $Re = 200$ could be considered, but the MSE is not large. Thus a total of $N_s = 68$ seems enough to construct reliable surrogates.

Must be noted that the optimisation process can be actually more efficient. When the initial data set of $N_s = 36$ is considered, if the optimisation algorithm is deployed, no optimal candidates are observed at low Reynolds numbers analysing the cases in the Pareto front, so the refinement at $Re = 140$ could be discarded. Also, especially for larger design spaces where by intuition is complicated to figure out the best regions, the RF predictor can be used to explore whether the number of $VS = 1$ cases is increasing or decreasing when Re (or any parameter) is varied. In this case, the number of cases with $VS = 1$ is decreasing as we approach to $Re = 140$. Said this, similarly to the abovementioned process, $AR = 0.25$ and $Re = 180$ combinatorial cases only, together with their $VS = 1$ predicted cases by the RF algorithm, can be used to refine the surrogate data set. These two refinements lead to a total number of just $N_s = 59$ samples, yielding identical results than the previous analysis with $N_s = 68$ and similar to the reference results with $N_s = 80$ (see Figure 20).

Finally, can be concluded that some computational resources are saved with the MLADO approach. To quantify the total costs in a comparison with traditional optimisation, the full

cost of each simulation will be considered. Simulations that led to steady flows ($VS = 0$) took approximately 1800 iterations at a rate of 2 seconds per iteration to converge, thus the total elapsed time was approximately 1 hour per simulation. On the other hand, the simulations that exhibited an oscillatory flow ($VS = 1$) were unsteady computations, which converged after several time steps with approximately 40E3 iterations at a rate of 2 seconds. That is, a total elapsed time of 22.2 hours per simulation. The total costs are compared in Table 10, where $N_s = 100$ would be the total cost of the full surrogate with $Re = \{120, 140, 160, 180, 200\}$, $BR = \{0.2, 0.3, 0.4, 0.5\}$ and $AR = \{0.125, 0.25, 0.5, 0.8, 1\}$.

N_s	Simulations $VS = 0$	Simulations $VS = 1$	Total hours of simulation
100	41	59	1350.8h
80	27	53	1203.6h
68	15	53	1191.6h
59	15	44	991.8h

Table 10: Computational costs (measured as simulation time) as function of the number of simulations.

To these costs one should add those related to the training of the Random Forest algorithm. However, in other studies with external data available (empirical correlations, relevant data sets of experimental data, etc.) this cost is not included (the computational resources in the training of a RF predictor are negligible). As seen in Section 4, even an algorithm with few samples is accurate. For instance, the algorithm trained with the initial data set of $N_s = 36$ full simulations (RF1) misclassified only 7 out of the 100, and would be trained at no extra costs. The algorithm trained with $N_p = 48$ (RF2) misclassified only 5 samples out of 100, and just required 12 additional samples obtained by steady partially converged simulations, as $VS = 1$ cases show oscillations in steady simulations that can be detected prior solid convergence. As the detection of vortex shedding in these simulations required approximately 2/3 of the time necessary to obtain a fully converged steady simulation, the method is still more efficient than the traditional approach. To use a predictor trained with more than $N_p = 36$ is not necessary, as it has been also seen in Section 4 that the misclassifications take place in the separation limit, which is not a region of interest in terms of high accuracy in the classification. Thence, from Table 10 can be observed that a MLADO with a final set of $N_s = 59$ may be reduce the computational time up to a 27% with respect to a standard optimisation with $N_s = 100$, and a 18% with respect to a standard

optimisation with $N_s = 80$. However, we outline that in the present application, the avoidable simulations are not the expensive ones ($VS = 0$ are steady). For other applications, i.e. designs to suppress vortex shedding, the optimisation process may be intended to designs with $VS = 0$ and i.e. reduce drag. In this scenario, the MLADO approach is recommended, as $VS = 1$ simulations would be saved in the construction of surrogates with more relevant computational savings.

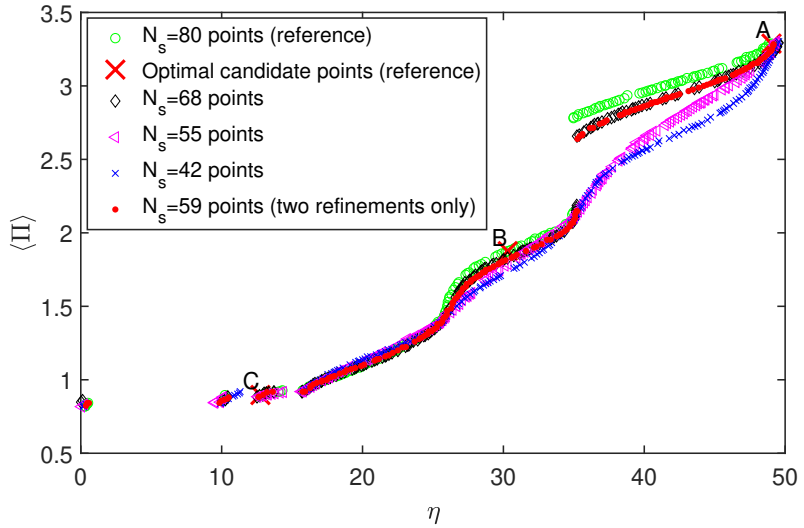


Figure 20: Pareto fronts from multi-optimisation with different number of surrogate samples N_s.

