This is a preprint of the following article, which is available from http://mdolab.engin.umich.edu
R. P. Liem, C. A. Mader, and J. R. R. A. Martins. Surrogate Models and Mixtures of Experts in Aerodynamic Performance Prediction for Mission Analysis. *Aerospace Science and Technology*, 2015 (to be published).
The published article may differ from this preprint.

Surrogate Models and Mixtures of Experts in Aerodynamic Performance Prediction for Mission Analysis

Rhea P. Liem[1]
University of Toronto Institute for Aerospace Studies, Toronto, ON, Canada
Charles A. Mader[2], Joaquim R. R. A. Martins[3]
University of Michigan, Department of Aerospace Engineering, Ann Arbor, MI

Abstract Accurate aircraft fuel burn evaluation over a complete mission is computationally expensive, as it requires up to millions of aerodynamic performance evaluations. Thus, it is advantageous to use surrogate models as approximations of high-fidelity aerodynamic or aerostructural models. Conventional surrogate models, such as radial basis function and kriging, are insufficient to model these functions accurately, especially in the transonic regime. To address this issue, we explore several ways to improve the accuracy of surrogate models. First, we employ an adaptive sampling algorithm to complement a traditional space-filling algorithm. Second, we improve the kriging surrogate performance by including gradient information in the interpolation (a form of gradient-enhanced kriging—GEK), and by introducing a known trend in the global model component (kriging with a trend). Lastly, we propose a mixture of experts (ME) approach, which is derived based on the divide-and-conquer principle. We validate the developed surrogate models using aerodynamic data for conventional and unconventional aircraft configurations, and assess their performance in predicting the mission ranges by performing analyses on ten mission profiles. Our results show that the proposed ME approach is superior to the traditional models. Using a mixture of GEK models to approximate drag coefficients give us approximation errors of less than 5% with less than 150 samples, whereas the adaptive sampling fails to converge when training a global model. However, when we have a simple function profile, such as the lift and moment coefficients, using a conventional surrogate model is more efficient than an ME model, due to the added computational complexity in the latter. The range estimation errors associated with the ME models are all less than 2% for all the test mission profiles considered, whereas some traditional models yield errors as high as $20\% - 80\%$. We thus conclude that the ME technique is both necessary and sufficient to model the aerodynamic coefficients for surrogate-based mission analysis.

1 Introduction

Fuel efficiency has become an increasingly important metric in aircraft design due to increases in fuel prices and environmental concerns [1, 2]. However, evaluating aircraft fuel burn accurately is not an easy task. Several disciplines contribute to the calculation, including: the aerodynamic performance of the aircraft, the aircraft's weight distribution and the performance of the engines. The calculation is also affected by the speed of the aircraft and the atmospheric conditions at the altitude where the aircraft is flying. To account for the coupling in such a multidisciplinary system, multidisciplinary design optimization (MDO) should be used as it can automatically perform the optimal interdisciplinary trade-offs [3]. While effective, MDO frameworks can be computationally expensive. Completing such a computation in an optimization process (which requires many iterations, prior to reaching optimality) using pure physics-based models quickly becomes computationally intractable. The most common approaches for reducing the cost of these fuel-burn computations involve a simplification of either the physics in the model or in the mission profile considered. The classical Breguet range equation is a popular example of such an approach [4, 5, 6]. Kenway and Martins [7] used this equation to analyze the aircraft performance at each operating point in multipoint high-fidelity aerostructural optimization problems to minimize fuel burn and takeoff gross weight. The multipoint objective is the weighted combination of the objective functions evaluated at five operating points (perturbations of the nominal cruise

condition), assuming an equal weight for each point. Other simplified models include using fuel fractions to represent the individual segment fuel burn values [6] or using either simplified analytical or empirical models [8] to represent the physics. These simplifications and assumptions reduce the computational time, albeit at the expense of accuracy and generality.

Recent work has shown that surrogate models can significantly reduce the computational cost of performing a detailed fuel-burn computation in a design optimization setting. Surrogate models, or metamodels, are commonly used as simpler approximations of the physical systems to reduce the cost of computationally intensive analysis and optimization tasks [9, 10, 11]. Surrogate models have previously been shown to assist various optimization procedures in aerospace engineering. Chung and Alonso [12, 13] used a gradient-enhanced kriging method in a supersonic business jet design optimization, Toal and Keane [14] used a cokriging method to perform a multipoint drag minimization, Zimmermann and Görtz [15] developed and used a POD-subspace restricted least squares model for solving the governing fluid flow equations, and Amsallem *et al.* [16] performed offline precomputations to construct fluid reduced order bases (ROB) and structural reduced order models (ROM) database for aeroelastic computations. Fossati and Habashi [17] employed a reduced order modeling approach, based on proper orthogonal decomposition (POD) and kriging interpolation, to reduce the computational cost in steady and unsteady three-dimensional viscous turbulent aero-icing simulations. In the context of mission analysis, Koko [18] used a Lagrangian interpolation as a surrogate to model the aerodynamic forces at different points along the flight mission of interest in a trajectory optimization problem aiming to minimize fuel consumption of morphing wingtip devices.

The authors have previously used kriging models to approximate the aerodynamic data required in a detailed mission analysis procedure, to give an accurate estimation of the amount of fuel burned during a mission [19, 20]. This surrogate-based mission analysis approach significantly reduces the required number of aerodynamic performance evaluations from millions to the number of samples required to build the kriging models, thus enabling the integration of mission analysis in aerostructural optimization cases. Using this procedure, a new strategy was derived to formulate multipoint design aerostructural optimization problems to maximize the aircraft performance over a large number of different missions [19]. This strategy was demonstrated in a fuel burn minimization problem for a long-range wide-body aircraft configuration, where only the cruise portion was modeled in detail. In this multipoint optimization strategy, the required number of high-fidelity aerostructural solutions at each optimization iteration is reduced from millions to 25. A similar approach was demonstrated in a direct operating cost (DOC) minimization problem for a 100-passenger regional jet configurations [20]. In this DOC minimization problem, a shorter range mission was considered. This necessitated the inclusion of the climb and descent segments in the mission, as the cruise segment was no longer the only dominant mission segment.

Expanding the input space of the surrogate model to include the flight conditions involved in climb and descent makes training the surrogate model significantly more challenging. In addition to requiring a larger input space, the model needs to capture the high drag gradient region outside the cruise regime, which causes problems for some of the simpler surrogate modeling techniques. Handling unconventional configurations such as a blended-wing-body (BWB) configuration, presents a similar challenge. In these cases, the challenge comes from a higher degree of correlation between drag and trim, causing more nonlinearity in the drag profile with respect to the tail angle variable [21]. These are the challenges that have motivated this work, where we develop a surrogate modeling technique that is sufficiently general to handle the full range of flight conditions and aircraft configurations that may be of interest to an aircraft designer. Specifically, we explore and analyze the performance of various surrogate models in the context of performing surrogate-based mission analysis. Based on our specific requirements, which will be discussed in Section 3, we have limited our selection of surrogate models to kriging and radial basis function (RBF) models. Several variants of kriging models are considered, in particular those that allow the incorporation of "extra knowledge" to further fine tune the surrogate models.

The first is the gradient-enhanced kriging (GEK) model, which incorporates gradient information at sample points so the surrogate model can have better approximations of the curvature around the sample points. GEK is a very well-established technique and has been shown to improve kriging performance; see, [12, 13, 22, 23, 24], for example, for some aerospace applications of GEK.

Second, we consider the "kriging with a trend" model, where we specify the basis functions for the global model of kriging [25]. Instead of using the commonly used low-order polynomials, we select the basis functions based on the system physics, e.g., by setting a quadratic trend in a certain direction. This second approach has been demonstrated in a previous work by the authors [20].

