Aerospace Science and Technology 84 (2019) 399–411

Contents lists available at ScienceDirect

Aerospace Science and Technology

www.elsevier.com/locate/aescte

Surrogate model of complex non-linear data for preliminary nacelle design

Alexander Heidebrecht *, David G. MacManus

Cranfield University, MK43 0AL Cranfield, Bedfordshire, UK

ARTICLE INFO

Article history:
Received 11 February 2018
Received in revised form 4 August 2018
Accepted 18 August 2018
Available online 3 October 2018

Keywords:
Nacelle drag response
Surface model
Surrogate model
Preliminary design

ABSTRACT

Most response surface methods typically work on isotropically sampled data to predict a single variable and fitted with the aim of minimizing overall error. This study develops a metamodel for application in preliminary design of aircraft engine nacelles which is fitted to full-factorial data on two of the eight independent variables, and a Latin hypercube sampling on the other six. The specific set of accuracy requirements for the key nacelle aerodynamic performance metrics demand faithful reproduction of parts of the data to allow accurate prediction of gradients of the dependent variable, but permit less accuracy on other parts. The model is used to predict not just the independent variable but also its derivatives, and the Mach number, an independent variable, at which a certain condition is met. A simple Gaussian process model is shown to be unsuitable for this task. The new response surface method meets the requirements by normalizing the input data to exploit self-similarities in the data. It then decomposes the input data to interpolate orthogonal aerodynamic properties of nacelles independently of each other, and uses a set of filters and transformations to focus accuracy on predictions at relevant operating conditions. The new method meets all the requirements and presents a marked improvement over published preliminary nacelle design methods.

© 2018 The Authors. Published by Elsevier Masson SAS. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Response surface models (RSM) are interesting in the design of engineering systems, as they promise not just near-instant evaluation of complicated designs but also to maximize the use that can be made of existing data. For this reason, they have become frequently used in various engineering studies, both to augment and accelerate optimisation processes [1] and as substitutes for the higher-fidelity methods used to generate the training data, particularly in preliminary design [2]. Since the development of the original Kriging method [3], several variants were developed, and improvements have been found, leading to a generalization in the form of Gaussian processes. Standardized modelling approaches have been developed for a range of applications [4]. However, several shortcomings have also been noted [5]. These include the simplicity of academic problems used to demonstrate RSM performance as well as the assumption of univariate output. Although the methods are continuously being developed [6], there are many real-world applications to which the current best practices for Kriging models cannot be directly applied. This may be because data cannot be easily sampled in the recommended way or because the data characteristics change very quickly from continuous to highly non-linear. In some cases the metric of interest is not the predicted variable itself but its derivative with respect to an independent variable, or the value of the independent variables at which the dependent variable exhibits a given feature. At the same time, preliminary design applications permit comparably low accuracy on absolute performance figures as long as the effects of design changes are modelled correctly. This allows the use of methods with very selective accuracy, where data of low interest can be simplified to a large extent.

This study investigates a preliminary design problem where all of the above difficulties arise and develops an RSM to predict multiple performance criteria. Since the study was limited to established methods with readily-available implementations, it was chosen to use a combination of several Gaussian process models. Due to the resources needed to produce training data, and specific research interests, the modelling techniques developed in this study are used exclusively on the problem of aero-engine nacelle design. The resulting RSM is intended for use in preliminary nacelle design studies, and in trade-off studies investigating the wider field of engine design and integration. However, it is believed that the general approach to establishing an interpolation model, or some of the RSM components may be useful for other

* Corresponding author.
E-mail address: a.heidebrecht@cranfield.ac.uk (A. Heidebrecht).

https://doi.org/10.1016/j.ast.2018.08.020
1270-9638/© 2018 The Authors. Published by Elsevier Masson SAS. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

Abbreviations

CFD	Computational Fluid Dynamics
CST	Class-Shape Transformation
DoE	Design of Experiments
GCI	Grid Convergence Index
LHS	Latin Hypercube Sampling
OK	Ordinary Kriging
RSM	Response Surface Method

Symbols

A	Area
c_D	nacelle drag coefficient, related to highlight area
c'_D	normalized drag coefficient
$\overline{c'_D}$	mean normalized drag coefficient
$c_{D,ref}$	reference drag
$g_1 ... g_6$	non-dimensional nacelle design variables
f_{max}	non-dimensional position of maximum nacelle radius
f_{nx}	coefficient in nugget equation
l_{nac}	nacelle length
l_{nx}	limit in nugget equation
M	flight Mach number
M'_T	normalized, transformed Mach number
M_{DR}	drag rise Mach number
M_{ref}	reference Mach number
$MFCR$	mass-flow capture ratio
$\dot{m}_{intake}$	intake mass-flow
n	nugget
r	radial coordinate
r_{hi}	nacelle highlight radius
r_{if}	initial forebody radius: radius of curvature at nacelle highlight
r_{max}	maximum nacelle radius
r_{TE}	nacelle trailing edge radius
v	flow velocity
x	independent variable
y	dependent variable
z	dependent variable
β_{nac}	boat-tail angle
γ	isentropic exponent
ρ	density
σ	standard deviation

Modifiers

X_{cruise}	cruise condition
X_{EOC}	end-of-cruise condition
X_{hi}	nacelle highlight
X_i	index
X_{max}	largest value
X_{min}	smallest value
X_T	transformed quantity
X_∞	state at upstream far-field position
X'	normalized quantity
Δx	difference
$\mathrm{Q}x$	quotient

problems, as it deals with challenges which are frequently found in engineering.

1.1. Aerodynamic nacelle design

The problem of the preliminary design of turbofan nacelles is becoming increasingly relevant, due to the increase in fan diameter which is typically associated with low-specific-thrust aero-engines with high bypass ratios. This increase is a consequence of the demand for quieter and more fuel-efficient engines and the fact that both goals require an increase in propulsive efficiency, leading to reduced specific thrust [7]. This generally means an increase of the outer engine radius and thus the surface of the nacelle. The increased weight and drag from the nacelle counteracts the efficiency gained by reducing specific thrust [8,9]. This eventually outweighs the propulsive efficiency benefits [10]. For this reason it is important to assess the aerodynamic properties of potential nacelles early in the engine design process.

In this context, the precise achievable drag, although important, is not the only focus of attention. A nacelle must also be able to function at a range of operating conditions without exhibiting adverse behaviour. Geometrical constraints for nacelle design are imposed by the intake and nozzle design and the need to provide sufficient volume to accommodate auxiliary units and thrust reversers. The task for a nacelle designer in this scenario is to determine the minimal overall dimensions required for a nacelle to fulfil all mechanical and aerodynamic requirements, and to estimate the resulting drag of such a design [11].

1.2. Performance metrics of nacelles

Fig. 1a shows the drag of an example nacelle design as a function of the aerodynamic operating conditions (M, $MFCR$). The drag map is not rectangular since the maximum achievable $MFCR$ is limited by the intake capacity, which depends on the flight Mach number and the intake design (see Section 1.1).

Nacelle drag coefficient varies only little at low flight Mach numbers and large $MFCR$, and stays at a low baseline level, compared to other operating conditions. Towards both large flight Mach numbers and small $MFCR$, the drag increases, due to compressibility effects and the acceleration around the profile leading edge, respectively. From a nacelle design perspective, this drag increase should not take place at the conditions for which a nacelle is being designed, both to avoid high nacelle drag itself and because a nacelle with a very large shock or separated flow may also influence the flow over the wing and cause additional drag or loss of lift. Fig. 1b shows so-called drag-rise curves which highlight how the sharp increase to nacelle drag as a function of M is influenced by $MFCR$. For many designs, wave drag can become more than one order of magnitude larger than the other nacelle drag components. Fig. 1c shows that the effect of $MFCR$ on nacelle drag is more gradual than the wave drag rise but is also non-linear.

During a typical flight, the aircraft will take-off at $M \approx 0.25$, $MFCR \approx 2$, then accelerate and climb at close to maximum $MFCR$ until it has reached cruise, usually at $M > 0.8$. For the purpose of this study, it is assumed that there is a margin of 0.02 between the cruise Mach number (M_{cruise}) and the drag-rise Mach number (M_{DR}), at which wave drag starts to rise sharply (Eq. (2)). There is no universal definition for M_{DR}, and this study uses the gradient-based criterion defined in Eq. (1). Not all response surfaces used in this study provide a straightforward way to compute derivatives. In order to keep a consistent approach, all drag gradients from response surfaces were evaluated using finite differences, with $\Delta M = 10^{-4}$. Once at cruise condition, the engines are throttled back slightly to sustain altitude and velocity. While $MFCR$ can vary significantly between different engines, it usually stays at or above 0.7 at the start of cruise. For this study, a representative

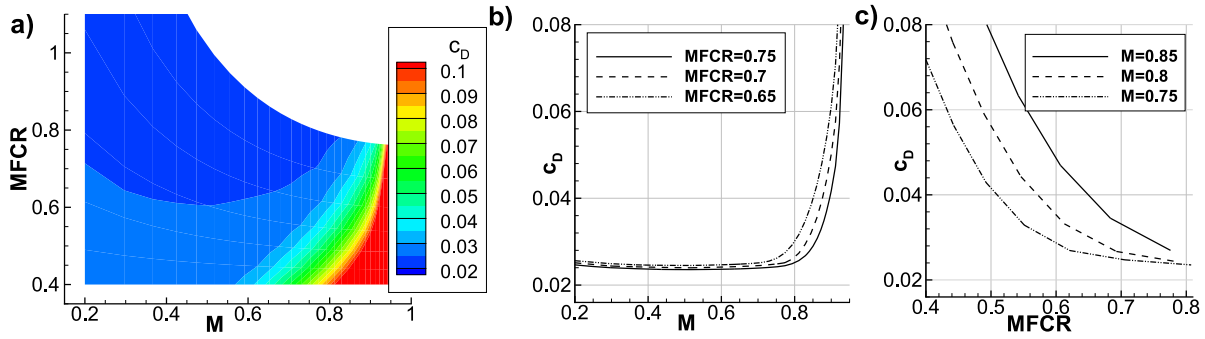


Fig. 1. (a) Drag map of an example nacelle. (b) c_D as a function of M ("drag-rise curves") at constant $MFCR$. (c) c_D as a function of $MFCR$ ("spillage curves") at constant M. Drag coefficients use highlight area as reference area.