As seen in Figure 20, the selected optimal points A, B and C are the same with the MLADO. Before proceeding to simulate the selected cases, it is useful to check by means of the predictive model whether vortex shedding may take place. If there is no vortex shedding, then the performance of low efficiency configurations may be debatable. The configurations of the three points A, B and C were input to the Random Forest predictor and it predicted vortex shedding for all. The description of the three points is given in Table 11 (which includes also the simulated performance via CFD), and their representation on the surrogates is depicted in the contour plots in Fig. 21, where for the sake of visualisation, it has been assumed that for point A, B and C the optimal value of Re is exactly $Re = 200$. In Table 11 can be seen that the Kriging surrogates predicted a performance very close to the simulated scenario. This is a good indicator of the accuracy of the surrogates to model the expected performance of non-simulated configurations. The CFD relevant visualisations of the selected optimal points are shown in Fig. 22.

To conclude, a comparison of the performance with respect to previous works in the literature

Optimal candidate	AR	BR	Re	Predicted η	Predicted $\langle\Pi\rangle$	CFD η	CFD $\langle\Pi\rangle$
Point A	0.130	0.500	199.956	49.032 %	3.288	48.96 %	3.284
Point B	0.131	0.370	199.875	30.305 %	1.875	30.2 %	1.85
Point C	0.251	0.203	199.879	12.751 %	0.896	13.4 %	0.891

Table 11: Selected optimal points from the Multi-Objective optimisation problem. Predicted η and $\langle\Pi\rangle$ corresponds to the values provided by the evaluation of the surrogates. CFD η and $\langle\Pi\rangle$ are the values upon the CFD simulation of the optimal point candidates.

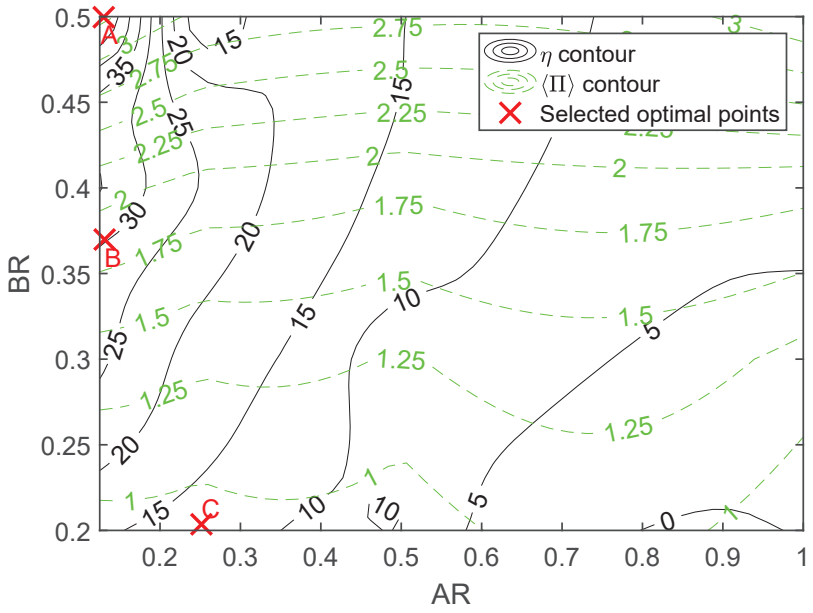


Figure 21: Representation of the selected optimal points on the η and $\langle\Pi\rangle$ surrogates for a constant value of $Re = 200$.

(which are related to different geometries) is shown in Table 12. Due to the fact that previous works used for comparison purposes are about mixing fluids with different mass fraction concentrations, Γ_c would be the Schmidt number: $\Gamma_c \equiv Sc$. Therefore, and for a fair comparison, both the Reynolds and the Schmidt numbers in the studies to be compared should be the same, but some authors used different values. Nevertheless, the comparison can still be valuable to understand the goodness of our optimal design. To support the analysis, a new simulation with Schmidt number $Sc = 10^3$, and with geometry configuration of *Point A* (the optimal candidate with the highest mixing efficiency and pumping power consumption), has been run. If the present study is compared to previous works with Schmidt numbers of 10^4 (see Table 12), the efficiency of our