Third, we propose to use multiple surrogate models in the input space, instead of just a single global model. The main rationale is that we let each local surrogate model perform well in a smaller subset of the input space, instead

of forcing one model to approximate the entire problem domain, which might have contrasting profiles in the input space (e.g., when the function profile in one region is more nonlinear than the others). We adopt the mixture of experts approach [26] which uses a cluster-based preprocessing step first proposed by Tang *et al.* [27]. In this approach, the problem domain is first partitioned into several subregions by using clustering algorithms, followed by local expert training within each subregion. In this case, the local experts are surrogate models. The local predictions are then combined probabilistically to yield the final prediction.

In this work, we compare the performance of these surrogate models in approximating the aerodynamic lift, moment and drag coefficients of two Boeing 777-size aircraft configurations: one conventional and one unconventional. We then assess the amount of error that this introduces into the estimated values of range across ten benchmark missions to assess how well the various techniques work for the surrogate-based mission analysis.

A part of the work presented in this paper was presented as a conference paper [28]. We start the remainder of this paper by describing the surrogate-based mission analysis procedure in Section 2. In Section 3, we first discuss the surrogate modeling classification, to select the techniques that are suitable for our purpose. We then explain the details of the selected techniques, namely kriging and RBF models, and their comparison. Our proposed *mixture of experts* model is presented in Section 4. The description of our case studies are given in Section 5. We then discuss our results and findings in Section 6, followed by the conclusion in Section 7.

2 Surrogate-based Mission Analysis

The classical Breguet range equation is commonly used to compute the amount of fuel required to fly a given range [4, 5, 6]. This widely used range equation was derived and published independently in 1920 by Coffin [29] and later in 1923 by Breguet [30]. This equation has since become a basic model describing the physics of aircraft, encompassing the three dominant disciplines within an aircraft system: engine (by the thrust specific fuel consumption, or TSFC), aerodynamics (by the lift to drag ratio, L/D), and structural technologies (by the structural weight). This equation, however, is only applicable under the assumption that the product of the inverse of TSFC, L/D, and flight speed are constant. One important implication of this is that the takeoff, climb, and descent segments are not properly modeled by this equation [31].

Simple fuel fractions (the ratio of the aircraft total weight at the end of a flight segment to the weight at the start of the same segment) are typically used to compute the amount of fuel burned in flight segments other than cruise. See, for example, Roskam [6] for values of suggested fuel-fractions corresponding to several mission phases for various aircraft types. Lee and Chatterji [32] presented the approximation functions for total fuel burn in climb, cruise, and descent phases. To compute fuel burn during climb, they applied a climb fuel increment factor, which was defined as the additional fuel required to climb the same distance as it was for cruise, normalized with respect to the takeoff weight [33].

Henderson *et al.* [34] presented an object-oriented aircraft conceptual design toolbox, `pyACDT`, which analyzed a given mission profile to estimate the mission fuel burn and point performance parameters. The Breguet range equation was used to calculate the cruise range. This toolbox uses a potential flow panel method to predict the aerodynamic performance. The Program for Aircraft Synthesis Studies (PASS) is a conceptual design tool that evaluates all aspects of mission performance [35]. This software package can incorporate several analyses, including linear aerodynamic models for lift and inviscid drag, sonic boom prediction for supersonic cases, weight and center of gravity estimation, and full mission analysis. These rapid analyses are coupled with optimization tools (gradient or non-gradient based) to perform aircraft design optimizations.

The fuel-burn computations mentioned above are done with simplifications of the aircraft performance and mission profile, which can reduce the accuracy of the predicted total aircraft fuel burn. For example, the constant product of L/D, inverse of TSFC, and flight speed assumed in the Breguet range equation do not reflect the actual aircraft operation, as their values vary across the flight operating points in the mission profile. Moreover, most fuel-burn computations focus on the cruise portion, which is critical for long-range missions, but not necessarily so for shorter-range missions. For shorter-range missions, the climb segments contribute significantly to the total fuel consumption. For a more accurate fuel-burn computation that is suitable for both short-range and long-range flight missions, performing a detailed mission analysis that include all phases in the mission profile is thus necessary. Instead of using the Brequet range equation, the range now needs to be evaluated via a numerical integration procedure. However, performing such a detailed mission analysis with physics-based models is computationally expensive due to the many performance evaluations required in the procedure. The computational issue is further exacerbated when we use the mission analysis in optimizations, or uncertainty quantification (e.g., using the Monte Carlo method), both of which require many

function evaluations.

We now describe the mission analysis procedure to compute the fuel weight W_{fuel}, range R, and time t, by numerically integrating a given mission profile. This mission analysis procedure has also been used in the previous work by the authors [20]. As inputs we have the mission profile parameters (such as altitude and Mach number for cruise segments; flight speed, initial and final altitudes for climb and descent segments), the initial takeoff weight, and the final zero-fuel weight (ZFW) for each mission.

The weight, mission segment range and time are then solved iteratively using an *all-at-once* approach. Using this approach, a set of residual equations, $\mathcal{R}$, is set up using the endpoint weights of each segment as states. To set up the residual equations, we need to match the endpoint weights of two adjacent segments: $W_{f_j} - W_{i_{j+1}} = 0$, where $j = 1, \ldots, N_{\text{seg}}$ denotes the segment index; W_i and W_f denote the segment's initial and final weight, respectively. Similarly, at the boundaries, $W_{i_1} = W_{\text{TO}}$ and $W_{f_{N_{\text{seg}}}} = W_{\text{ZF}}$, where W_{TO} and W_{ZF} refer to the takeoff and zero-fuel weights. The residual is set to zero to determine the characteristics for the entire mission profile, which is solved using the Newton–Krylov algorithm. This forces the weights of the various segments to be consistent with each other, providing a valid and continuous mission profile. Each segment can then be analyzed independently, based only on the current states of the system, until the residual equations are solved.

The amount of fuel burned during startup, taxi, takeoff, and landing is computed using the fuel fraction method, where $W_f = (1 - \zeta) W_i$, where ζ is the fuel fraction. The numerical integration to compute the fuel burn for the climb, cruise, and descent segments is derived from the range equation. Since TSFC (c_T) is the weight of fuel burned per unit time per unit thrust, we can compute the rate of reduction of aircraft weight as $dW/dt = -c_T T$, where W and T denote aircraft weight and thrust, respectively. Using this relation and the generic integral equation for range, $R = \int_{t_i}^{t_f} V dt$, the numerical integrations for range are given below. The subscripts i and f in the integration limits correspond to the initial and final values, respectively. For the cruise segment, the integration is done with respect to weight,

$$R = \int_{W_i}^{W_f} -\frac{V}{c_T T} dW. \tag{1}$$

For the climb and descent segments, the range equation is integrated over the change in altitude,

$$R = \int_{h_i}^{h_f} \frac{V \cos\gamma}{RC} dh, \tag{2}$$

where h and γ denote the altitude and the flight path angle. The rate of climb, RC, is derived from the equation of motion, $T_{\text{av}} \cos(\phi_T + \alpha) - D - W \sin\gamma = (W/g)(dV/dt)$, and that $RC = V \sin\gamma$. The symbol T_{av} denotes the available thrust, D denotes drag, and g is the gravitational acceleration. The thrust inclination angle is denoted by ϕ_T (typically assumed to be zero [36]), and α refers to the angle of attack. With small angle approximations, this equation yields

$$RC = \frac{(T_{\text{av}} - D) V}{W \left(1 + \frac{V}{g} \frac{dV}{dh}\right)}. \tag{3}$$

We have the information of flight speed and altitude for each segment interval from the mission specification. TSFC is a property of the aircraft engine, which can be estimated using an engine model or assumed constant. We then need to compute T to evaluate (1) and (2), which we can find once we know D. We can evaluate drag upon determining the angle of attack, α, and the tail rotation angle, η, that satisfy the lift (e.g., level flight, $L = W$, for cruise) and trim ($C_M = 0$) constraints simultaneously. These two angles are found using a Newton algorithm.