mass-flow capture ratio of 0.75 was assumed for the cruise condition (Eq. (2)). As the aircraft burns fuel during cruise, it becomes lighter and requires less thrust towards the end of the cruise segment. Depending on nacelle design and the mission, $MFCR$ may be reduced to 0.6 at end-of-cruise. During descent, $MFCR$ is further decreased to 0.6 or below. Values of 0.5 or less are usually only reached in situations like engine failure. In such cases, the additional drag from the nacelle causes a yawing moment, which means that there is a limit to how much additional drag such a reduction in MFCR may cause [10]. The resulting requirements for a preliminary design method are accurate prediction of nacelle drag at $MFCR > 0.6$, with reduced accuracy acceptable below this mark. The lower bound for the regarded $MFCR$ range was chosen to be 0.4

$$M_{DR} = M \left|_{\frac{dc_D}{d_M}=0.1}\right. \tag{1}$$

$$MFCR_{cruise} = 0.75, \qquad M_{cruise} = M_{DR}\,(MFCR_{cruise}) - 0.02 \tag{2}$$

1.3. Data characteristics and challenges for RSM

A response surface method models a variable as a function of a set of independent variables. Often, an implicit assumption is that these independent variables are created equal, i.e. that an isotropic sampling method can be used [5], such as Latin hypercube sampling (LHS) [12,13]. Another frequent implicit assumption is that predictions need to be equally accurate for all combinations of input variables, and that noise characteristics are similar across the sampled region. Lastly, most response surfaces are only used to predict the value of one dependent variable [5], whereas the RSM at hand needs to also provide values of independent variables at which a certain condition is fulfilled, and multiple pieces of information derived from the dependent variable.

For the response surface model developed in this study, the assumptions outlined above are violated. The following Sections 1.3.1–1.3.3 describe these challenges in more detail.

1.3.1. Data sampling

The response surface regarded in this work models nacelle drag coefficient, c_D, as a function of a set of six geometrical design variables (g_1 through g_6), as well as the two variables which describe the operating conditions, M and $MFCR$. From this data, other performance metrics such as M_{DR} are then derived.

The generally recommended sampling plan for RSMs is Latin hypercube sampling (LHS) [4,13]. However, this would require that for each nacelle design, the drag is sampled at only one single operating condition. Once a nacelle geometry is generated, a mesh created and a converged solution is obtained, it is much more computationally efficient to generate multiple solutions for the same nacelle design at different operating conditions than it is to repeat the full CFD process for each data point. A much larger amount of data can be generated for a given computational budget by densely sampling nacelle drag along the M and $MFCR$ axes and sampling more coarsely in geometrical design space. Another reason to sample anisotropically is the need to compute gradients of drag as a function of flight Mach number, which is required to evaluate M_{DR} (Eq. (1)). Since the drag gradient changes rapidly within the transonic range ($0.75 < M < 1.0$), intervals between samples need in this range need to be on the order of 0.01. Similarly dense sampling on the other independent variables would be computationally prohibitive. This means that not all independent variables can be treated equally by the RSM, and that the two variables determining the operating condition need to use sampling plans than the six geometrical design variables.

Since the space of valid operating conditions is not rectangular (Fig. 1), the sampling plan in (M, $MFCR$) cannot be a simple full factorial or a LHS. Either a non-rectangular sampling plan needs to be used, or an affine transformation must be applied to map a rectangular sampling plan onto the space of valid operating conditions in a drag map.

1.3.2. Varying data characteristics and accuracy requirements

Fig. 2 shows how nacelle drag characteristics can change rapidly with operating conditions, and across different designs. Accurate drag prediction is mostly relevant at cruise Mach number at the cruise condition and $MFCRs$ down to 0.6 (Section 1.2, Eq. (2)). The cruise Mach number is below drag rise by definition (Eq. (2)), and additional drag incurred while reducing $MFCR$ from this condition is low for all acceptable nacelle designs. For Nacelle designs which do not meet these acceptability requirements, an accurate performance prediction is not required. This means that, effectively, the operating conditions for which accurate predictions are needed are always within the part of the dragmap with only small variations in drag. Outside of this range, drag prediction is only important as it needs to be known whether a design meets the acceptability requirements or not, but not the exact margin by which the requirements are missed.

The drag-rise criterion (Eq. (1)) also results in an additional requirement on the RSM, to predict the range of operating conditions for which a particular nacelle design provides sufficient drag-rise margin. By testing this range against the mission profile of a particular aircraft, the designer can determine the viable range of nacelle designs to choose from. The resulting requirement on the RSM is accurate prediction of M_{DR}, as a function of $MFCR$ and the design variables. Since M_{DR} depends on drag gradients, this requires an especially consistent prediction of c_D up to drag-rise Mach number, but not at $M > M_{DR}$.

The required accuracy varies not just with operating conditions but also across different regions in the geometric design space. If a design has particularly poor aerodynamic characteristics, it is suffi-

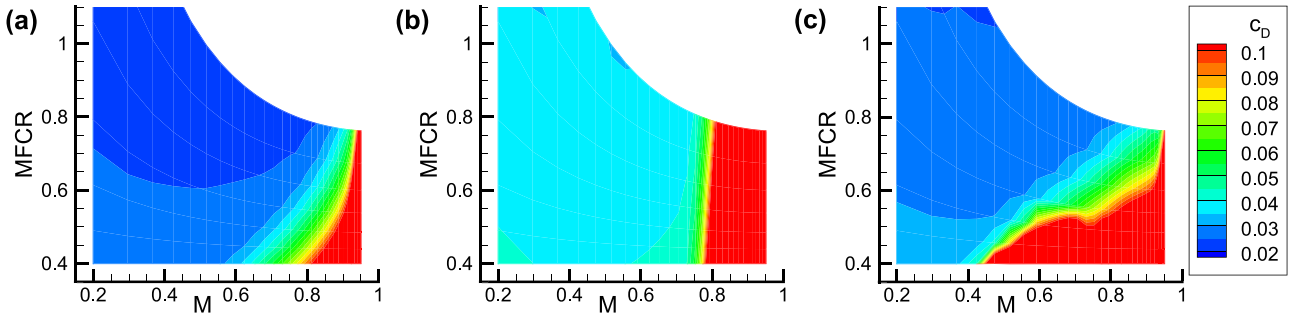


Fig. 2. Drag maps of several example nacelles designs which feature (a) low overall drag, moderate spillage and good M_{DR}, (b) high drag, very low M_{DR}, (c) extreme spillage.

cient to know that a threshold has been crossed. For example, the very abrupt drag increase at low *MFCR* in Fig. 2c does not need to be represented accurately, as long as it is known that spillage drag is past the limit of acceptability.

Training a response surface model to model such highly nonlinear parts of the input data means tuning the model parameters to represent the very rapid and erratic changes to drag seen in Fig. 2c at $MFCR < 0.6$, or the considerable drag gradients at flight Mach numbers above drag rise. Such a model would not be optimally suited to represent the much smoother drag maps of desirable nacelles (e.g. Fig. 2a), and would be expected to have reduced accuracy in more pertinent regions of the drag map ($M < M_{DR}$, $MFCR > 0.6$). One way to address such a problem is to reduce the sampled space to exclude regions of undesirable data [14]. However, this is not an option in the current study since the extent of the problematic region is not known a priori, since the range of accurate predictions needs to include the transition regions toward undesirable data, and because even for undesirable designs or unrealistic operating conditions, an indication of drag is still needed. A suitable RSM must thus prioritize the modelling of desirable features, while still providing indicators when performance is not acceptable.

1.3.3. Data shift on independent variables

One of the main performance figures for nacelles is the drag-rise Mach number (M_{DR}), as it determines the maximum flight Mach number at which the nacelle can operate with an acceptable drag. This poses a challenge: Fundamentally, M_{DR} is not the dependent variable predicted by the RSM, but rather the value of an independent variable at which the characteristics of the dependent variable change. This very distinctive feature of every drag map shifts along the M axis as the design changes, similar to the effect shown in Fig. 1b, where M_{DR} changes as a function of *MFCR*.

The dense sampling on the M axis required to correctly determine M_{DR} cannot be feasibly used on the other axes, but the resulting RSM still needs to provide the necessary level of detail to determine M_{DR} at points in the design space at which no input data has been sampled. This cannot be expected of a standard interpolation method which only regards amplitude of the independent variable (c_D). As an example, a hypothetical Kriging method trained on the two drag-rise curves at $MFCR = 0.65$ and $MFCR = 0.75$ in Fig. 1b, is regarded. Although the main difference between both input curves is a shift of about 0.05 along the M axis, an interpolation to $MFCR = 0.7$ would be expected to produce a blend between the input curves, where drag begins to rise early but more gradually than in either of the input curves.

To solve this problem, a suitable RSM must model not just the nacelle drag accurately but also has to include some measure of the shift of the drag-rise feature along the Mach number axis which occurs as a result of changes to *MFCR* or the design variables. A naïve interpolation method which simply models drag as a function of all independent variables can not be expected to perform well in this regard.

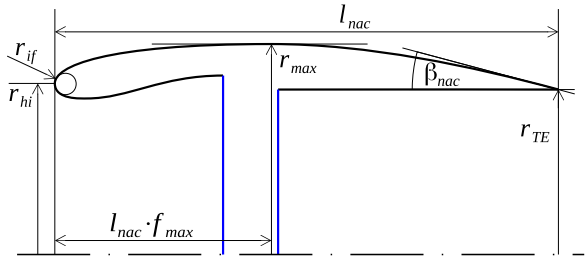


Fig. 3. The parameters used to specify the nacelle geometry.