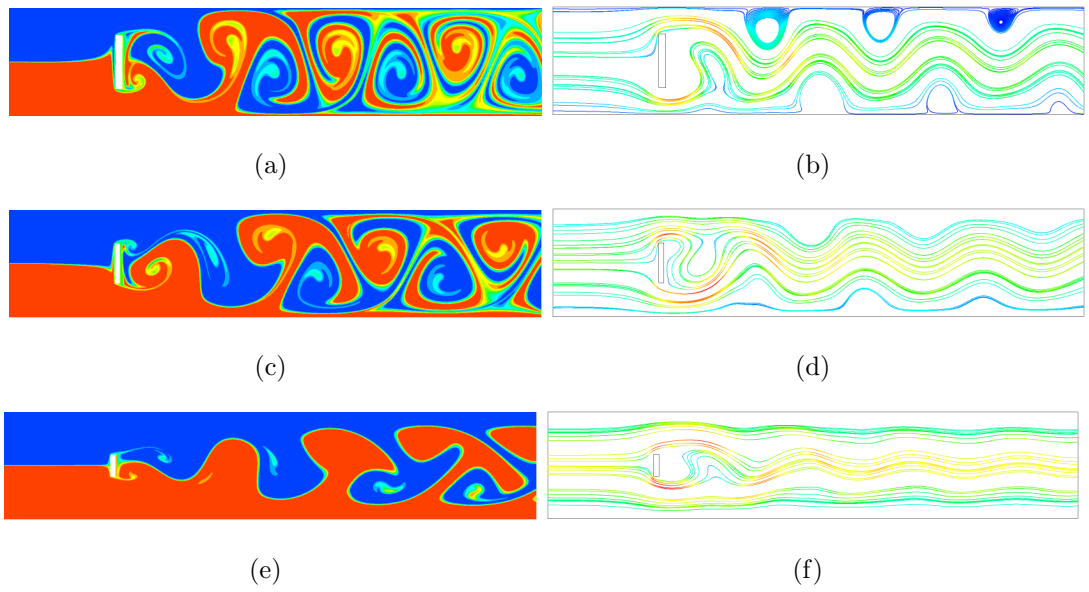


Figure 22: Config. A: a) Mass fraction. b) Streamlines. Config. B: c) Mass fraction. d) Streamlines. Config. C: e) Mass fraction. f) Streamlines. For all the figures, the streamlines are coloured according to max total velocity (red) and min total velocity (blue).

proposed design outperforms them all. In terms of pressure drop (Δp is provided in the previous works and is made dimensionless here with ρU^2), can be seen that also exhibits a low pressure drop to achieve a high efficiency. Although the pressure drop in [30] exhibits lower values, the efficiency is much lower too. Actually, if one really needs to reduce Δp, points B and C are just 1.85 and 0.891, respectively, and still outperforming the efficiency. The works in the literature related to $Sc = 10^3$, despite the good mixing performance achieved, they report large values of the pressure drop. It is important to consider the pressure drop (or the pumping power) in the comparison, because each design in the literature survey is different in terms of elements and configuration, so to focus only on the mixing would not be reliable since the energy requirements may be large. Therefore, in terms of mixing-pumping power trade-off, our microdevice seems to outperform them all, because achieves a good mixing at low pumping power. However, to clarify this statement, the comparison must be put into context. Although a mixing efficiency of $\eta \sim 50\%$ may seem small compared to other mixing efficiencies in the literature, the investigated optimal device is designed for a very high Peclet-like fluid, a very short microchannel and with just one obstacle interaction. If the length of the microchannel and/or the number of obstacles is increased, the mixing efficiency will be also increased notably, but the pressure drop will grow excessively, as seen for instance in [109, 110, 111, 90, 112]. In these works it is especially relevant the strong impact

	[90], 50, 10^3	[113], 0.29, 10^4	[112], 200, 700	[30], 200, 10^4	PR, ∼200, 10^4	PR, ∼200, 10^3
η [%]	13-90	12-14	∼90	12.8-17.8	∼49	∼52
Δp [-]	9.3-35.9	∼1800	24.6-470.5	0.74-1.96	3.284	3.284

Table 12: Comparison of mixing efficiency got by different authors. Each column header stands for: Author, Re, Sc. PR stands for Present Research.

on pressure drop by the aggregation of many obstacles. As to add more obstacles to our design would modify the vortex shedding mechanics and increase pressure drop remarkably, to extend the length of the microchannel would be an easy option to increase mixing efficiency. To give evidence of this, an additional computation has been conducted: the impact of the increase in one unit length has been tested for the optimal configuration of *Point A* with $Sc = 10^4$. The results of this length $L = 6$ are compared to the baseline optimal design of length $L = 5$. An important increase from $\eta = 49\%$ to $\eta = 58.6\%$ is observed, with a very gently increase of the dimensionless pressure drop from 3.284 to 3.348. Thus, the optimal mixer designed in this work is a very efficient option, especially when the microchannel must be short and pressure drop (pumping power) must be low. Thus, if the length is extended, the efficiency of the vortex shedding-based micromixer is expected to be notably improved with decent cost, as no additional objects are placed along the channel. Future work may be oriented to develop the present investigation as a multi-objective optimisation problem including L as design parameter, which should not be excessively long, in order to avoid a large increase in pressure drop.