This procedure computes the mission range given the fuel weight ($W_{\text{TO}} - W_{\text{ZF}}$). When the mission range is specified, we perform a secant algorithm to find the corresponding fuel weight, W_{fuel}. Following this procedure, the required number of aerodynamic performance evaluations would be equal to the product of the number of missions, number of secant iterations, number of iterations to solve the residual equations, number of integration intervals, and number of Newton iterations to solve for the angles. Therefore, a mission analysis would require millions of aerodynamic solutions, which would be computationally prohibitive. For this reason, surrogate models are built to approximate the aerodynamic force and moment coefficients (C_L, C_D, and C_M) to be used in the mission analysis. When a sample-based surrogate modeling technique is used, the required number of aerodynamic performance evaluation calls is reduced to the number of samples used to build the surrogates, making the procedure computationally tractable.

The surrogate-based mission analysis procedure described here allows us to perform mission optimizations, where we set some parameters (e.g., cruise Mach number and altitude) as design variables; aerostructural optimizations

(e.g., to minimize fuel burn or DOC) with an accurate fuel burn computation; and coupled mission and aerostructural optimizations. Liem *et al.* [20] used this procedure in a DOC minimization problem for a regional jet aircraft configuration, by first optimizing the mission profiles, and then used these optimized profiles in an aerostructural optimization problem. Coupling high-fidelity aerostructural optimization framework [7, 37, 38, 39, 40, 41, 42] with this mission analysis procedure can lead to a more realistic aircraft design process. For example, we can further extend the multipoint high-fidelity aerostructural optimization strategy [19] to include detailed analyses during the climb and descent segments, in addition to cruise segment. To obtain meaningful results from these optimizations, however, we first need to have reliable and accurate surrogate models.

3 Surrogate Modeling

A surrogate model uses mathematical models to provide a simpler approximation of a physical system, thereby reducing the computational expenses of analyses and optimizations [9, 10, 11]. Essentially, surrogate models are used as low-cost substitutes to replace expensive evaluations when the original physics-based models are used in any computational-intensive tasks (e.g., analysis or optimization) [43]. These approximation models are also known as *metamodels* [9], or *models of models* [44, 45]. Before we can use a surrogate model to perform surrogate-based mission analyses, it is imperative to have reliable surrogate models that can approximate the aircraft performance over the entire flight operating regime (from takeoff to landing) of the various mission profiles considered. For our purposes, the modeling techniques also need to be flexible enough to be used with different aircraft configurations (conventional and unconventional), for both short and long range missions. To select suitable surrogate models for our surrogate-based mission analysis procedure, we briefly discuss the available surrogate modeling techniques.

Eldred *et al.* [46] classified surrogate models into three categories: data-fits, reduced-order models [47], and hierarchical models [48, 49]. Reduced-order and hierarchical surrogate models can be classified as *physics-based* approaches, since they exploit and simplify the governing equations [50]. These models are thus considered as *intrusive* methods. Data-fit surrogate models, on the other hand, belong to the *black-box* approach category, where the derivations are only based on the inputs and outputs of the high-fidelity models, without necessarily knowing the underlying governing equations. Black-box approaches are *non-intrusive* and typically approximate a function at a point in the N_d-dimensional input space $\mathbf{x}_0 \in \mathbb{R}^{N_d}$ based on the available N_s sample information, including the sample locations $\mathbf{x}_s \in \mathbb{R}^{N_s \times N_d}$ and values $\mathbf{y}_s \in \mathbb{R}^{N_s}$,

$$y\left(\mathbf{x}_0\right) \approx \hat{y}\left(\mathbf{x}_0, \mathbf{x}_s, \mathbf{y}_s, \boldsymbol{\alpha}\right). \tag{4}$$

The symbol $\boldsymbol{\alpha}$ denotes a vector of model parameters, i.e., the undetermined coefficients that are typically derived based on the available training sample set.

Black-box models can be categorized into *regression* and *interpolation* models. Regression models are derived in a least-squares sense, so they are more suitable to approximate functions with inherent random error components, such as measurement data. Interpolation models, on the other hand, reproduce the function values exactly at sample locations. These models are thus suitable to model deterministic computer experiments, where repeated experiments with the same input settings return exactly the same outputs [51]. Two widely used interpolation surrogate models are kriging and RBF models.

To construct and use black-box surrogate models, we first need to generate data (samples), select model structure, and estimate model parameters. The model is then assessed to evaluate the goodness of fit (for regression models) and approximation accuracy at untested data. The sampling and model assessment procedures are presented next. The model structure selection and parameter estimation are model dependent; they will be discussed when kriging and RBF models are described.

3.1 Sampling Methods

Surveys of various *sampling plans*, also referred to as a *design of experiments*, are give by Simpson *et al.* [52], as well as Wang and Shan [53]. When building surrogate models on unknown landscapes, a sampling plan that is uniform, irregular, and space-filling is favorable [54]. The random Monte Carlo simulation (MCS) method is a popular choice in industry, mainly due to its simplicity [53]. Another popular choice is the Latin hypercube sampling (LHS) [55], as its projections onto each variable axis are uniform. Since there are no specific guidelines to determine the "appropriate" sampling size *a priori*, sequential and adaptive sampling plans have become more popular recently. The new points (*infill points*) are selected based on some *infill criteria* to improve the model's predictive capability. There are two main categories for the infill criteria: *exploitation* and *exploration* [54]. Exploitation criteria are used mostly in surrogate-based optimization (i.e., when surrogate models are used to approximate the objective function),

to help finding the optimum point. Some examples include the minimizing the predictor approach and the trust-region method. Exploration criteria aim to "fill the gaps" between existing sample points to ensure that the samples are evenly distributed spatially. This category consists of sequential space-filling sampling plans such as Sobol' [56] and Halton sampling sequence [57], as well as an adaptive approach that locates infill points with the highest estimated error (e.g., using the kriging variance as a metric). In general, maximizing variance when adding samples tends to maximize the inter-site distances (*D-optimality*) [58]. In this work we use a Halton sequence sampling method and an adaptive sampling procedure based on an exploration criterion, which is described in Section 5.3.

3.2 Surrogate-Model Validation Methods

We need to validate the models before using them as surrogates in computationally-intensive analyses and optimizations. One popular approach used to validate surrogate models is the cross-validation method, an overview of which can be found in Meckesheimer *et al.* [45]. In the p-fold cross validation approach, the sample set is first divided into p subsets. Then, we reconstruct the metamodel p times, by omitting one of the subsets each time, to compute the approximation errors. When each subset contains only one sample point, this procedure is called the *leave-one-out* cross-validation [59]. However, the cross-validation approach tends to be biased towards over-represented regions. Due to this limitation, a more reliable model validation approach that employs additional test points to compute the approximation errors is preferred [53]. One of the most commonly used error measures is the root mean square error (RMSE),

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2}, \tag{5}$$

where m denotes the number of validation (test) points. The normalized RMSE is also often used, especially when the function value has large variations within the input space of interests. In this error measure, each error component, $(y_i - \hat{y}_i)$, is normalized with respect to its actual value, y_i, before computing RMSE, to yield the relative approximation error. In this work, we use the normalized RMSE to assess our surrogate models.

3.3 Surrogate-Modeling Techniques

In this work, we want to use surrogate models to approximate the aerodynamic force and moment coefficients with data obtained from solving physics-based aerodynamic models, which are deterministic computer experiments. Thus, we consider only interpolative models, i.e., kriging and RBF models. These two models are described in more details below.