1.4. Input data generation

The data used in this study was obtained with the process for axisymmetric CFD simulations for nacelles described by Heidebrecht et al. [15]. The process is built around a parametric nacelle geometry (Fig. 3). The nacelle is constructed from a CST curve [16], constrained by an analytical approach [17] to conform to a set of six intuitive design variables. The geometrical set-up and the dimensional design variables are shown in Fig. 3, and are similar to the set-up used in other CFD simulations with the aim of determining nacelle drag [18]. The design variables are mapped to six non-dimensional degrees of freedom to define a non-dimensional design space (Table 1). The non-dimensional variables define the nacelle shape relative to the highlight radius, r_{hi}. r_{hi} was kept constant for all designs in the input dataset, as it only affects the scale and has therefore only a small effect on aerodynamic coefficients, in the form of Reynolds number effects.

The CFD simulations were conducted with a compressible implicit flow solver [19], with the assumption of a perfect gas with a constant isentropic exponent of $\gamma = 1.4$. They employed the k–ω SST turbulence model. The domain inlet boundary condition was set to the farfield total pressure, to generate an efflux which matches the farfield velocity. This was done to generate a generic exhaust condition independent of any specific engine design, and to minimize post-exit forces, which are not present in ideal pressure-matched exhausts [20]. The mesh consisted of approximately 39,000 cells, and a mesh sensitivity study found a grid convergence index (GCI) [21] of 1% on nacelle drag, based on meshes with halved and doubled resolution (11,000 and 155,000, respectively). A domain sensitivity study used domains between 50 and 90 nacelle radii, and found the influence of domain size on nacelle drag to be on the order of 0.01%, for the chosen domain size of 80 nacelle radii. Nacelle drag was extracted as modified standard drag [20], using the method developed by Christie [9]. Fig. 4 illustrates a typical convergence history, a mesh and a flow solution, for the same design used to produce the data in Fig. 1. The

Table 1
Dimensional and non-dimensional design variables for the parametric nacelle geometry, and the design space limits.

Variable	Description	Non-dimensional	min	max
r_{max}	Maximum radius	$g_1 = \frac{(r_{max} - r_{hi})}{l_{nac}}$	0.017	0.126
l_{nac}	Nacelle length	$g_2 = \frac{l_{nac}}{r_{hi}}$	2.6	4.5
f_{max}	Location of maximum radius	$g_3 = f_{max}$	0.2	0.5
r_{if}	Initial forebody radius	$g_4 = f_{if} = \frac{r_{if} \cdot f_{max} l_{nac}}{(r_{max} - r_{hi})^2}$	0.6	1.5
f_{te}	Trailing edge radius	$g_5 = \frac{r_{te} - r_{hi}}{l_{nac}}$	−0.05	0.035
β_{nac}	Boat-tail angle	$g_6 = \beta_{nac}/^\circ$	9.5	14.5

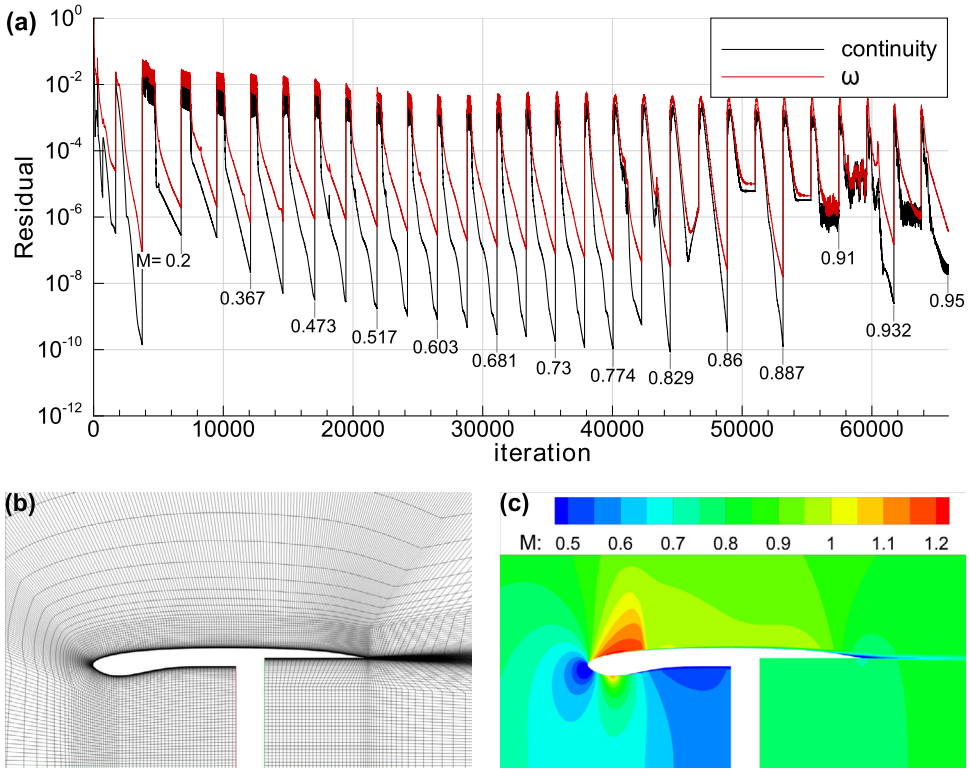


Fig. 4. Data from the CFD simulation for a representative nacelle design. (a) Convergence history for one Mach number sweep, at maximum MFCR; (b) mesh; (c) flow field at $M = 0.845$, $MFCR = 0.78$.

CFD approach was also previously validated against measurements [22]. Fig. 4 shows an example of convergence, mesh and flow solution for a representative nacelle design.

2. Methodology

Although it is computationally possible to generate tens of thousands of CFD results for the simple axisymmetric CFD model employed, the requirement to evaluate many operating conditions for each nacelle design means that the number of designs sampled in the 6-dimensional design space is much lower than the total number of sampled nacelle drags. This may hinder the use of methods which might otherwise be useful for large numbers of samples, and it favours an approach which treats the design space separately from the 2-dimensional space which defines the operating conditions.

Consequently, the response surface modelling problem was decomposed into several components. First, an interpolation method for individual drag maps is established for the full-factorial data generated for each input design. This is used to extract two scaling references from each of the drag maps in the input dataset. These scaling references to the two main outputs of the full model, drag at subcritical Mach numbers, and the onset of wave drag. They help exploit the self-similarities between drag maps of different designs, and allow the data to be normalized in such a way that the main features of drag maps are mapped to the same normalized input variables for all designs. The remaining part of the RSM process uses normalized variables, and it models the individual differences in the shapes of drag maps, relative to their commonalities, as a function of the design variables.

2.1. Analysis and interpolation for single design

The following sections describe how the data obtained for a single design is sampled, how a drag map RSM is created, and how the scaling references are determined and used to non-dimensionalize the drag map data.

2.1.1. Data sampling and non-dimensionalisation

The aim of the sampling method for drag within the space of operating conditions (M and $MFCR$) is to provide data suitable to subsequently establish an accurate interpolation model for a drag map. It needs to provide sufficient sampling density where drag changes sharply, while not expending too many sampling points in areas of the drag map where drag changes very little. Particularly around the onset of wave drag, the sampling plan needs to provide

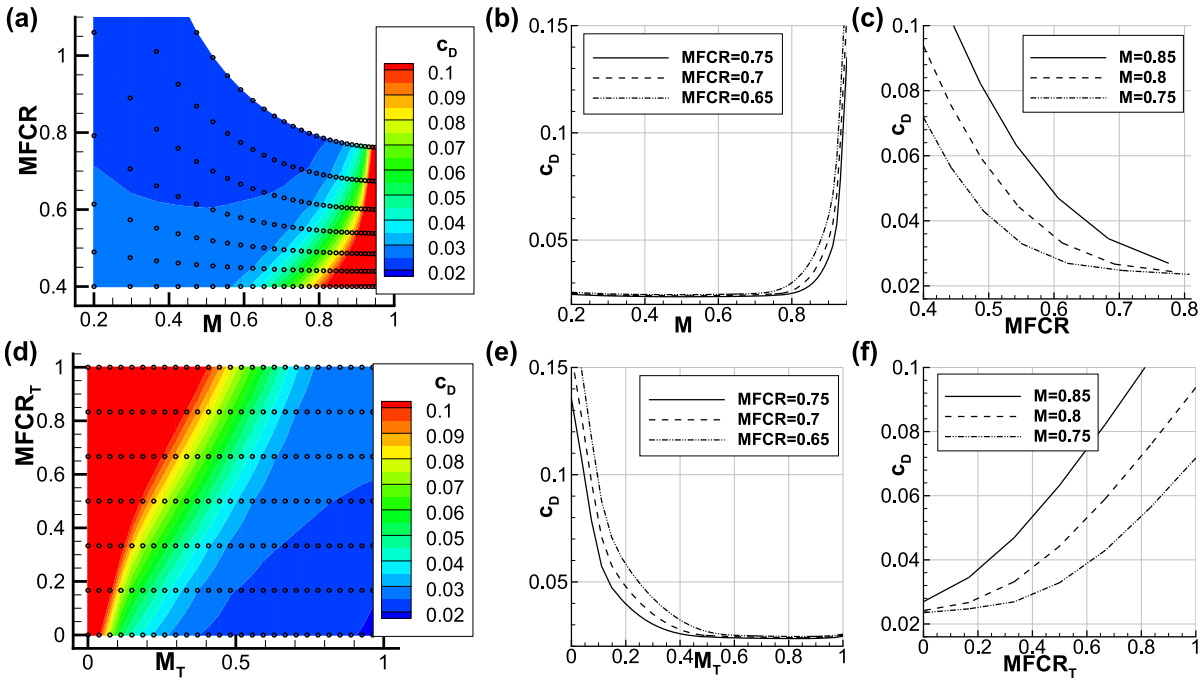


Fig. 5. (a) Drag map of an example nacelle. Symbols indicate sample locations. (b) c_D as a function of M at for a range of fixed $MFCR$. (c) c_D as a function of $MFCR$ at different M. (d), (e), (f) The same data as functions of M_T and $MFCR_T$.

sufficient data to accurately compute the drag gradient information required to find M_{DR} (Eq. (1)).