6. Conclusions

In this work, an optimal design of a micromechanical device which consists of a pillar confined in a microchannel to generate vortex shedding is engineered. For this objective, a proposed Machine Learning-aided design optimisation (MLADO) framework is followed, which consists of using CFD simulations, a Random Forest predictive model and Gaussian Processes (Kriging). The MLADO framework is an approach easily extendable to other engineering applications, and can be automatised if shape parametrisation strategies are included (for instance, Computer Aided Engineering-based or mesh morphing-based). The framework has demonstrated to be an efficient way to design the device: since the surrogate predictors may need extra training data to update, the Random Forest vortex shedding predictor can be used to decide whether the new points are

necessary or not. Hence, computational resources can be saved fully simulating only configurations that lead to vortex shedding and the creation of the surrogates becomes data-driven. The use of a Genetic Algorithm (the NSGA-II algorithm) on the final surrogates has determined the optimal configurations (Pareto front) for the multi-objective problem of minimising the power consumption and maximising the mixing efficiency. Three optimal candidates have been simulated and their performance matches with negligible differences with the predicted performance via the Kriging surrogates. The optimal candidates outperform existing devices in the literature, especially in mixing per unit length, as good mixing is achieved even for a short microchannel of $L = 5$ and a single object interaction .

Supplementary Material

The trained Random Forest predictor is provided. Extension: .RData (Rstats only).

Acknowledgments

The authors want to acknowledge the financial support from the UMA18-FEDERJA-184 grant and The Andalusian Research, Development and Innovation Plan (PAIDI - Junta de Andalucia) fundings. The authors want to express also a warm acknowledgement to the anonymous reviewers, who suggested important improvements to enhance the quality of this work.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Fan Zhan, Yubo Fan, Xiaoyan Deng, and Zaipin Xu. The beneficial effect of swirling flow on platelet adhesion to the surface of a sudden tubular expansion tube: its potential application in end-to-end arterial anastomosis. ASAIO journal, 56(3):172–179, 2010.

[2] Norbert Kockmann, Michael Gottsponer, Bertin Zimmermann, and Dominique M Roberge. Enabling continuous-flow chemistry in microstructured devices for pharmaceutical and fine-chemical production. Chemistry–A European Journal, 14(25):7470–7477, 2008.

[3] Sulabh K Dhanuka, Jacob E Temme, James F Driscoll, and Hukam C Mongia. Vortex-shedding and mixing layer effects on periodic flashback in a lean premixed prevaporized gas turbine combustor. Proceedings of the Combustion Institute, 32(2):2901–2908, 2009.

[4] HE Ahmed, HA Mohammed, and Mohd Zamri Yusoff. An overview on heat transfer augmentation using vortex generators and nanofluids: approaches and applications. Renewable and Sustainable Energy Reviews, 16(8):5951–5993, 2012.

[5] Xinyan Zhao, Tao Dong, Zhaochu Yang, Nuno Pires, and Nils Høivik. Compatible immuno-nasba loc device for quantitative detection of waterborne pathogens: design and validation. Lab on a Chip, 12(3):602–612, 2012.

[6] Amir M Foudeh, Tohid Fatanat Didar, Teodor Veres, and Maryam Tabrizian. Microfluidic designs and techniques using lab-on-a-chip devices for pathogen detection for point-of-care diagnostics. Lab on a Chip, 12(18):3249–3266, 2012.

[7] Curtis D Chin, Vincent Linder, and Samuel K Sia. Lab-on-a-chip devices for global health: Past studies and future opportunities. Lab on a Chip, 7(1):41–57, 2007.

[8] Chutikarn Chaimayo, Bualan Kaewnaphan, Nattaya Tanlieng, Niracha Athipanyasilp, et al. Rapid SARS-CoV-2 antigen detection assay in comparison with real-time RT-PCR assay for laboratory diagnosis of COVID-19 in thailand. Virology Journal, 17(1):1–7, 2020.

[9] Centers for Disease Control and Prevention. Pfizer-BioNTech COVID-19 Vaccine. https://www.cdc.gov/vaccines/covid-19/info-by-product/pfizer/index.html, 2021. [Online; accessed 09-April-2021].

[10] Centers for Disease Control and Prevention. Manual for Pfizer-BioNTech COVID-19 Vaccine. https://www.cdc.gov/vaccines/covid-19/info-by-product/pfizer/downloads/prep-and-admin-summary.pdf, 2021. [Online; accessed 09-April-2021].

[11] J Ortega-Casanova. Enhancing mixing at a very low Reynolds number by a heaving square cylinder. Journal of Fluids and Structures, 65:1–20, 2016.

[12] J Ortega-Casanova. CFD study on mixing enhancement in a channel at a low Reynolds number by pitching a square cylinder. Computers & Fluids, 145:141–152, 2017.

[13] J Ortega-Casanova and C-H Lai. Cfd study on laminar mixing at a very low Reynolds number by pitching and heaving a square cylinder. Computers & Fluids, 168:318–327, 2018.

[14] Henry Reginald Arnulph Mallock. On the resistance of air. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 79(530):262–273, 1907.