3.3.1 Radial Basis Function Model

RBF is a black-box surrogate model which emulates complicated design landscapes using a weighted sum of simple functions,

$$\hat{y}\left(\mathbf{x}_0, \mathbf{x}_s, \mathbf{y}_s, \boldsymbol{\alpha}\right) = \Psi_0^T \mathbf{w} = \sum_{i=1}^{N_c} w_i \psi\left(\|\mathbf{x}_0 - \mathbf{c}_i\|\right) \tag{6}$$

where the function $\psi\left(\|\cdot\|\right)$ is the kernel function centered at $\mathbf{c}_i$. The norm $\|\cdot\|$ is the Euclidean distance. Typically, the training sample points are used as the centers, thus $\mathbf{c} = \mathbf{x}_s$ and $N_c = N_s$. The vector of unknown coefficients, $\mathbf{w}$, is determined by solving the following system of linear equations,

$$\Psi \mathbf{w} = \mathbf{y}_s, \tag{7}$$

where Ψ is the *gram matrix*, defined as $\Psi_{ij} = \psi\left(\left\|\mathbf{x}_{s_i} - \mathbf{x}_{s_j}\right\|\right)$. Thus, the gram matrix is the the kernel function evaluated at the Euclidean distance between the i^{th} and j^{th} samples. In this work, we use the thin plate splines, cubic, and square-exponential (Gaussian) kernel functions.

3.3.2 Kriging Model

The kriging surrogate model was initially developed in the field of geostatistics by Danie G. Krige (after whom the method is named) [60]. The term "kriging" was first coined by Matheron [61], who was also the first to formulate kriging mathematically. When first derived in the geostatistics field, kriging was used to model continuous and uniquely defined functions relating numbers (e.g., measurement data) to a domain of geographic coordinates (in one-, two-,

or three-dimensional domains) [62]. The foundation of using kriging models in the design and analysis of computer experiments (DACE) was first developed by Sacks *et al.* [63], where points in the input space are analogous to the spatial (geographical) coordinates.

In kriging models, we assume that the deterministic response $y(\mathbf{x})$ is a realization of a stochastic process $Y(\mathbf{x})$ [63, 64],

$$Y(\mathbf{x}) = \sum_{k=1}^{N_f} f_k(\mathbf{x})\beta_k + Z(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}). \tag{8}$$

The first term is the global model component, where $\mathbf{f}(\mathbf{x}) = \left[f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_{N_f}(\mathbf{x})\right]^T$ is a vector of N_f basis functions and $\boldsymbol{\beta} = \left[\beta_1, \beta_2, \ldots, \beta_{N_f}\right]$ is a vector of the unknown coefficients. The stochastic component $Z(\mathbf{x})$ is treated as the realization of a stationary Gaussian random function with zero expected value, $\mathrm{E}[Z(\mathbf{x})] = 0$, and covariance

$$\mathrm{Cov}\left[Z(\mathbf{x}_i), Z(\mathbf{x}_j)\right] = \sigma^2 R(\mathbf{x}_i, \mathbf{x}_j), \tag{9}$$

where $R(\cdot)$ denotes the correlation function with $R(0) = 1$. Therefore, kriging models give exact prediction at sample points, with increasing error variance as we go further from these sample points. In other words, in kriging models the data are assumed to be exact but the function is a realization of a Gaussian process [65]. This second term is called the *localized deviation* [52], *bias*, or *systematic departure* from the linear model [64]. A *stationary* correlation function is typically assumed in kriging models, where the correlation between any two points in the input space, $y(\mathbf{x}_i)$ and $y(\mathbf{x}_j)$, depends only on the difference vector $\Delta\mathbf{x} = \mathbf{x}_i - \mathbf{x}_j$, thus $R(\mathbf{x}_i, \mathbf{x}_j) = R(\mathbf{x}_i - \mathbf{x}_j)$.

For higher-dimensional problems, the correlation function in a kriging model typically satisfies the *product correlation rule*, where the correlation function can be expressed as a product of stationary, one-dimensional correlations,

$$R_{ij}(\boldsymbol{\theta}, d) = \prod_{k=1}^{N_d} R\left(\theta^{(k)}, d_{ij}^{(k)}\right). \tag{10}$$

The vector of correlation parameters is denoted as $\boldsymbol{\theta} = \left\{\theta^{(k)}\right\}$, $k = 1, \ldots N_d$. The notation $d_{ij}^{(k)}$ is the distance between two points in the k^{th} dimension, $\left|x_i^{(k)} - x_j^{(k)}\right|$. These correlation parameters (kriging hyperparameters) are also referred to as *length scales* or *distance weights*, and are typically found via the maximum likelihood estimation (MLE) approach. Large θ values correspond to weak spatial correlation, whereas small values correspond to strong spatial correlation [66]. When each variable has a distinct physical meaning, it makes sense to use an anisotropic correlation function, i.e., having different $\theta^{(k)}$ values in different dimensions. In that case, we have more flexibility in the modeling, but at the expense of a more complex MLE [58, 67]. We will use the Gaussian and cubic spline correlation functions. These two correlation functions exhibit a parabolic behavor near the origin ($R(\theta, d) \propto d^2$ for small d) [67], which are suitable for continuously differentiable functions.

When the global model is assumed known, kriging models produce the *best linear predictor* (BLP). When the known global model is a constant, we have a *simple kriging* [68]. On the other hand, when the global model is unknown and thus needs to be derived, the model is referred to as the *best linear unbiased predictor* (BLUP). When a constant global model is assumed, a BLUP model is called *ordinary kriging*, whereas when a set of basis functions is used (typically low-order polynomials, e.g., linear or quadratic), it is referred to as *universal kriging* [68] or *kriging with a trend* [25]. Ordinary kriging models are more popular and commonly used, as the *a priori* knowledge of the trends in the data is typically unknown [54].

The kriging equation can be derived via the mean square error (MSE) minimization approach, with the unbiasedness constraint [63]. Another alternative for the derivation is the Bayesian approach [58, 64, 69, 70, 71, 72]. In the derivation, the model assumes that the prior variance σ^2, the family and parameters of the correlation function, $R(\cdot)$, are known. Typically, the designers will determine the correlation function, and apply the *empirical Bayes* approach to find the parameters to be most *consistent* with the observed data [58, 64], in particular by employing the MLE. Some correlation functions have tunable parameters $\boldsymbol{\theta}$ that still need to be determined. Since there is no closed-form solution for these optimum parameters, $\boldsymbol{\theta}$ is typically solved by performing a constrained iterative search. The equation for kriging approximation at a test point $\mathbf{x}_0$, $\hat{y}(\mathbf{x}_0)$, can generally be expressed as,

$$\hat{y}(\mathbf{x}_0) = \mathbf{f}^T(\mathbf{x}_0)\hat{\boldsymbol{\beta}} + \mathbf{r}^T(\mathbf{x}_s, \mathbf{x}_0)\mathbf{R}_s^{-1}\left(\mathbf{y}_s - \mathbf{F}_s\hat{\boldsymbol{\beta}}\right), \tag{11}$$

where $\mathbf{f}$ and $\boldsymbol{\beta}$ are as previously defined, $\mathbf{r}(\mathbf{x}_s, \mathbf{x}_0)$ denotes the correlation vector between the test and sample points and $\mathbf{R}_s$ denotes the correlation matrix between samples. The symbol $\mathbf{y}_s$ refers to the vector of function values at sample locations, and $\mathbf{F}_s$ is a matrix containing the basis functions evaluated at each sample location.