The maximum possible $MFCR$ at a given Mach number, $MFCR_{choke}$ can be computed for isentropic flow (Eq. (3)). Because in real flows the intake boundary layer effectively reduces the intake throat area, and because a flow very close to choking may cause divergence in a CFD solver, the maximum value was set to 95% of the isentropic choking value. The lowest sampled MFCR was chosen as 0.4 (Eq. (4)), to allow prediction for values of $MFCR$ that may be reached in engine-out scenarios (Section 1.2). Rather than distributing sampling points evenly within the interval, it was decided to use a mapping which is linear with $\sqrt{MFCR}^{-1}$, mapped to the interval [0; 1] (Eq. (5)). The inverted square root function increases the sampling density towards lower $MFCR$, where larger drag gradients are expected, and the linear mapping allows the drag map interpolation to work on a unit square.

$$MFCR_{choke} = \left(\frac{1+0.2M^2}{1.2}\right)^3 \frac{1}{CR \cdot M} \tag{3}$$

$$MFCR_{min} = 0.4; \qquad MFCR_{max} = 0.95 MFCR_{choke} \tag{4}$$

$$MFCR_T = \frac{\sqrt{MFCR}^{-1} - \sqrt{MFCR_{max}}^{-1}}{\sqrt{MFCR_{min}}^{-1} - \sqrt{MFCR_{max}}^{-1}} \tag{5}$$

The sampled Mach number interval was chosen to be 0.2–0.95 (Eq. (6)). Since it is known that aerodynamic effects which are dependent on Mach number mainly appear at high Mach numbers, a sampling distribution was chosen which takes this into account. In this work, the Prandtl–Glauert factor, $\sqrt{1-M^2}$, which is used in the Prandtl–Glauert transformation [23], is used to linearize some of the compressibility effects. This biases the sampling density towards larger Mach numbers. Like $MFCR_T$, the transformed Mach number is linearly mapped to the interval [0; 1] (Eq. (7)).

$$M_{min} = 0.2; \qquad M_{max} = 0.95 \tag{6}$$

$$M_T = \frac{\sqrt{1-M^2} - \sqrt{1-M_{max}^2}}{\sqrt{1-M_{min}^2} - \sqrt{1-M_{max}^2}} \tag{7}$$

Within the square created by M_T and $MFCR_T$, a linear full factorial sampling was generated. Since the nacelle drag is influenced much more by Mach number than by $MFCR$, the sampling plan comprises 28 samples along the M_T axis and 7 samples along the $MFCR_T$ axis. Fig. 5 shows the sampling plan as a function of the original (M and $MFCR$) and transformed (M_T and $MFCR_T$) operating conditions. It can be seen that the transformation successfully biases the sampling towards the conditions where the largest changes in drag occur, and reduce some of the non-linearities in the data. For example, lines of drag at constant Mach number show less curvature when plotted as a function of $MFCR_T$ (Fig. 5c, f). The drag-rise curves as a function of M_T show a much more gradual onset of wave drag than the same curve plotted against M, with the segments on either end of the curve approaching linear trends (Fig. 5b, e).

2.1.2. Interpolation of drag maps

To interpolate within a drag map, a Gaussian process model [24] is generated. This uses the implementation in *scikit-learn* [25]. The implementation offers the ability to specify a custom regression model, as well as a local nugget variable to express the measure of confidence in the accuracy of a given value. The definition for the nugget is shown in Eq. (8), where σ_i is the standard deviation on the input value at index i. The base nugget value chosen for all input data was $n_{base} = 4 \cdot 10^{-6}$, which corresponds to a mean prediction error of 0.2% in the CFD results.

$$n = \left(\frac{\sigma_i}{y_i}\right)^2 \tag{8}$$

It was found that attempting to model all of the potentially irregular input data (e.g. Fig. 2c) with this accuracy can corrupt the RSM, as it leads to a low estimated correlation between the data points, which results in oscillations throughout the entire map.

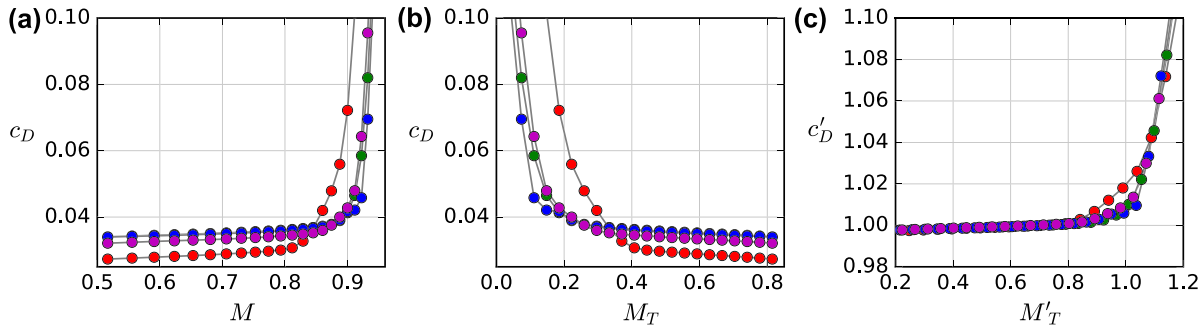


Fig. 6. Four example drag-rise curves, sampled at maximum *MFCR*. (a) Raw input data, c_D against M; (b) c_D against M_T; (c) transformed and normalized to $c_{,ref}$ and M_{ref}, c'_D as function of M'_T.

Therefore, the nugget values for some of the irregular input data were increased to reduce their impact on the model parameters. Such an increase can be justified for data from parts of the drag map where the respective design would not conceivably be required to perform well. This is the case for data above drag rise or data points with very high spillage drag. Such data does not need to be accurately reproduced by the RSM, as long as it is possible to provide an indication that drag is unacceptably high. Even at conditions below drag-rise, small-scale irregularities may be smoothed out. In the context of preliminary design, this corresponds to the assumption that it is possible to remove such small-scale features from the drag map at a later stage in the design process.

To deal with non-smooth data, the nugget was increased by two additional terms which depend on the smoothness of the drag map, as indicated by the second-order finite differences in the input data, $\frac{\delta^2 c_D}{\delta M_T^2}$ and $\frac{\delta^2 c_D}{\delta^2 MFCR_T^2}$. For both indicators, acceptability limits were determined by analysing the dragmaps for nacelles with acceptable performance. Should the indicator fall outside these limits for a given data point, the nugget for this data point is increased proportionally to the square of the margin by which each limit is violated.

A Gaussian process model is then generated to predict c_D as a function of M_T and $MFCR_T$. It uses a squared exponential correlation model [24] and employs a custom regression function, to make use of existing knowledge about the general shape of drag maps. The custom regression model consists of second-order polynomials for both M_T and $MFCR_T$, in addition to exponential terms dependent on M_T, as well as all combinations of interaction terms between the two input variables. The exponential terms allow the model to account for a quickly increasing drag gradient with increasing Mach numbers.

2.1.3. Scaling references and normalisation

To prepare the data for interpolation between the drag maps of different nacelle designs, two scaling references are determined which are then used to normalize the data and to make it easier to exploit some fundamental self-similarities between aerodynamic properties of different nacelle designs. The first scaling reference, $c_{D,ref}$ was chosen to be the nacelle drag at $M = 0.6$, $MFCR = 0.75$ (Eq. (9)). This is representative of the level of the typical drag plateau at low Mach numbers and high MFCR. $c_{D,ref}$ also indicates the level of drag at which accuracy is most important, as there should not be large changes in drag between the reference operating conditions (Eq. (9)), and the cruise condition (Eq. (2)).

The second scaling reference, M_{ref}, is defined using the gradient criterion in Eq. (10). This uses a very similar definition as the drag-rise Mach number (Eq. (1)) but using a larger threshold for the drag gradient. The reason for this is that M_{ref} can be determined with greater accuracy than M_{DR} because any inaccuracy in determining the drag gradient has a smaller influence on the result if the gradient is larger.

$$c_{D,ref} = c_D\,(M = 0.6, MFCR = 0.75) \tag{9}$$

$$M_{ref} = M \left| \frac{dc_D\,(MFCR_{max})}{d_M} = 0.5 \right. \tag{10}$$

Each drag map is then normalized with respect to the two scaling references. To do this, two new variables are introduced, the normalized drag, c'_D (Eq. (11)), and the normalized transformed Mach number, M'_T (Eq. (12)). This normalization maps both the reference drag ($c_{D,ref}$) and the reference Mach number ($M_{ref,T}$) to nominal values of 1. The different stages of the transformation from M and c_D to M'_T and c'_D are shown for four sample drag-rise curves in Fig. 6. As can be seen, many of the systematic differences between the curves are removed by the normalization. Most importantly, the bulk of the wave drag increase appears at the same position on the M'_T axis for every design, and the slopes of the curves at $M'_T > 1$ are almost equal. The main remaining differences between the drag-rise curves are at $0.8 < M'_T < 1.1$ (Fig. 6c). These differences are of much smaller amplitude than in the input data (Fig. 6a) and concern the shape of the curves during the onset of wave drag rise.

$$c'_D = c_D - c_{D,ref} + 1 \tag{11}$$

$$M'_T = \frac{1 - M_T}{1 - M_{ref,T}} \tag{12}$$

This normalisation step allows the subsequent process to use separate models for the different aerodynamic properties of nacelle designs, represented by the scaling references ($c_{,ref}$ and M_{ref}) and the normalized input data (c'_D as a function of M'_T, $MFCR_T$ and the design variables).