[15] Henri Bénard. Formation de centres de giration a l'arrière d'un obstacle en movement. Comptes Rendus Academie des Sciences, 147:839–842, 1908.

[16] Henri Benard. Sur la zone de formation des tourbillons alternés derriere un obstacle. CR Acad. Sci. Paris, 156:1003–1005, 1913.

[17] Th Von Karman. Über den mechanismus des widerstandes, den ein bewegter körper in einer flüssigkeit erfährt. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 1911:509–517, 1911.

[18] Kai Zhang, Hiroshi Katsuchi, Dai Zhou, Hitoshi Yamada, and Zhaolong Han. Numerical study on the effect of shape modification to the flow around circular cylinders. Journal of Wind Engineering and Industrial Aerodynamics, 152:23–40, 2016.

[19] Wengang Chen, Xintao Li, and Weiwei Zhang. Shape optimization to suppress the lift oscillation of flow past a stationary circular cylinder. Physics of Fluids, 31(6):063604, 2019.

[20] Ersegun D Gedikli, David Chelidze, and Jason M Dahl. Observed mode shape effects on the vortex-induced vibration of bending dominated flexible cylinders simply supported at both ends. Journal of Fluids and Structures, 81:399–417, 2018.

[21] Benoît Pier. On the frequency selection of finite-amplitude vortex shedding in the cylinder wake. Journal of Fluid Mechanics, 458:407–417, 2002.

[22] Wagner J Pinto, Florent Margnat, and Yves Gervais. Effect of cross-section on flow three-dimensionality for prismatic bodies and the associated noise emission. In 25th AIAA/CEAS Aeroacoustics Conference, page 2531, 2019.

[23] Laura Kamps, Thomas F Geyer, Ennes Sarradj, and Christoph Brücker. Vortex shedding noise of a cylinder with hairy flaps. Journal of Sound and Vibration, 388:69–84, 2017.

[24] Éric Lamballais and Jorge Hugo Silvestrini. Direct numerical simulation of interactions between a mixing layer and a wake around a cylinder. In TSFP DIGITAL LIBRARY ONLINE. Begel House Inc., 2001.

[25] Vedant Kumar, Hemanshul Garg, Gaurav Sharma, and Rajneesh Bhardwaj. Harnessing flow-induced vibration of a D-section cylinder for convective heat transfer augmentation in laminar channel flow. Physics of Fluids, 32(8):083603, 2020.

[26] Aditya Desai, Sanchit Mittal, and Sanjay Mittal. Experimental investigation of vortex shedding past a circular cylinder in the high subcritical regime. Physics of Fluids, 32(1):014105, 2020.

[27] MR Islam and A Mohany. Vortex shedding characteristics in the wake of circular finned cylinders. Physics of Fluids, 32(4):045113, 2020.

[28] Tingting Tang, Peng Yu, Xiaowen Shan, Jianhui Li, and Shimin Yu. On the transition behavior of laminar flow through and around a multi-cylinder array. Physics of Fluids, 32(1):013601, 2020.

[29] Md Reyaz Arif and Nadeem Hasan. Vortex shedding suppression in mixed convective flow past a square cylinder subjected to large-scale heating using a non-Boussinesq model. Physics of Fluids, 31(2):023602, 2019.

[30] J Ortega-Casanova. On the onset of vortex shedding from 2D confined rectangular cylinders having different aspect ratios: Application to promote mixing fluids. Chemical Engineering and Processing-Process Intensification, 120:81–92, 2017.

[31] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160:3–24, 2007.

[32] Leo Breiman. Random forests. Machine learning, 45(1):5–32, 2001.

[33] Stephan Dreiseitl and Lucila Ohno-Machado. Logistic regression and artificial neural network classification models: a methodology review. Journal of biomedical informatics, 35(5-6):352–359, 2002.

[34] Christopher KI Williams and Carl Edward Rasmussen. Gaussian processes for machine learning, volume 2. MIT press Cambridge, MA, 2006.

[35] Karl W Schulz and Yannis Kallinderis. Numerical prediction of the hydrodynamic loads and vortex-induced vibrations of offshore structures. J. Offshore Mech. Arct. Eng., 122(4):289–293, 2000.

[36] BA Younis and TH Yang. Prediction of the effects of vortex shedding on UV disinfection efficiency. Journal of Water Supply: Research and Technology—AQUA, 60(3):147–158, 2011.

[37] Jochen Fröhlich and Wolfgang Rodi. Les of the flow around a circular cylinder of finite height. International journal of heat and fluid flow, 25(3):537–548, 2004.

[38] Eduard Naudascher and Donald Rockwell. Flow-induced vibrations: an engineering guide. Courier Corporation, 2012.

[39] RD Blevins. Review of sound induced by vortex shedding from cylinders. Journal of Sound and Vibration, 92(4):455–470, 1984.

[40] Kimon Roussopoulos. Feedback control of vortex shedding at low Reynolds numbers. Journal of Fluid Mechanics, 248:267–296, 1993.