A gradient-enhanced kriging model (GEK) interpolates gradient information, in addition to the function value, at each sample location, thus achieving a first-order-consistency requirement, in addition to the zeroth-order-consistency achieved by gradient-free kriging [22]. Depending on how the gradient information is used, there are two types of GEK, namely the indirect GEK and direct GEK. The former uses the gradient information to generate new samples around the available samples via a Taylor series expansion around those samples,

$$y(\mathbf{x}_{N_s+ik}) = y(\mathbf{x}_i) + \frac{\partial y(\mathbf{x}_i)}{\partial x^{(k)}} \Delta x^{(k)}. \tag{12}$$

In the direct GEK approach, the gradients are now directly included in the formulation as additional observations or sample data, as shown below,

$$\mathbf{y}_s = \left[y(\mathbf{x}_1), y(\mathbf{x}_2), \ldots, y(\mathbf{x}_{N_s}), \frac{\partial y(\mathbf{x_1})}{\partial x^{(1)}}, \frac{\partial y(\mathbf{x_1})}{\partial x^{(2)}}, \ldots, \frac{\partial y(\mathbf{x_1})}{\partial x^{(N_d)}}, \frac{\partial y(\mathbf{x_2})}{\partial x^{(1)}}, \ldots, \frac{\partial y(\mathbf{x}_{N_s})}{\partial x^{(N_d)}} \right] \tag{13}$$

This method requires augmenting the correlation matrix with its derivative terms, which considerably increases the order of the correlation matrix to $N_s(N_d + 1)$ from kriging's N_s. Consequently, the computational cost to build and use GEK models is higher than that of the original kriging models. See, for example, [22, 73, 74], for more details on the formulations and uses of GEK.

3.3.3 Kriging Compared to RBF

There is no clear consensus as to which of the kriging and RBF surrogate models has a better predictive performance. Mathematically, an ordinary kriging model can be reduced to RBF with an offset [75]. Wang and Shan [53] claimed that RBF is a compromise between kriging models and polynomial regressions, as it can interpolate the sample points (generally more accurate than polynomial regressions) and at the same time easier to construct than kriging models. Forrester and Keane [54] argued that kriging is the least assuming method, which provides a greater flexibility in the modeling. The flexibility comes mainly due to the parameters in the covariance function; however, it comes at the expense of the estimation of hyperparameters [75]. Jin *et al.* [44], on the other hand, concluded that RBF has the best performance overall in terms of accuracy, robustness (the most robust model is the one that is the least problem-dependent), efficiency (the amount of computational effort required for the surrogate model construction), transparency (the capability to provide information on model sensitivity to input variables and the inter-variable interactions), and conceptual sensitivity (ease of implementation). The comparison was performed with 13 analytical problems and one vehicle handling problem, with varying non-linearity, scale (dimensionality), and smoothness. With the varying opinions regarding the two models, it is safe to conclude that their predictive performance is essentially problem-dependent.

4 Mixture of Experts

There has been a growing interest in using multiple surrogates instead of a single model in isolation [65]. Combining models in some way has been shown to improve the approximation performance of the surrogates [76, 77]. One of the main motivations of using multiple models is to overcome the limited modeling flexibility of using one global model when there is heterogeneity in the function profile [78]. Black-box surrogate model training seeks to find a set of model parameters that fit the observations over the entire input space, which might be inadequate when the function complexity is input dependent. In this section, we will first present an overview of the mixture of experts approach, followed by the description of our proposed approach.

4.1 Overview of Mixture of Experts

Using multiple surrogate models can be done by combination or selection. For example, in *committees* we take the average of predictions from different trained models. Viana *et al.* [65] proposed using cross-validation error in both model selection and combination. When multiple surrogates are present, we can either select one with the lowest cross-validation error, or to use the cross-validation errors to create a weighted surrogate by minimizing the integrated square error. *Decision tree models* use a sequence of binary selections to select one model as the predictor (a winner-take-all strategy) [79]. *Bayesian model averaging* seeks to find one model among several models, by assessing the posterior probability of each candidate [80].

The divide-and-conquer approach, which seeks to solve a complex problem by dividing it into simpler problems, has become increasingly popular in combining multiple models for regression and classification [81]. Some classical examples include the classification and regression tree (CART) algorithm [82], the multivariate adaptive regression splines (MARS) algorithm [83], and the iterative dichotomiser 3 (ID3) algorithm [84]. The CART and ID3 algorithms are derivations of the decision tree models. The mixture of experts (ME) model offers a statistical approach of decision tree modeling, by using a probabilistic framework to combine models [77]. Jacobs *et al.* [26] introduced the original ME model, which adopts the divide-and-conquer strategy. This model relies on three main components, namely the experts, a gating function that facilitates soft splits of data (allowing data to lie simultaneously in multiple regions), and a probabilistic model to combine the experts and gating function. In ME, the experts can either be classifiers or regression functions. When used in regression, ME performs better than other techniques in modeling nonstationary, piecewise continuous data [77]. Bettebghor *et al.* [85] used an ME approach as a surrogate for a discontinuous problem domain in a structural optimization problem. For a comprehensive overview of the ME methods, including the development, advances, and applications, readers are referred to a survey presented by Yuksel *et al.* [77].

In the classical ME models described above, the partitioning and learning of the problem domain are based on the same algorithm. Tang *et al.* [27] proposed another approach, which relies on a cluster-based preprocessing step, therefore separating the partitioning and learning processes. Tang *et al.* [27] used the Kohonen's self-organizing feature map (SOM) to partition the input data as a preprocessing step in their ME approach. Xing and Hu [86] initialized their ME model using K-means. Nguyen–Tuong *et al.* [87] used a distance-based measure in partitioning the training data, assuming the same kernel width for all local kriging models. Bettebghor *et al.* [85] implemented the Gaussian mixture model (GMM) with conjoint data to partition their input space and derive the mixing proportion. This new approach is typically referred to as the mixture of explicitly localised experts (MELE), whereas the classical ME model is categorized as the mixture of implicitly localised experts (MILE). A thorough literature survey of different ME methods based on their classification to MILE and MELE approaches is presented by Masoudnia and Ebrahimpour [88], where the authors also compare the advantages and disadvantages of MILE and MELE.

One of the main challenges in ME modeling is the automatic determination of the number of experts *a priori* [77]. In fact, this is a difficult problem in data clustering in general [89]. Research on addressing this issue is ongoing. Yuksel *et al.* [77] categorized the proposed approaches into four main categories, namely growing, pruning, exhaustive search, and Bayesian models. Ueda and Ghahramani [90, 91] used a variational framework to simultaneously estimate the parameters and model structure of an ME, by treating the number of experts as a random variable. In the infinite mixture of Gaussian Processes proposed by Rasmussen and Ghahramani [92], the ME model was assumed to have an infinite number of experts, thus eliminated the needs to specify the number of experts explicitly.

4.2 Proposed Approach

In this work, we propose a means to combine surrogate models using the general formulation of ME approach, which is shown below

$$\hat{y}\left(\mathbf{x}_0\right) = \sum_{k=1}^{K} \pi_k\left(\mathbf{x}_0\right) \hat{y}_k\left(\mathbf{x}_0\right), \quad \text{with} \quad 0 \leq \pi_k\left(\mathbf{x}_0\right) \leq 1 \quad \text{and} \quad \sum_k \pi_k\left(\mathbf{x}_0\right) = 1. \tag{14}$$

The notation $\hat{y}_k\left(\mathbf{x}_0\right)$, $k = 1, \ldots, K$ is the local surrogate model (i.e., the local expert), where K denotes the total number of experts in the mixture. $\pi_k\left(\mathbf{x}_0\right)$ is the mixing proportion, which depends on the evaluation point location $\mathbf{x}_0$ in the input space. The final prediction $\hat{y}\left(\mathbf{x}_0\right)$ is therefore a linear superposition, or weighted combination, of multiple surrogate models. The basic idea is to have several surrogate models to be responsible for different parts of the input space, to enable modeling the heterogenous complexity in the function profile.