2.2. Interpolation between multiple designs

2.2.1. Sampling in design space

To populate the design space, a Latin hypercube sampling (LHS) [12,26], was generated, which covers the entire six-dimensional design space (g_1–g_6, Table 1). Some combinations of design variables lead to nacelle geometries which are not fully convex, and such designs were rejected since they are expected to have unacceptable drag characteristics. This process generated 235 valid nacelle designs from a LHS of over 500 samples. After constructing an initial RSM, an additional 55 infill samples were generated to improve the quality of predictions. The infill samples were selected based on the uncertainty estimator for a Kriging model trained to predict M_{ref} from the initial sample nacelle designs. For 5 of the 290 resulting designs, the CFD process failed due to extreme and

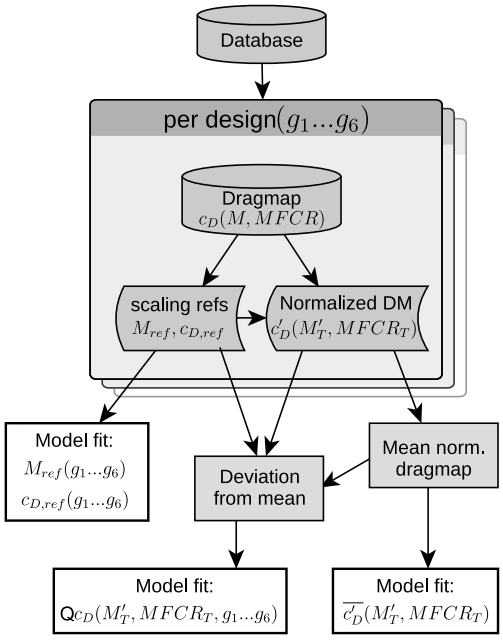


Fig. 7. Overview of the RSM construction process.

adverse aerodynamics. The total set of CFD results contained 285 designs. For each of these designs, the method described in Section 2.1.1 was used to determine drag at the combinations of M and $MFCR$ shown in Fig. 5. This data was then used to train a drag map interpolation model (Section 2.1.2) for each drag map to compute the scaling references ($c_{D,ref}$, M_{ref}) and normalized drag maps (Section 2.1.3). For 6 of the 285 sample designs, the drag map interpolation model generation failed due to highly erratic drag distributions associated with severely adverse designs, and they were removed from the set. The remaining set contains 279 designs with 196 CFD solutions per design, comprising 54684 CFD solutions.

For each design, the drag map processing (Section 2.1.3) yields the scaling references (M_{ref} and $c_{D,ref}$), a Gaussian process to interpolate the drag map, and the drag map itself, normalized to M'_T, $MFCR_T$ and c'_D.

2.2.2. Construction of RSM

A schematic overview of the process by which the RSM is constructed is shown in Fig. 7. It is explained in detail below.

After establishing the scaling references and computing the normalized dragmaps for each design in the database, two Gaussian process models are generated to interpolate the scaling references (M_{ref} and $c_{D,ref}$) across the design space. In both cases, the choice of correlation and regression model, as well as the constant nugget value were optimized to minimize the RMS of the prediction errors obtained through leave-one-out (LOO) auto-crossvalidation. LOO auto-crossvalidation is the best-practice method to obtain an unbiased measure of the accuracy of machine learning models [27].

The ability to predict the scaling references (M_{ref} and $c_{D,ref}$) at any point in the geometrical design space makes it possible to convert between dimensional (M, $MFCR$) and normalized (M'_T, $MFCR_T$) operating conditions for any design. This allows the RSM generation process to work on the normalized drag maps, whose drag-rise slopes and subcritical drag levels are aligned, as illustrated in Fig. 6c. The part of the data for which the drag prediction must be accurate is the subset where $c_D \approx c_{D,ref}$, and thus $c'_D \approx 1$. This is only a relatively small portion of the full range of c'_D in the input data.

Fitting an RSM to directly interpolate the normalized drag map data in design space is similar to generating a drag map interpolation model (Section 2.1.2), as there are regions of strong nonlinearities and regions in which the data changes little but must be modelled very accurately. To solve this problem, the normalized drag map data was decomposed further, into a mean normalized drag map, $\overline{c'_D} = f\left(M'_T, MFCR_T\right)$, and a deviation from this mean. This was done by using the Gaussian process models for each input drag map to interpolate all normalized drag maps onto one regular grid in M'_T and $MFCR_T$, and compute the mean normalized drag on each of the grid points. To prevent data from designs with particularly adverse aerodynamic properties from influencing the result, and to improve quality of the method for nacelles with beneficial aerodynamics, input data points were ignored by the averaging process if drag was more than two standard deviations above the mean normalized drag (c'_D) at the same normalized operating condition ($M'_T, MFCR_T$). The averaged drag data were then used to generate a drag map interpolation model in normalized coordinates (modelling $\overline{c'_D} = f(M'_T, MFCR_T)$). Similar to a custom regression model for a Gaussian process, this allows to split out large systematic trends from the data.

After computing the mean normalized drag map, the drag quotient $\mathrm{Q}c_D$ (Eq. (13)) was computed for each normalized input drag map. $\mathrm{Q}c_D$ is a function of M'_T and $MFCR_T$ and represents the quotient between the drag of each sampled nacelle design to the drag that would be obtained by re-dimensionalizing the mean normalized drag map.

$$\mathrm{Q}c_D\left(M'_T, MFCR_T\right) = \frac{c'_D\left(M'_T, MFCR_T\right) - 1 + c_{D,ref}}{\overline{c'_D}\left(M'_T, MFCR_T\right) - 1 + c_{D,ref}} \tag{13}$$

Fig. 8 illustrates the process using the example drag-rise curves shown in Fig. 6. The four example drag-rise curves in Fig. 8a are decomposed into the mean drag-rise curve of all nacelle designs in the dataset (Fig. 8b), and the drag quotient $\mathrm{Q}c_D$ (Fig. 8c). The mean drag-rise curve contains much of the systematic trends. With these systematic components removed, the individual differences between the drag-rise curves make up a much larger proportion of the overall signal in the $\mathrm{Q}c_D$ data (Fig. 8c), compared to the normalized curves (Fig. 8a).

The last component of the overall RSM method is a Gaussian process model which interpolates $\mathrm{Q}c_D$, as a function of all eight independent variables: six design variables ($g1$ through $g6$) and the normalized operating conditions M'_T and $MFCR_T$. It was found that this Gaussian process model produced the most accurate results when choosing a cubic correlation model [28]. No regression function was used, as systematic trends have already been extracted from the data as far as possible. Each input data point was assigned the same nugget as for the drag map interpolation model described in Section 2.1.2. Since there are as many input points as there are CFD solutions in the input dataset (ca. 55,000), this requires a significant amount of computational working memory. A set of three filters was applied before fitting the model, to remove data from the set which may make prediction of desirable performance less accurate, and at the same time reducing the memory requirements per nacelle design to allow inclusion of a larger number of sample designs. The first filter removes all data above M_{ref} (Eq. (14)). Since $M'_T = 1$ maps to M_{ref} and $M_{ref} > M_{DR}$, this only removes data above the Mach number range in which accurate drag prediction is required. The second filter removes all data points at which $\mathrm{Q}c_D$ is larger than three (Eq. (15)), i.e. drag data points which is more than three times as high as would be expected based on the mean normalized drag and the reference drag at the same normalized operating conditions (M'_T, $MFCR_T$). The justification for this filter is that designs which generate three times the drag as an average nacelle at the same nor-

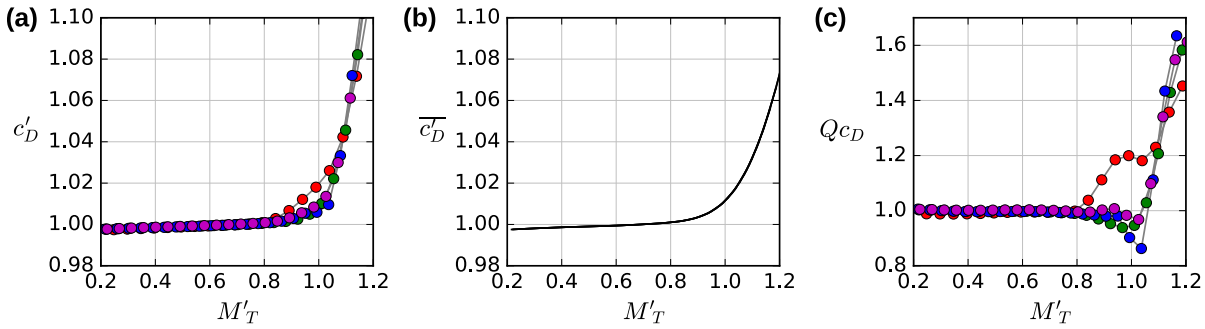


Fig. 8. (a) Normalized example drag-rise curves, (b) mean drag-rise curve, (c) the Qc_D function for the example drag-rise curves.

malized conditions are not suitable to operate at this condition, therefore an accurate prediction is not required. It is also assumed that there remains enough data in the input set with $Qc_D < 3.0$ at neighbouring operating conditions to inform the model that Qc_D is very large. The third filter removes all data points from the input set with a nugget value above 2.0 (Eq. (16)). According to Eq. (8), this means that the expected relative uncertainty on the input value is greater than $\sqrt{2}$ (i.e. larger than 140% of the input value itself), and the information lost from removing the input would not have notable impact on the model prediction.

$$M'_T < 1.0 \tag{14}$$

$$Qc_D < 3.0 \tag{15}$$

$$n < 2.0 \Rightarrow \frac{\sigma_{Qc_D}}{Qc_D} < \sqrt{2} \tag{16}$$

2.2.3. Drag interpolation method

The model building procedure generates one Gaussian process model to predict each of the scaling references (M_{ref} and $c_{D,ref}$) across the design space, one Gaussian process model to predict the normalized mean drag as a function of the normalized operating condition, specified by M'_T and $MFCR_T$, and a Gaussian process model to predict Qc_D (Eq. (13)), as a function of the normalized operating condition and the six design variables. To predict drag for a given nacelle geometry at a given operating condition, a set of six design variables and an operating condition, determined by M and $MFCR$, must be provided. First, the scaling references (M_{ref} and $c_{D,ref}$, Eqs. (10), (9)) are predicted from the design variables, using the Gaussian process models for $c_{D,ref}$ and M_{ref}. The operating condition (M and $MFCR$) is converted to normalized variables (M'_T and $MFCR_T$), using Eqs. (12) and (5), and the interpolated M_{ref}. With the normalized operating condition, the normalized mean drag map model (Fig. 8b) can then be evaluated to predict the mean normalized drag coefficient, $\overline{c'_D}$. The drag quotient Qc_D is predicted from the same variables (Eq. (13), Fig. 8c). From this data, the predicted drag coefficient can be computed by inverting Eq. (13) and combining with Eq. (11) into Eq. (17).

$$c_D = \left(\overline{c'_D} - 1 + c_{D,ref}\right) Qc_D \tag{17}$$

To predict M_{DR} for a given design, at a given $MFCR$, one could potentially generate a separate RSM to model $M_{DR} = f(MFCR, g_1 \ldots g_6)$ directly. This was not pursued because the results of such an independent RSM would not necessarily be consistent with the derivatives of the predicted drag, and because M_{DR} is not guaranteed to be a continuous function of the other seven independent variables, and there may be more than one unique Mach number for a given design at one $MFCR$ which meets the drag-rise criterion (Eq. (1)). Instead, M_{DR} is found by predicting the drag-rise curve for a nacelle design and iteratively evaluating drag gradient to find the Mach number which satisfies Eq. (1).