[41] Robert W Paterson, Paul G Vogt, Martin R Fink, and C Lee Munch. Vortex noise of isolated airfoils. Journal of Aircraft, 10(5):296–302, 1973.

[42] Fei Tao, He Zhang, Ang Liu, and Andrew YC Nee. Digital twin in industry: State-of-the-art. IEEE Transactions on Industrial Informatics, 15(4):2405–2415, 2018.

[43] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. Applied logistic regression, volume 398. John Wiley & Sons, 2013.

[44] Scott Menard. Applied logistic regression analysis, volume 106. Sage, 2002.

[45] David G Kleinbaum, K Dietz, M Gail, Mitchel Klein, and Mitchell Klein. Logistic regression. Springer, 2002.

[46] Turgay Ayer, Jagpreet Chhatwal, Oguzhan Alagoz, Charles E Kahn Jr, Ryan W Woods, and Elizabeth S Burnside. Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. Radiographics, 30(1):13–22, 2010.

[47] Yasuyuki Umeda, Fujimaro Ishida, Masanori Tsuji, Kazuhiro Furukawa, Masato Shiba, Ryuta Yasuda, Naoki Toma, Hiroshi Sakaida, and Hidenori Suzuki. Computational fluid dynamics (cfd) using porous media modeling predicts recurrence after coiling of cerebral aneurysms. PloS one, 12(12), 2017.

[48] Yi-Qing Zheng, Bi-Ru Zhang, Wei-Yang Su, Jian Gong, Man-Qiong Yuan, Yuan-Lin Ding, and Shao-Qi Rao. Laryngeal aerodynamic analysis in assisting with the diagnosis of muscle tension dysphonia. Journal of voice, 26(2):177–181, 2012.

[49] Stephen John Gorman. Aerodynamic effects of vocal function exercises in elderly men. PhD thesis, University of Cincinnati, 2002.

[50] J Ross Quinlan. Decision trees and decision-making. IEEE Transactions on Systems, Man, and Cybernetics, 20(2):339–346, 1990.

[51] Jehad Ali, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. Random forests and decision trees. International Journal of Computer Science Issues (IJCSI), 9(5):272, 2012.

[52] Peter Sollich and Anders Krogh. Learning with ensembles: How overfitting can be useful. In Advances in neural information processing systems, pages 190–196, 1996.

[53] Robin Genuer, Jean-Michel Poggi, Christine Tuleau-Malot, and Nathalie Villa-Vialaneix. Random forests for big data. Big Data Research, 9:28–46, 2017.

[54] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Random forests. In The elements of statistical learning, pages 587–604. Springer, 2009.

[55] GÃŠrard Biau. Analysis of a random forests model. Journal of Machine Learning Research, 13(Apr):1063–1095, 2012.

[56] Gérard Biau and Erwan Scornet. A random forest guided tour. Test, 25(2):197–227, 2016.

[57] Maziar Raissi, Zhicheng Wang, Michael S Triantafyllou, and George Em Karniadakis. Deep learning of vortex-induced vibrations. Journal of Fluid Mechanics, 861:119–137, 2019.

[58] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378:686–707, 2019.

[59] Maziar Raissi, Alireza Yazdani, and George Em Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. Science, 367(6481):1026–1030, 2020.

[60] Jaehoon Cho, Hyunseung Kim, Addis Lulu Gebreselassie, and Dongil Shin. Deep neural network and random forest classifier for source tracking of chemical leaks using fence monitoring data. Journal of Loss Prevention in the Process Industries, 56:548–558, 2018.

[61] Peyman Zahedi, Saeid Parvandeh, Alireza Asgharpour, Brenton S McLaury, Siamack A Shirazi, and Brett A McKinney. Random forest regression prediction of solid particle erosion in elbows. Powder Technology, 338:983–992, 2018.

[62] Alfredo Aranda and Alvaro Valencia. Study on cerebral aneurysms: Rupture risk prediction using geometrical parameters and wall shear stress with CFD and machine learning tools. Machine Learning and Applications: An International Journal (MLAIJ) Vol, 5, 2018.

[63] Anh Tran, John M Furlan, Krishnan V Pagalthivarthi, Robert J Visintainer, Tim Wildey, and Yan Wang. Weargp: A computationally efficient machine learning framework for local erosive wear predictions via nodal gaussian processes. Wear, 422:9–26, 2019.

[64] Michael Joly, Soumalya Sarkar, and Dhagash Mehta. Machine learning enabled adaptive optimization of a transonic compressor rotor with pre-compression. In ASME Turbo Expo 2018: Turbomachinery Technical Conference and Exposition. American Society of Mechanical Engineers Digital Collection, 2018.

[65] Sethuraman Sankaran, Leo Grady, and Charles A Taylor. Impact of geometric uncertainty on hemodynamic simulations using machine learning. Computer Methods in Applied Mechanics and Engineering, 297:167–190, 2015.

[66] Anand Pratap Singh, Shivaji Medida, and Karthik Duraisamy. Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils. AIAA Journal, pages 2215–2227, 2017.