In this formulation, we adopt the MELE approach. Implementing this approach gives us the liberty to use different surrogate modeling types for different partitions in the input space. This is possible since the partitioning and learning processes are independent of each other, unlike in the MILE approach. We choose to partition the training data based on an attribute that reflects the function profile to be modeled, such as its function values or derivative information. This translates to a one-dimensional problem in the clustering algorithm, which is simpler. For this purpose, we perform an unsupervised learning algorithm, which can learn the pattern or hidden structure of unlabeled data. The mixing proportion is typically derived based on the cluster posterior probability, i.e., the probability that the k-th cluster is active at $\mathbf{x}_0$. Bettebghor *et al.* [85] chose the cluster with the highest posterior probability (hard-split) to give the approximation. Tresp [93] used the softmax function of the Gaussian processes as the mixing proportion. Another distribution type can also be used, such as the Dirichlet distribution as used by Shi *et al.* [78]. In our approach,

however, the initial clustering of the training data is performed in the y-space, i.e., the training data are partitioned based on the function values or derivatives, without regard to their locations in the input space. To enable computing the corresponding cluster posterior probability at $\mathbf{x}_0$, we therefore need to map the clustering in the y-space to that in the x-space. For this purpose, we can perform a supervised learning algorithm, since the training data are now already labeled upon completing the unsupervised learning algorithm in the first stage. In this work, we use the Gaussian mixture models (GMM) [94] as the unsupervised learning algorithm, and the regularized Gaussian classifier [76] as the supervised learning algorithm.

We found that both hard-split and using the softmax function in the model combination have some limitations in the approximation accuracy of the ME models, which will be further elaborated in Section 6. To address this issue, we derive a "modified" softmax function as mixing proportion in our ME model, which is described below. The cluster posterior probability can be expressed as

$$p\left(z_k = 1 | \mathbf{x}\right) = \frac{p\left(\mathbf{x} | z_k = 1\right) p\left(z_k = 1\right)}{\sum_j p\left(\mathbf{x} | z_j = 1\right) p\left(z_j = 1\right)}, \tag{15}$$

where $z_k\left(\mathbf{x}_0\right)$ is an unobservable latent indicator which assigns data points to local experts in the mixture [78]. The K-dimensional binary random variable $z_k\left(\mathbf{x}_0\right)$ is a 1-of-K encoding where $z_k\left(\mathbf{x}_0\right) \in \{0, 1\}$ and $\sum_k z_k\left(\mathbf{x}_0\right) = 1$. In other words, the k^{th} local model is active when $z_k = 1$. When there are only two clusters ($K = 2$), this cluster posterior probability becomes a *sigmoid function*, an S-shaped curved with values ranging from 0 to 1,

$$p\left(z_1 = 1 | \mathbf{x}\right) = \frac{1}{1 + \exp\left(-a\right)} = \sigma(a), \quad \text{where} \quad a = \ln \frac{p\left(\mathbf{x} | z_1 = 1\right) p\left(z_1 = 1\right)}{p\left(\mathbf{x} | z_2 = 1\right) p\left(z_2 = 1\right)} \tag{16}$$

The *cluster boundary* is defined at the point where $p\left(z_k = 1 | \mathbf{x}\right) = 0.5$. We can modify the sigmoid function by introducing weight (ω) and bias (λ), $\sigma\left(\omega a + \lambda\right)$. Altering λ shifts the cluster boundary, whereas altering ω changes the slope of the S-shaped curve around the cluster boundary. Since we want to maintain the cluster boundary position, we set λ to the default value 0. Increasing ω drives the sigmoid function to be closer to a step function, or $p\left(z_k = 1 | \mathbf{x}\right) = \{0, 1\}$ as $\omega \to \infty$. For cases where $K > 2$, we use a softmax function,

$$p\left(z_k = 1 | \mathbf{x}\right) = \frac{\exp\left(a_k\right)}{\sum_j \exp\left(a_j\right)}, \quad \text{where} \quad a_k = \ln\left[p\left(\mathbf{x} | z_k = 1\right) p\left(z_k = 1\right)\right]. \tag{17}$$

The previous discussion on the effects of adding ω and λ to the sigmoid function also applies to this softmax function. In our proposed approach, we use this "modified" softmax function as the mixing proportion,

$$\pi_k\left(\mathbf{x}_0\right) = \frac{\exp\left(\omega a_k\left(\mathbf{x}_0\right)\right)}{\sum_j \exp\left(\omega a_j\left(\mathbf{x}_0\right)\right)} \tag{18}$$

We will vary ω and discuss how it affects the predictive performance of the mixture of experts, which will be presented in Section 6.

The procedure for the mixture of experts we developed can be summarized in the following steps:

1. Implement the Gaussian mixture model as the unsupervised learning algorithm to cluster the training data. The designers need to decide on the clustering criterion and the number of clusters prior to performing this step. The training data set for clustering, $\mathcal{T} = \{\mathbf{x}_n, y_n\}_n$, is now partitioned into K clusters, $\mathcal{T}_k = \{\mathbf{x}_n, y_n\}_{n \in \mathcal{C}_k}$, $k = 1, \ldots, K$, where $\mathcal{C}_k$ denotes the set of clustering training data indices that correspond to the k^{th} cluster.

2. Map the clustering of training data to the clustering in the input space ($\mathbf{x}$-space) by implementing the regularized Gaussian classifier as the supervised learning algorithm.

3. Build a separate local surrogate model within each cluster, $\hat{y}_k\left(\mathbf{x}\right)$, $k = 1, \ldots, K$.

4. Compute the cluster posterior probability, i.e., the probability that $\mathbf{x}_0$ belongs to the k^{th} cluster, using (15).

5. Compute the corresponding mixing proportion, $\pi_k\left(\mathbf{x}_0\right)$, using (18).

6. Compute the mixture of experts estimation, $\hat{y}\left(\mathbf{x}_0\right)$, following (14), using the local experts and mixing proportions obtained in steps 3 and 5, respectively.

To further improve the approximation accuracy of the ME models, we complement the local surrogate model training with adaptive sampling procedure. This is possible since each local surrogate model is independent of each other, and built based on exclusive sample set. Some traditional ME models, on the other hand, use the same training data for all local experts and gating functions [27]. At each cluster, this adaptive procedure starts with a subset of $\mathcal{T}_k$, until the termination criterion is achieved, which will be described in more details in Section 5.3.

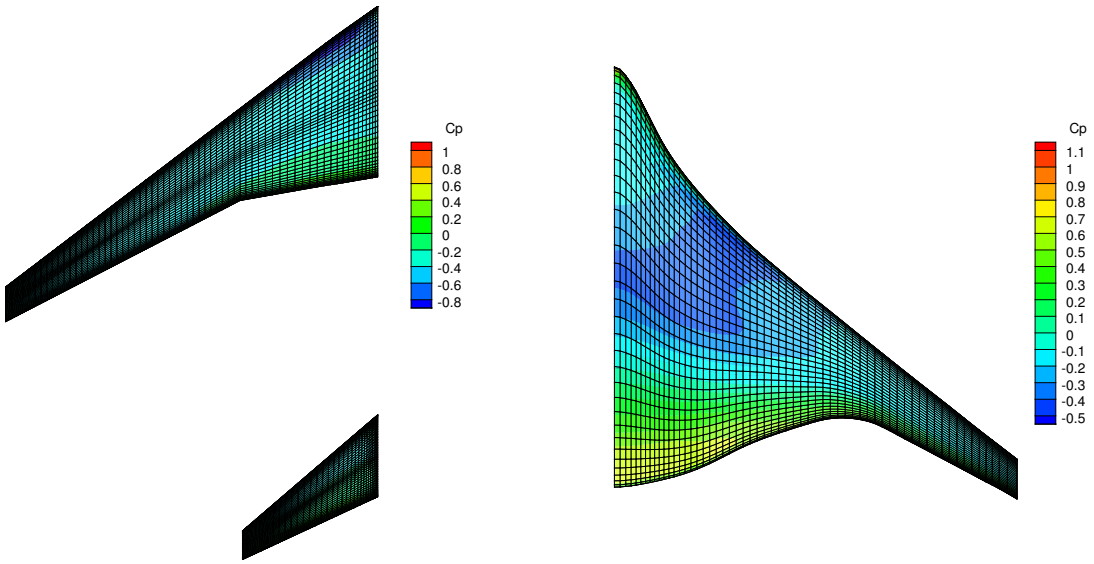
In this paper, we demonstrate the effectiveness of the proposed mixture of experts procedure in creating a surrogate model for the aerodynamic force and moment coefficients of aircraft configurations. The performance is then compared to those of some conventional global surrogate models.