3. Results

Several tests of components of the RSM and of the overall model were carried out. An investigation into the distribution of nacelle drag prediction errors across different operating conditions was also conducted. Predictions from the drag map interpolation model were tested against a high-resolution drag map generated from CFD. The interpolation models for the scaling references (M_{ref} and $c_{D,ref}$) were tested through LOO auto-crossvalidation, and the overall RSM was tested against an independent CFD dataset, which allows for statistical computation of the model accuracy. Finally, the distribution of nacelle drag prediction errors for different operating conditions was characterized by comparing predictions from the RSM and CFD for several example nacelle designs.

3.1. Drag map interpolation

Both the computation of the scaling references ($c_{D,ref}$ and M_{ref}) from training data and the evaluation of mean normalized drag, $\overline{c'_D}$, in the drag prediction process (Section 2.2.3) depend on the drag map interpolation method described in Section 2.1.2. The drag map interpolation thus needs to provide not just accurately interpolated drag but must be consistent enough to provide reliable estimates of the derivative of drag in regard to Mach number. To test the drag map interpolation model, a CFD drag map with a resolution of 125 Mach numbers and 30 values for $MFCR$ was generated for one specific nacelle design. This was in addition to a drag map with standard resolution of 28 Mach numbers and 7 $MFCR$s, generated for the same nacelle design.

Fig. 9 shows contour levels of drag obtained from CFD data sampled at high resolution (Fig. 9a) and standard resolution (Fig. 9b). The comparison of drag-rise criteria (solid lines in Fig. 9b) based on each dataset shows an offset of 0.0024 for all $MFCR > 0.6$, when using first-order finite differences to compute the drag gradient (Eq. (1)). The same offset is observed with respect to the criterion for M_{ref} (Eq. (10), dashed lines in Fig. 9b). Another visible difference between both maps are the oscillations of the line for $dc_D/dM = 0.1$ in the high-resolution data (Fig. 9a). A direct comparison between the two maps found that these originate with small-scale oscillations in the CFD results for nacelle drag at $MFCR < 0.55$, $M > 0.6$, with an amplitude of less than 0.5% of nacelle drag. At these low $MFCR$ values, drag in transonic flow is very sensitive to small changes in the highly loaded boundary layer interacting with a sonic shock. This effect causes a reduction in numerical convergence and amplifies its effect on the computed nacelle drag. Although the amplitude of these oscillations is very small, the calculation of dc_D/dM is affected due to the high sampling rate used for the high-resolution drag map, while the effect is not noticeable at the normal sampling resolution. Both the inaccurate M_{DR} and M_{ref} obtained by linear interpolation from the standard resolution drag map and the irregular gradients obtained

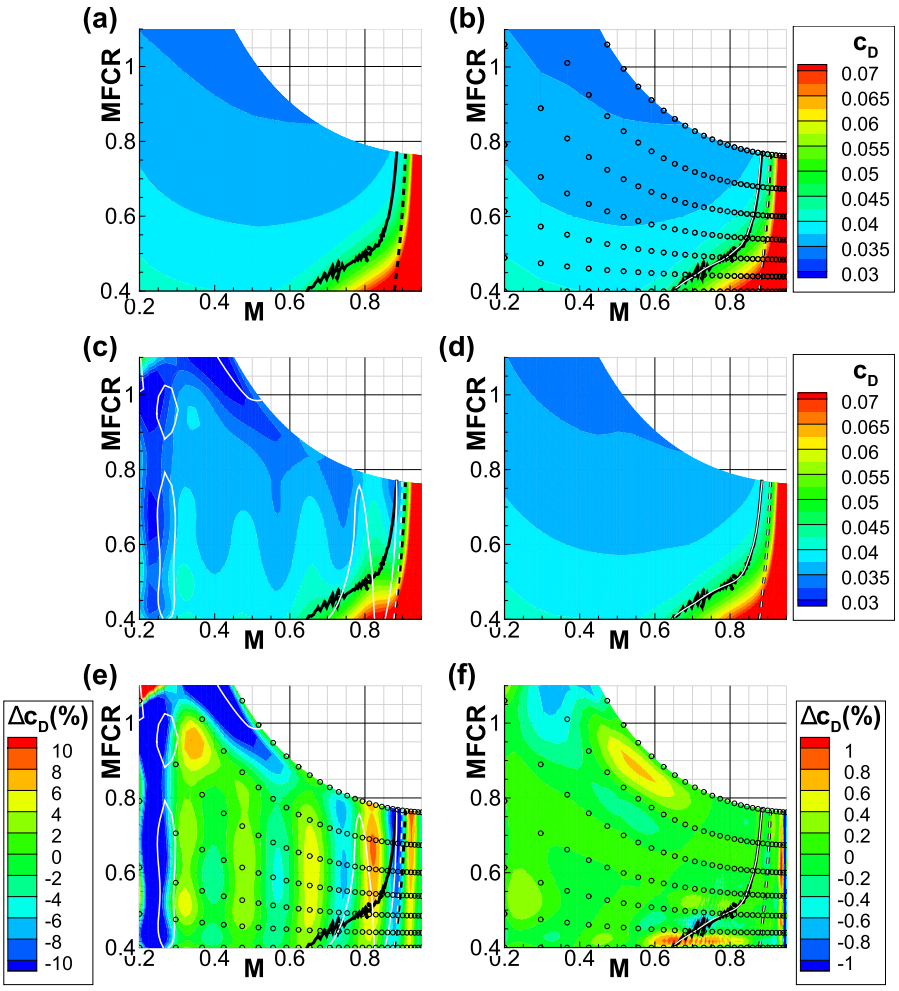


Fig. 9. Drag map from (a) high-resolution CFD data (125×30 samples) and (b) standard resolution (28×7). Drag maps interpolated from standard resolution to high resolution using (c) ordinary Kriging (OK) and (d) the proposed new drag map interpolation method (Section 2.1.2). Difference between high-resolution data and maps interpolated by (e) OK and (f) the proposed new method. Solid lines mark $dc_D/dM = 0.1$, dashed lines mark $dc_D/dM = 0.5$. Black lines represent data from the high-resolution CFD drag map. Symbols denote sampling locations.

from the high-resolution drag map confirm the requirement for a more sophisticated interpolation model to obtain accurate, smooth drag gradients.

To test the drag map interpolation method used by the nacelle drag prediction RSM, the CFD nacelle drag data at standard resolution in Fig. 9b was interpolated to the operating conditions which had been sampled in the high-resolution drag map (Fig. 9a). For comparison, this was done with an ordinary Kriging (OK) process, in addition to the method described in Section 2.1.2.

Fig. 9a shows the results of the naïve application of OK to model c_D as a function of M and $MFCR$, based on the standard resolution drag map CFD data. This approach differs from the method described in Section 2.1.2 in that it does not use a regression model ("constant regression"), and uses a constant nugget value of 4×10^{-6}, corresponding to the baseline nugget employed by the drag map interpolation method. The resulting interpolated drag map shows that the ordinary Kriging (OK) model produces considerable errors even at low Mach numbers, up to the drag-rise Mach number (Fig. 9c). At the same time, the high-drag portion of the drag map is comparably well reproduced, including the curve for $dc_D/dM = 0.5$. The CFD-derived and OK-derived curves for $dc_D/dM = 0.5$ show a Mach number difference of 0.01 at the maximum $MFCR$, which indicates a less accurate interpolation of drag gradient than achieved by linear interpolation (Fig. 9b). Compared to this, the lines indicating $dc_D/dM = 0.1$ (Eq. (1)) bear almost no resemblance to the original data, as the oscillations in the low-drag range of the drag map cause the gradient of $dc_D/dM = 0.1$ to be exceeded frequently, even at $M < 0.3$. Relative interpolation errors (Fig. 9e) reach or exceed 10% in several places, even at conditions where there is little variation in the drag data, but the amplitude of errors does not increase in the high-drag regions of the drag map. The large errors at low-drag operating conditions are the consequence of the model adapting to the high drag gradients and strong non-linearities in the drag data at Mach numbers above M_{DR}. The ordinary Kriging interpolation does not allow useful predictions of drag or drag gradients at the relevant operating conditions. The fact that higher gradients are represented more accurately by OK confirms that the OK model adjusts to the stronger gradients and larger drag amplitudes found in the data above drag-rise Mach number, to the detriment of accuracy at conditions with small drag gradients.