[67] Rohit K Tripathy and Ilias Bilionis. Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification. Journal of computational physics, 375:565–588, 2018.

[68] Peter J Diggle and Paulo J Ribeiro. Gaussian models for geostatistical data. Model-based geostatistics, pages 46–78, 2007.

[69] Timothy W Simpson, Timothy M Mauery, John J Korte, and Farrokh Mistree. Kriging models for global approximation in simulation-based multidisciplinary design optimization. AIAA journal, 39(12):2233–2241, 2001.

[70] Francisco-Javier Granados-Ortiz, Carlos Perez Arroyo, Guillaume Puigt, Choi-Hong Lai, and Christophe Airiau. On the influence of uncertainty in computational simulations of a high-speed jet flow from an aircraft exhaust. Computers & Fluids, 180:139–158, 2019.

[71] Shinkyu Jeong, Mitsuhiro Murayama, and Kazuomi Yamamoto. Efficient optimization design method using kriging model. Journal of aircraft, 42(2):413–420, 2005.

[72] Julien Laurenceau and Matthieu Meaux. Comparison of gradient and response surface based optimization frameworks using adjoint method. AIAA Paper, 1889:2008, 2008.

[73] Azfarizal Mukhtar, Khai Ching Ng, and Mohd Zamri Yusoff. Optimal design of opening ventilation shaft by Kriging metamodel assisted multi-objective genetic algorithm. Int. J. Model. Optim, 7:92–97, 2017.

[74] Enrica Bernardini, Seymour MJ Spence, Daniel Wei, and Ahsan Kareem. Aerodynamic shape optimization of civil structures: A CFD-enabled Kriging-based approach. Journal of Wind Engineering and Industrial Aerodynamics, 144:154–164, 2015.

[75] F Duchaine, T Morel, and LY M. Gicquel. Computational-fluid-dynamics-based Kriging optimization tool for aeronautical combustion chambers. AIAA journal, 47(3):631–645, 2009.

[76] Yu Zhang, Sanbo Hu, Jinglai Wu, Yunqing Zhang, and Liping Chen. Multi-objective optimization of double suction centrifugal pump using Kriging metamodels. Advances in Engineering Software, 74:16–26, 2014.

[77] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE transactions on evolutionary computation, 6(2):182–197, 2002.

[78] Nidamarthi Srinivas and Kalyanmoy Deb. Multiobjective optimization using nondominated sorting in genetic algorithms. Evolutionary computation, 2(3):221–248, 1994.

[79] N Srinivas and K Deb. Multiobjective optimization using NSGA. Evolutionary Computing, 2(3), 1995.

[80] Tushar Goel, Rajkumar Vaidyanathan, Raphael T Haftka, Wei Shyy, Nestor V Queipo, and Kevin Tucker. Response surface approximation of pareto optimal front in multi-objective optimization. Computer methods in applied mechanics and engineering, 196(4-6):879–893, 2007.

[81] Claudio Comis Da Ronco, Rita Ponza, and Ernesto Benini. Aerodynamic shape optimization of aircraft components using an advanced multi-objective evolutionary approach. Computer Methods in Applied Mechanics and Engineering, 285:255–290, 2015.

[82] Carlo Poloni, Andrea Giurgevich, Luka Onesti, and Valentino Pediroda. Hybridization of a multi-objective genetic algorithm, a neural network and a classical optimizer for a complex design problem in fluid dynamics. Computer Methods in Applied Mechanics and Engineering, 186(2-4):403–420, 2000.

[83] XD Wang, C Hirsch, Sh Kang, and C Lacor. Multi-objective optimization of turbomachinery using improved NSGA-II and approximation model. Computer Methods in Applied Mechanics and Engineering, 200(9-12):883–895, 2011.

[84] Mohammad Darvish Damavandi, Mostafa Forouzanmehr, and Hamed Safikhani. Modeling and pareto based multi-objective optimization of wavy fin-and-elliptical tube heat exchangers using cfd and nsga-ii algorithm. Applied Thermal Engineering, 111:325–339, 2017.

[85] KL Vasudev, R Sharma, and SK Bhattacharyya. A multi-objective optimization design framework integrated with CFD for the design of auvs. Methods in Oceanography, 10:138–165, 2014.

[86] Stefan Jakobsson, Michael Patriksson, Johan Rudholm, and Adam Wojciechowski. A method for simulation based optimization using radial basis functions. Optimization and Engineering, 11(4):501–532, 2010.

[87] Marco Evangelos Biancolini. Fast radial basis functions for engineering applications. Springer, 2017.

[88] Marco Evangelos Biancolini, Emiliano Costa, Ubaldo Cella, Corrado Groth, Gregor Veble, and Matej Andrejašič. Glider fuselage-wing junction optimization using CFD and RBF mesh morphing. Aircraft Engineering and Aerospace Technology, 2016.

[89] A Serani, R Pellegrini, J Wackers, C-E Jeanson, P Queutey, Michel Visonneau, and M Diez. Adaptive multi-fidelity sampling for CFD-based optimisation via radial basis function meta-models. International Journal of Computational Fluid Dynamics, 33(6-7):237–255, 2019.