5 Problem Description

In this section we describe the two aircraft configurations considered in this study and the aerodynamic solver used to generate the aerodynamic coefficient data. We then provide more details on the surrogate models that we benchmark, as well as the selected sampling and model validation procedures.

5.1 Aircraft Configurations

Two Boeing 777-size configurations are considered in this work to demonstrate the proposed surrogate-based mission analysis procedure: one conventional and one unconventional. For the conventional configuration, we use the wing-tail from the Common Research Model (CRM) [95]. This aircraft exhibits design features typical of a transonic, wide-body, long-range aircraft, with overall dimensions similar to those of the Boeing 777-200ER. For the unconventional configuration, we consider a BWB configuration with the sizing parameters used by Lyu and Martins [21]. Figure 1 shows the layouts for both aircraft configurations, and the grid to be used in the aerodynamic solver, which is described next.



(a) CRM (conventional)

(b) BWB (unconventional)

Figure 1: Aircraft configurations considered in this study, showing the grid used in the aerodynamic solver.

5.2 Aerodynamic Solver

An aerodynamic panel code, TriPan, is used to generate the aerodynamic force and moment coefficient data (both samples and validation points) in this work. This solver, developed by Kennedy and Martins [96, 97], calculates the aerodynamic forces and moments of inviscid, incompressible, external lifting flows on unstructured grid using surface pressure integration, with constant source and double singularity elements. The induced drag is computed using a Trefftz plane integration and the code includes a skin friction and compressibility drag estimate as well. TriPan can

be used with both quadrilateral and triangular panels. In this study, quadrilateral panels are used for the surface discretization of the two aircraft configurations, as shown in Figure 1.

5.3 Surrogate Models

The surrogate models to approximate the lift, drag, and pitching moments coefficients (C_L, C_D, C_M) are constructed in a four-dimensional space with input variables: Mach number (M), angle of attack (α), flight altitude (h), and tail rotation angle (η). Due to the varying magnitudes of the input variables (in particular between the flight altitude and other input variables), the input variables are scaled to be between 0 and 1 prior to constructing the surrogate models.

We consider both global and ME models. As previously mentioned in Section 3, we consider black-box (data-fit) surrogate models that are interpolative, i.e., kriging and RBF models. For the RBF models, three kernel functions are used, namely the thin plate splines, cubic, and square-exponential (Gaussian). For the kriging models, since we know that the aerodynamic force and moment coefficients are continuously differentiable, we consider only the correlation functions that exhibit parabolic behavior, namely the Gaussian and cubic spline functions. Direct GEK and universal kriging models are also benchmarked.

For the universal kriging model (kriging with a trend), the global model is modeled by an analytical expression, which takes different values in space, to be the trend component [25]. Using this approach, instead of being restricted to using low-order polynomials as the basis functions, we select the basis functions that reflect the physics of the system, to assist the prediction. Since we know that drag coefficient profiles are expected to have a steep gradient in the high Mach (M) and high angle of attack (α) region, we set the trend to be

$$\psi\left(M,\alpha\right)=\begin{cases} 1/\left(1-M^2\right) & \text{if} \quad \alpha\leq 1.0 \\ \alpha^2/\left(1-M^2\right) & \text{if} \quad \alpha\geq 1.0 \end{cases}. \tag{19}$$

The constant numerator for $\alpha \leq 1.0$ is used to remove the quadratic profile (in the α dimension) in the low α region, to be consistent with the C_D profile obtained from the aerodynamic solver. The basis function vector, $\mathbf{f}\left(\mathbf{x}\right)$, and the coefficient vector, $\boldsymbol{\beta}$, are thus expressed as follows,

$$\mathbf{f}\left(\mathbf{x}\right)=\left[1,\psi\left(M,\alpha\right)\right] \quad \text{and} \quad \boldsymbol{\beta}=\left[\beta_0,\beta_1\right]^T. \tag{20}$$

Thus at an evaluation point $\mathbf{x}_0=[M_0,\alpha_0,h_0,\eta_0]$, the kriging equation can be expressed as

$$\hat{y}\left(\mathbf{x}_0\right)=\beta_0+\beta_1\psi\left(M_0,\alpha_0\right)+\mathbf{r}\left(\mathbf{x}_s,\mathbf{x}_0\right)^T\mathbf{R}^{-1}\left[\mathbf{y}_s-\beta_0-\beta_1\psi\left(M_0,\alpha_0\right)\right]. \tag{21}$$

The basis function coefficients, β_0 and β_1, are obtained by computing their least squares estimates.

The Halton sampling sequence, which is a space-filling low-discrepancy method [57], is used to generate training samples to construct the surrogate models, as well as to generate the clustering training data for constructing the ME models. The *discrepancy* in this case refers to the departure of the sampling points from a uniform distribution, thus ensuring an even distribution of samples over the input space. Moreover, Halton sample generation is done in an incremental fashion. That is, when we increase the size of training samples (N_s), we reuse the points from the smaller sample set. With this incremental sampling, we can compare the surrogate modeling performance with different sizes of sample sets more fairly, compared to other sampling method like LHS, which generates a new set of samples for each sample size.

For the ordinary kriging and GEK models, we also use the adaptive sampling procedure following the exploration infill criterion. At each iteration, we select a point with the maximum index of dispersion, or *variance-to-mean ratio* (VMR), σ^2/μ, of the kriging prediction as the next sample. Using the maximum VMR instead of the more commonly used maximum variance criterion takes into account the varying magnitudes of kriging predictions at different parts of the input space. An initial set of points is required to perform this adaptive sampling procedure. For this purpose, we use 15 Halton points for the global models, and the first 15 points in each clustering data set $\mathcal{T}_k$ for the ME approach. This number of initial samples is deemed sufficient to have a reasonable surrogate model to start with for the kind of problem complexity we consider here, while giving the adaptive sampling procedure enough "room" to be effective.

The adaptive sampling procedure is terminated when the convergence criterion is achieved (maximum VMR $<$ tolerance), or when the specified maximum number of samples (sampling budget) is reached. For simplicity, we select the next sample out of a set of 10 000 candidate points, which are distributed uniformly in the input space. Note that no actual function evaluations are required to compute the VMR values at those points, since the variance σ^2 and mean μ of the kriging prediction come out naturally from the kriging derivation and can be expressed analytically.

The actual function evaluation is only required at the selected sample location, to update the sample set $\mathcal{S}$. We found that using this approach significantly accelerates the convergence of the VMR maximization, compared to using some optimization techniques, resulting in much smaller sample sizes.

To validate the surrogate models, we generate $10\,000$ validation points with the aerodynamic solver, as the truth set data. These data are used to compute the normalized RMS error, with which we assess and compare the accuracy of surrogate models tested.

For the ME models, we follow the procedure presented in Section 4. Ordinary kriging and direct GEK models with adaptive sampling are used as the local experts. With the divide-and-conquer approach, the computational cost required to build and use the multiple kriging models is reduced. The correlation matrix for the ordinary kriging model is $\mathcal{O}\left(N_s^2\right)$ in size and its inversion is $\mathcal{O}\left(N_s^3\right)$ in cost. Even when the total number of samples used are the same for the global model and the ME model, the total computational cost is lower for the latter. When the job is distributed to local experts, we can disregard the correlation between samples that belong to different subregions. Moreover, each local expert is free to select the best model parameters to better reflect the characteristics of the underlying function in the input subregion it is responsible for (e.g., by having different length scales, $\boldsymbol{\theta}$, for each local kriging model). In short, this divide-and-conquer approach allows us to distribute a complex task into multiple simpler tasks. The numbers of clustering training data are 100 for the two-dimensional test cases, and 500 for the four-dimensional test cases. In this work, we try several numbers of clusters until there are "empty" clusters, i.e., clusters with zero or very few training data. We then compare and discuss the performance of the ME models with different numbers of clusters.