The dragmap interpolation method described in Section 2.1.2 produces a much more accurate approximation of the high-resolution drag map, based on the standard resolution data (Fig. 9d). The lines for $dc_D/dM = 0.1$ and for $dc_D/dM = 0.5$ are in excellent agreement. The error is small enough that drag-rise Mach numbers (M_{DR}, Eq. (1)) can be determined from the interpolated data with an error of less than 10^{-3}, at $MFCR > 0.55$, and the error for M_{ref} (Eq. (10)) is less than 5×10^{-4}. Both error

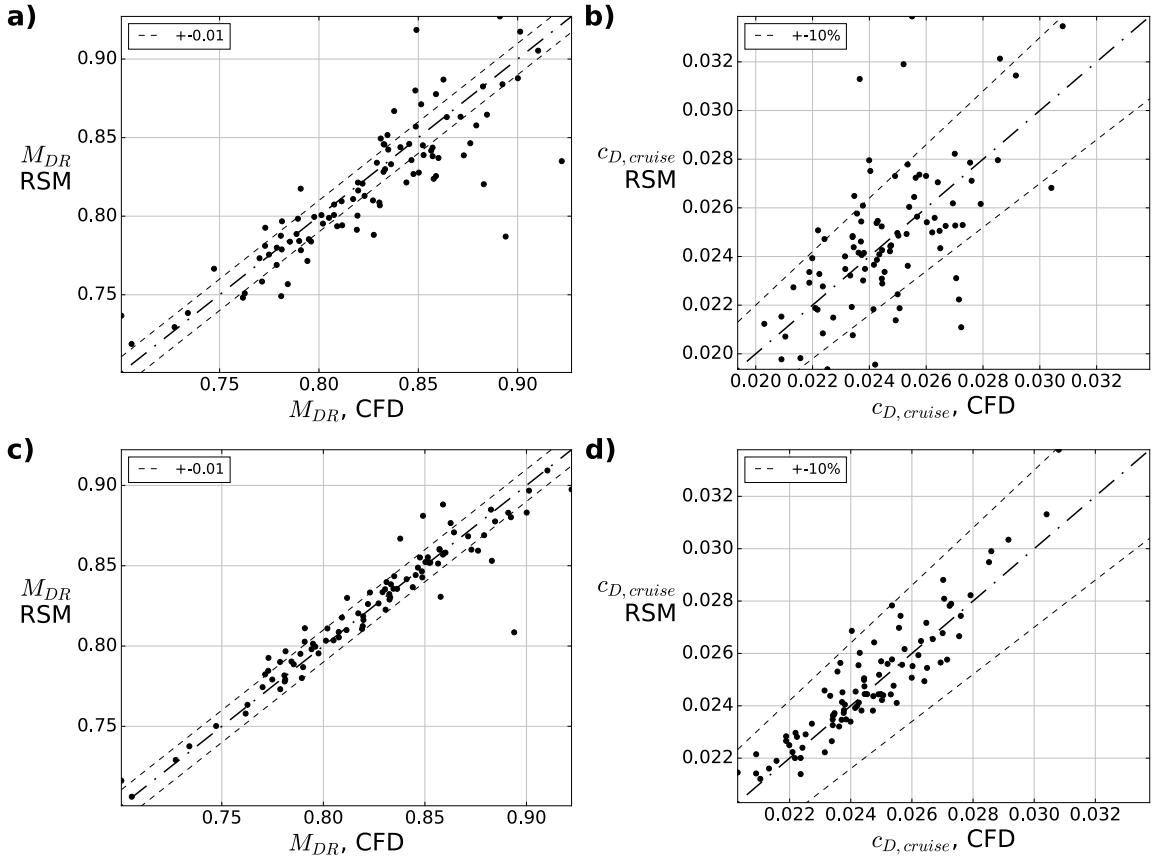


Fig. 10. Validation against an independent data set, predictions from RSM compared to CFD results. (a) Drag-rise Mach number at $MFCR = 0.75$ from a simple Gaussian process model; (b) drag at cruise conditions (Eq. (2)) from a simple Gaussian process model; (c) and (d) use the combined RSM developed in this paper.

margins are significantly lower than with either linear interpolation or OK, and also an order of magnitude below the errors required for the nacelle drag prediction RSM. The differences between the interpolated map (Fig. 9d) and the high-resolution drag map from CFD (Fig. 9a) are below 1% at all conditions where $M < M_{ref}$ (Fig. 9f). The error only exceeds 1% at the highest sampled Mach numbers which have no relevance for the application at hand. This is the intended result of de-prioritizing input data points with very large drag gradients. The small-scale oscillations visible around $M = 0.7$ at the lowest *MFCR*s are not due to the interpolation but originate with the high-resolution CFD, as already explained in the discussion to Fig. 9a and b. In a preliminary design context, it is desirable that the interpolation model produces smooth gradients in place of any small-scale oscillations.

3.2. Interpolation of scaling references

The interpolation models for the scaling references (M_{ref} and $c_{D,ref}$, Section 2.1.3), were tested by leave-one-out (LOO) auto-cross-validation [27]. The RMS of the prediction error was found to be 2.6×10^{-4} for $c_{D,ref}$. With typical values of 0.025 of $c_{D,ref}$, this error is well below 1% of the $c_{D,ref}$ value. This means that the interpolation model for $c_{D,ref}$ does not add any significant amount of uncertainty to the overall method.

The LOO validation for the M_{ref} model found an RMS of 4.7×10^{-3}. For a representative case of $M_{ref} = 0.9$, $M_{DR} = 0.85$, this leads to an error of about 5.5×10^{-3} in the estimate of M_{DR}. This is an acceptable result, and it is significantly better than the uncertainty of nacelle drag-rise prediction in published preliminary methods [29], which is around 0.02.

3.3. Full model validation

3.3.1. Validation dataset and validation criteria

Generating a Gaussian process model for Qc_D requires a notable amount of computational effort, therefore it is impractical to conduct LOO cross-validation for the Qc_D model or for the full RSM. To obtain reliable indicators of the accuracy of the full nacelle drag prediction model (Section 2.2.3), an independent CFD dataset was used. This consisted of 93 designs within the same design space (Table 1), sampled according to a Latin hypercube distribution. These nacelle designs were analysed using the same CFD procedure as the input data for the RSM.

The validation regarded the prediction of two performance metrics which are defined in Section 1.2. The first metric was drag-rise Mach number (M_{DR}, Eq. (1)) at $MFCR = 0.75$. The second metric was drag at the assumed representative cruise condition (Eq. (2)), which is specific to each nacelle design. Each of the two performance metrics was extracted from the CFD validation dataset, using the drag map interpolation method described in Section 2.1.2. The same metrics were then predicted from the full RSM, and compared to the results from the CFD dataset.

3.3.2. Prediction of performance metrics

To compare the results of the RSM model to the output of a more conventional process, the Gaussian process used to predict Qc_D was trained on data of drag coefficient (c_D) directly as a function of M, *MFCR*, and the six design variables. Since the full dataset is too large to train a Gaussian process with the available working memory, this model used the reduced dataset left after applying the filters in Eqs. (14)–(15). This introduces some information from the complete RSM to the simple comparison model, and gives it an advantage over an entirely naïve method with fewer designs and

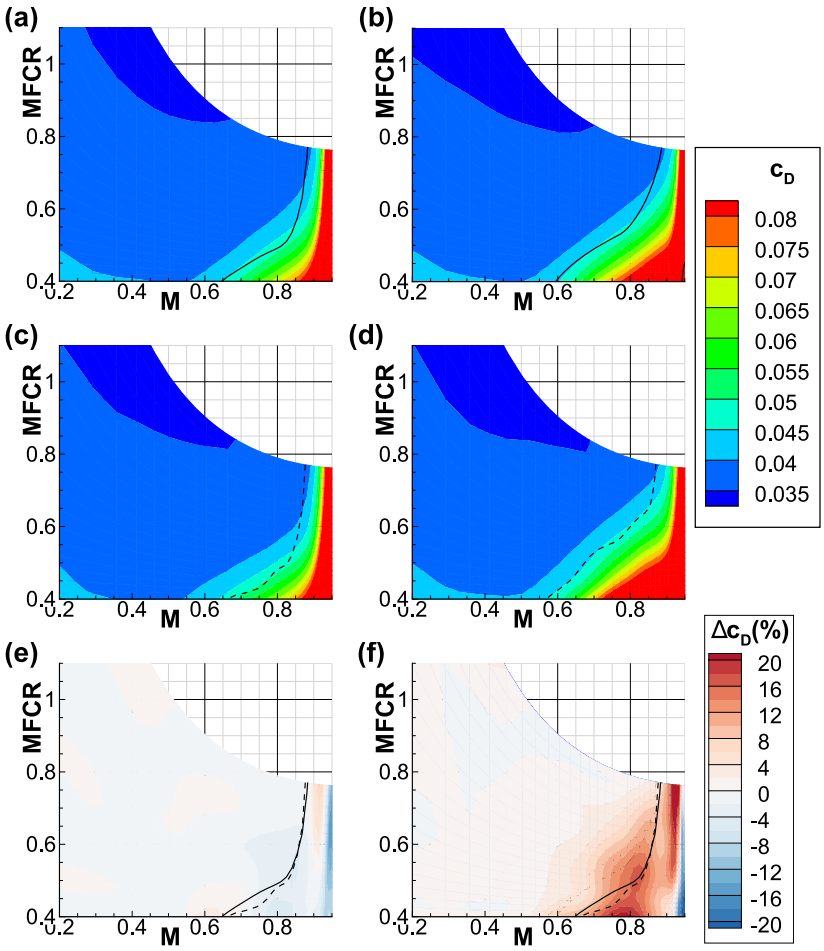


Fig. 11. Comparison of CFD results and predictions from RSM. Lines show M_{DR} as a function of *MFCR* computed from CFD data (solid), and predicted from the RSM (dashed). (a), (b) Drag maps from CFD; (c), (d) drag maps for the same nacelles, from RSM; (e), (f) error of the RSM prediction relative to the CFD results. Positive Δc_D indicate drag over-prediction.

no filtering of undesirable input data. The nugget for this simplified RSM was set to a constant value of 4×10^{-6}, the baseline nugget used for the Qc_D and the drag map interpolation models.

Validation was conducted against the independent validation dataset of 93 nacelle designs (Section 3.3.1). Both the simplified and the full RSM were used to predict the performance metrics M_{DR} and $c_{D,cruise}$ (Eqs. (1), (2)). Fig. 10 shows the results of the validation for both models. To quantify the goodness of fit, the standard error was computed for each predicted variable for each model, according to Eq. (18).

$$\sigma_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_{predicted} - X_{CFD}\right)^2} \tag{18}$$

For the simplified RSM, the standard error for M_{DR} is 0.023, with a considerable number of outliers in the validation graph (Fig. 10a). Within the context of this study, this is an unacceptably large error, as the majority of model evaluations would result in errors larger than the typical design margin of 0.02 between M_{cruise} and M_{DR}. The standard error for $c_{D,cruise}$ (Eq. (2)) is 2.3×10^{-3} or 9.5% of the correct cruise drag (see also Fig. 10b). In contrast to this, the full response surface method presented in this paper achieves a standard error for drag-rise prediction of 0.014, with fewer outliers (Fig. 10c) than observed in the results from the simple model (Fig. 10a) and a standard error for cruise drag prediction of 3.6%, almost a third of the error achieved by the simple Gaussian process. Both results clearly surpass the error margins quoted for estimation of c_D and M_{DR} with the preliminary nacelle design method published by ESDU [29], of 8% on subcritical nacelle drag, 20% on wave drag at cruise and 0.02 on prediction of M_{DR}.