[90] Sourav Sarkar, KK Singh, V Shankar, and KT Shenoy. Numerical simulation of mixing at 1–1 and 1–2 microfluidic junctions. Chemical Engineering and Processing: Process Intensification, 85:227–240, 2014.

[91] Suhas V Patankar and D Brian Spalding. A calculation procedure for heat, mass and momentum transfer in three-dimensional parabolic flows. In Numerical prediction of flow, heat transfer, turbulence and combustion, pages 54–73. Elsevier, 1983.

[92] Patrick J Roache. Perspective: a method for uniform reporting of grid refinement studies. 1994.

[93] Julio M Ottino. Mixing and chemical reactions a tutorial. Chemical Engineering Science, 49(24):4005–4027, 1994.

[94] Dieter Bothe, Carsten Stemich, and Hans-Joachim Warnecke. Computation of scales and quality of mixing in a T-shaped microreactor. Computers & Chemical Engineering, 32(1-2):108–114, 2008.

[95] S Turki, H Abbassi, and S Ben Nasrallah. Effect of the blockage ratio on the flow in a channel with a built-in square cylinder. Computational Mechanics, 33(1):22–29, 2003.

[96] Pratish P Patil and Shaligram Tiwari. Effect of blockage ratio on wake transition for flow past square cylinder. Fluid Dynamics Research, 40(11-12):753, 2008.

[97] Atul Sharma and V Eswaran. Heat and fluid flow across a square cylinder in the two-dimensional laminar flow regime. Numerical Heat Transfer, Part A: Applications, 45(3):247–269, 2004.

[98] Foster Provost. Machine learning from imbalanced data sets 101. In Proceedings of the AAAI'2000 workshop on imbalanced data sets, volume 68, pages 1–3. AAAI Press, 2000.

[99] Marco Evangelos Biancolini, Katia Capellini, Emiliano Costa, Corrado Groth, and Simona Celi. Fast interactive CFD evaluation of hemodynamics assisted by RBF mesh morphing and reduced order models: the case of aTAA modelling. International Journal on Interactive Design and Manufacturing (IJIDeM), 14(4):1227–1238, 2020.

[100] Andrew P Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition, 30(7):1145–1159, 1997.

[101] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. R news, 2(3):18–22, 2002.

[102] Gilles Louppe. Understanding random forests: From theory to practice. PhD thesis, University of Liege, 2014.

[103] Botros N Hanna, Nam T Dinh, Robert W Youngblood, and Igor A Bolotnov. Machine-learning based error prediction approach for coarse-grid computational fluid dynamics (cg-cfd). Progress in Nuclear Energy, 118:103140, 2020.

[104] Thais Mayumi Oshiro, Pedro Santoro Perez, and José Augusto Baranauskas. How many trees in a random forest? In International workshop on machine learning and data mining in pattern recognition, pages 154–168. Springer, 2012.

[105] Christian Borgelt, Frank Höppner, and Frank Klawonn. Guide to intelligent data analysis. Springer, 2010.

[106] Kennedy P Kusumo, Lucian Gomoescu, Radoslav Paulen, Salvador Garcia Munoz, Constantinos C Pantelides, Nilay Shah, and Benoit Chachuat. Bayesian approach to Probabilistic Design Space Characterization: A nested sampling strategy. Industrial & Engineering Chemistry Research, 59(6):2396–2408, 2019.

[107] Wang Kun, Chen Fu, Yu Jianyang, and Song Yanping. Nested sparse-grid Stochastic Collocation Method for uncertainty quantification of blade stagger angle. Energy, 201:117583, 2020.

[108] Søren N Lophaven, Hans Bruun Nielsen, Jacob Sondergaard, and A Dace. DACE. A matlab kriging toolbox. Technical University of Denmark, Lyngby, Technical Report No. IMMTR-2002, 12, 2002.

[109] Harrson S Santana, Deborah S Tortola, Joao L Silva Jr, and Osvaldir P Taranto. Biodiesel synthesis in micromixer with static elements. Energy conversion and management, 141:28–39, 2017.

[110] Yu-Cheng Lin, Yung-Chiang Chung, and Chung-Yi Wu. Mixing enhancement of the passive microfluidic mixer with J-shaped baffles in the tee channel. Biomedical microdevices, 9(2):215–221, 2007.

[111] Chien-Chong Hong, Jin-Woo Choi, and Chong H Ahn. A novel in-plane passive microfluidic mixer with modified Tesla structures. Lab on a Chip, 4(2):109–113, 2004.

[112] Nita Solehati, Joonsoo Bae, and Agus P Sasmito. Numerical investigation of mixing performance in microchannel T-junction with wavy structure. Computers & Fluids, 96:10–19, 2014.

[113] J Ortega-Casanova. Application of CFD on the optimization by response surface methodology of a micromixing unit and its use as a chemical microreactor. Chemical Engineering and Processing: Process Intensification, 117:18–26, 2017.