85% of all observed errors on M_{DR} are smaller than 0.014, and in 75% of all cases the error is smaller than 0.01. With a normally distributed error, one would expect to have only 68.1% of all samples within the standard deviation. This suggests that outliers contribute significantly to the standard error, and that in the majority of cases, predictions are already more accurate than the standard error indicates. The performance of the response surface model for predicting cruise drag of nacelles (Fig. 10d) is more than sufficient for a preliminary design application, where average errors of the order of 10% are acceptable. This limit is only exceeded for one of the 93 validation designs.

3.3.3. Prediction of drag maps

To assess the ability to predict the properties of likely nacelle designs, as opposed to the random designs in the validation dataset, and to test the prediction accuracy of drag prediction across different operating conditions, two example nacelle geometries were generated. These nacelles were chosen from the valid design space to have desirable performance characteristics. This was done to evaluate the model performance for cases which could reasonably be the results of a design process, and for which detailed drag prediction is relevant, rather than for arbitrary designs which are of less relevance and also more likely to have hard-to-interpolate extreme drag characteristics, such as shown in Fig. 2c. Fig. 11 shows the difference between the RSM prediction and CFD

results for the two nacelle designs. A comparison of Fig. 11a, b to Fig. 11c, d shows that the general shape of each drag map is correctly predicted by the RSM, including the characteristics of M_{DR} as a function of *MFCR*. No oscillations or systematic errors are noticeable in the predicted drag maps. The error maps in Figs. 11e, f show that the accuracy of drag prediction at subcritical Mach numbers is generally within 1%. At Mach numbers close to M_{DR}, this difference increases but generally stays below 10% at all operating conditions below M_{DR}. The fact that above M_{DR}, the error increases and can reach 20%, is acceptable for preliminary design, and is in part also intended consequence of prioritizing accuracy at cruise condition over high-drag conditions. The fact that the prediction errors shown in Fig. 11e, f change very gradually with operating conditions below M_{DR} means that the RSM is sufficiently consistent to compute drag gradients. The gradients of drag error increase above M_{DR}, but are still small enough around drag-rise to predict M_{DR} with good accuracy, with an error of less than 0.01 for both nacelle designs, for $MFCR > 0.55$.

4. Conclusions

The problem of nacelle drag prediction from RSM presents a unique combination of challenges which prevent the application of standardized methods. One challenge is that the data cannot be sampled isotropically, because not just drag but gradient of drag over Mach number needs to be represented accurately in the sampled input data. This prevents the use of a global Latin hypercube sampling (LHS). Another challenge is that both the input data and the accuracy requirements are inhomogeneous. At conditions of low drag and low drag gradients, modelling must be accurate, while the presence of high drags and large gradients at other conditions needs to be indicated but not modelled accurately. Even for a single nacelle design, an ordinary Kriging approach fails to model the inhomogeneous characteristics of the data. The third challenge is that the model is effectively used as a multivariate predictor since the immediate output, nacelle drag, is used to compute the gradient of drag over Mach number, which is then used to determine the drag-rise Mach number.

The presented method decomposes the problem into several components which are based on Gaussian processes. It exploits self-similarities between different drag maps and orthogonal properties of designs, such as drag-rise Mach number and the drag level at low Mach numbers. These are then interpolated, per design, by separate models. Another strategy that was applied is the consequent use of filtering and nuggets to focus accuracy onto the most relevant parts of the data. This allows prediction of nacelle drag with a standard error of 3.6% at cruise condition, while very sharp drag changes outside of the plausible operating range of a nacelle design are only approximately modelled. At the same time, the model retains the important ability to indicate the range of viable operating conditions. Predictions are consistent enough to obtain smooth first derivatives of nacelle drag with respect to Mach number. This allows prediction of drag-rise Mach number with a standard error of 0.014. The new method is a notable improvement on previously published methods for preliminary nacelle design, which claim standard errors of 0.02 on drag-rise prediction and 8% on the prediction of nacelle drag at cruise condition.

Conflict of interest statement

The authors declare that there is no conflict of interest.

Acknowledgements

The authors would like to thank Christopher Sheaf and Nicolas Grech from Rolls-Royce PLC for providing advice.

This work was co-funded by Innovate UK.

Data statement

Due to confidentiality agreements with commercial research collaborators, supporting data for this study cannot be shared.

References

[1] G. Venturelli, E. Benini, Kriging-assisted design optimization of S-shape supersonic compressor cascades, Aerosp. Sci. Technol. 58 (2016) 275–297.
[2] S. de Lucas, J.M. Vega, A. Velazquez, Surrogate model for viscous drag in aircraft empennage conceptual design, Aerosp. Sci. Technol. 31 (2013) 99–107.
[3] D.G. Krige, A statistical approach to some basic mine valuation problems on the Witwatersrand, J. Chem. Metall. Min. Soc. S. Afr. 53 (5) (1952) 159–162.
[4] A. Forrester, A. Sóbester, A. Keane, Engineering Design Via Surrogate Modelling: A Practical Guide, John Wiley & Sons Ltd, Hoboken, New Jersey, USA, 2008.
[5] J.P.C. Kleijnen, Kriging metamodeling in simulation: a review, Eur. J. Oper. Res. 192 (2009) 707–716.
[6] Z. Han, S. Görtz, R. Zimmermann, Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function, Aerosp. Sci. Technol. 25 (2013) 177–189.
[7] J. Whurr, J. Hart, A Rolls-Royce perspective on concepts and technologies for future green propulsion systems, in: Encyclopedia of Aerospace Engineering, John Wiley & Sons Ltd, Hoboken, New Jersey, USA, 2015, pp. 1–9.
[8] W. Burgsmuller, S. Thome, Close coupling of very high-bypass engines in combination with take-off/landing, Notes Numer. Fluid Mech. 60 (1997) 95–102.
[9] R. Christie, S. Ramirez, D.G. MacManus, Aero-engine installation modelling and the impact on overall flight performance, in: RAeS Applied Aerodynamics Conference: Advanced Aero Concepts, Design and Operations, 2014.
[10] D.L. Daggett, S.T. Brown, R.T. Kawai, Ultra-Efficient Engine Diameter Study, NASA, 2003, NASA/CR-2003-212309.
[11] M.J. Langley, The Design of Axisymmetric Cowls for Podded Nacelles for High By-pass Ratio Turbofan Engines, Aeronautical Research Council Reports and Memoranda, Aircraft Research Association Ltd, Bedford, 1979, R&M No. 3846.
[12] M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 21 (2) (1972) 239–245.
[13] J.C. Helton, F.J. Davis, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, Reliab. Eng. Syst. Saf. 81 (2003) 23–69.
[14] M.Y.M. Ahmed, N. Qin, Metamodels for aerothermodynamic design optimization of hypersonic spiked blunt bodies, Aerosp. Sci. Technol. 14 (2010) 364–376.
[15] A. Heidebrecht, T. Stankowski, D.G. MacManus, Parametric geometry and CFD process for turbofan nacelles, in: Proceedings of ASME 2016 Turbo Expo: Turbomachinary Technical Conference and Exposition, 2016, GT2016-57784.
[16] B.M. Kulfan, J.E. Bussoletti, Fundamental parametric geometry representations for aircraft component shapes, in: 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2006, AIAA2006-6948.
[17] R. Christie, A. Heidebrecht, D.G. MacManus, An automated approach to nacelle parameterization using intuitive class shape transformation curves, J. Eng. Gas Turbines Power 139 (6) (2017), GTP-16-1227.
[18] B. Malouin, et al., Engine pre-entry thrust and standard net thrust evaluation based on the far-field method, Aerosp. Sci. Technol. 45 (2015) 50–59.
[19] ANSYS, User's Guide, 2015.
[20] MIDAP Study Group, AGARDograph No. 237 Guide to In-flight Thrust Measurement of Turbojets, AGARD, 1979, pp. 34–61.
[21] P. Roache, Verification and Validation in Computational Science and Engineering, Hermosa Publishers, Albuquerque, 1998.
[22] M. Robinson, D.G. MacManus, C. Sheaf, Aspects of aero-engine nacelle drag, Proc. Inst. Mech. Eng., G J. Aerosp. Eng. (2018), https://doi.org/10.1177/0954410018765574.
[23] H. Glauert, The effect of compressibility on the lift of an aerofoil, Proc. R. Soc. London CXVIII (1928) 113–119.
[24] S.N. Lophaven, H.B. Nielsen, J. Søndergaard, DACE, A MATLAB Kriging Toolbox, IMM, University of Denmark, Kgs. Lyngby, 2002.
[25] F. Pedregosa, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[26] S. Wessing, Two-Stage Methods for Multimodal Optimization, Dissertation, Technische Universität Dortmund, 2015.
[27] A.I.J. Forrester, A.J. Keane, N.W. Bressloff, Design and analysis of 'noisy' computer experiments, AIAA J. 44 (10) (2006) 2331–2339.
[28] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, The MIT Press, Cambridge, Massachusetts, 2006.
[29] ESDU, Drag of Axisymmetric Cowls at Zero Incidence for Subsonic Mach Numbers, ESDU, London, UK, 1981, ESDU 81024.



https://dspace.org/

Surrogate model of complex non-linear data for preliminary nacelle design

Heidebrecht, Alexander; MacManus, David G.

2018-10-03
Elsevier
https://doi.org/10.1016/j.ast.2018.08.020
Heidebrecht A, MacManus DG. (2019) Surrogate model of complex non-linear data for preliminary nacelle design. Aerospace Science and Technology, Volume 84, January 2019, pp. 399-411

Downloaded from a DSpace repository. https://dspace.org/