5.4 Mission Analysis

In addition to verifying the modeling accuracy of surrogate models in the input space, we will also look into how this accuracy translates to the accuracy of surrogate models when used to evaluate the aircraft performance via the mission analysis procedure. For this purpose, we run a *reference mission analysis* using TriPan, instead of surrogate models, to compute the aerodynamic force and moment coefficients. Due to the high computational cost of running the reference mission analyses, we limit the comparison and verification to 10 mission profiles for the BWB configuration. For the same reason, the numerical integration is performed with only four intervals per segment.

The 10 benchmark mission profiles are randomly selected from the typical payload-range diagram for long-range aircraft configurations, such as a Boeing 777-200ER configuration[1]. The same payload-range diagram was also used in the multipoint high-fidelity aerostructural optimization problem, which minimizes fuel burn over a large number of different missions [19]. Although these 10 benchmark mission profiles correspond to the same vehicle configuration, randomization is introduced in the mission Mach number (M) and altitude (h) to account for the operational variation, since each flight mission is independent of each other. The Mach number is drawn randomly to be within $[0.7, 0.88]$ to account for the unknown operational demands that might require faster or slower flights, whereas the altitude is varied between $[28\,000\,\text{ft}, 41\,000\,\text{ft}]$ to simulate the variability in the altitudes assigned by air traffic control. The four mission parameters, namely mission payload, range, cruise Mach number, and cruise altitude for the 10 verification missions are summarized in Table 1 (sorted by the mission range), and visualized in the payload-range diagram shown in Figure 2. In this diagram, each scatter point, which represents a mission profile, is sized with the cruise Mach number, and is color coded by the cruise altitude.

Index	Payload (kg)	Range (nmi)	Cruise Mach	Altitude (ft)
1	37 969	580	0.77	39 939
2	29 531	1 704	0.76	28 796
3	23 203	2 025	0.82	38 347
4	9 844	2 747	0.81	29 061
5	43 594	3 469	0.74	34 633
6	18 281	3 870	0.80	30 653
7	11 953	4 191	0.85	40 204
8	26 719	4 593	0.70	36 224
9	7 031	5 315	0.73	39 673
10	15 469	6 759	0.74	38 082

Table 1: List of mission parameters for the ten benchmark mission profiles.

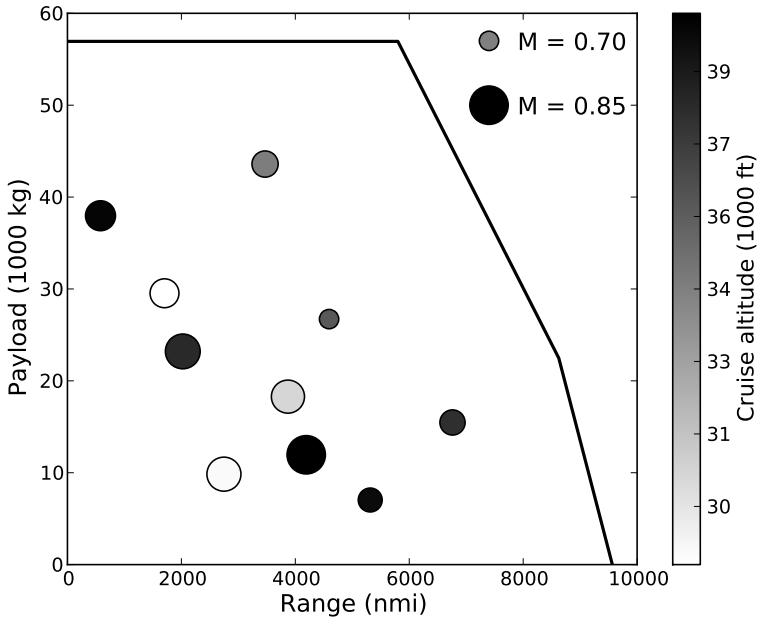


Figure 2: Selected mission parameters for the mission analysis verification. Each point in the payload-range diagram is sized to be proportional to the cruise Mach number, whereas the color is mapped to the cruise altitude

A typical mission profile is used in the analysis, as shown in Figure 3, and described in Table 2. We only include one main profile, without loiter or reserve profiles. For longer range missions, we assume a step climb procedure for every $2\,000\,\mathrm{nmi}$, with a $2\,000\,\mathrm{ft}$ altitude increment at each step. In this illustration we only show two cruise steps. The cruise altitude specified in the mission parameters is used for the first cruise segment. The entire cruise portion of the flight, from the initial cruise (segment 8), through the step climb (segment 9), to the final cruise (segment 10), is done at a constant Mach number M, which is specified in the mission parameters. From $10\,000\,\mathrm{ft}$, the climb is done at a constant KIAS, which is the indicated airspeed in knots (segment 6), until it intercepts the desired cruise Mach number, at which point the climb is done at a constant Mach number (segment 7). The altitude where the constant Mach segment starts is denoted as h_{cm}. The descent is also done in a similar fashion, with a constant Mach descent (segment 11) followed by a constant KIAS descent (segment 12). The same fuel fraction value of 0.01 is used for the startup, taxi, and landing segments, whereas 0.005 is used for takeoff segment. These values follow those suggested by Roskam [6], Raymer [98], and Sadraey [99].

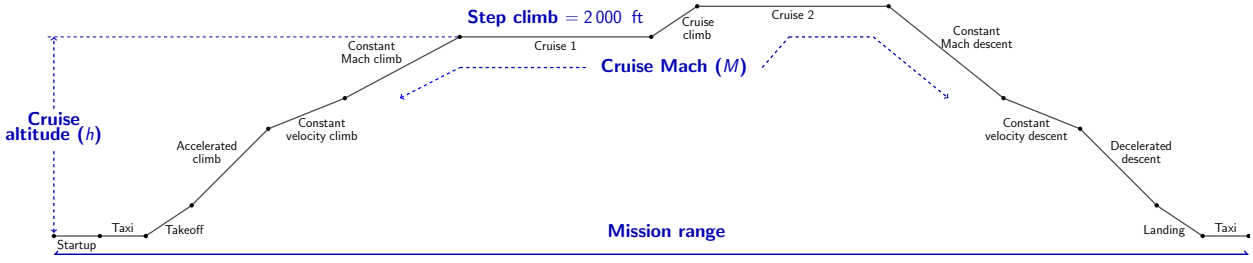


Figure 3: Typical mission profile for a long-range configuration (not to scale).

The main focus of this study is to investigate the performance of surrogate models in the mission analysis procedure. As such, the benchmarking case is kept simple by assuming a constant TSFC (0.53 lb/(lbf $\cdot$ h)), instead of using an engine model. However, a full integration of engine performance model to the mission analysis procedure is currently an ongoing work by the authors and their colleagues, to achieve a coupled aerostructural and mission parameter optimization. We use a weight and balance model with four components, namely the mission payload (20

	Segment	Altitude [ft]	Speed
1	Startup	—	—
2	Taxi	—	—
3	Takeoff	—	—
4	Climb	$1\,500 \rightarrow 10\,000$	$150\,\text{KIAS} \rightarrow 250\,\text{KIAS}$
5	Cruise	$10\,000 \rightarrow 10\,000$	$250\,\text{KIAS} \rightarrow 310\,\text{KIAS}$
6	Climb	$10\,000 \rightarrow h_{cm}$	$310\,\text{KIAS}$
7	Climb	$h_{cm} \rightarrow h$	M
8	Cruise	h	M
9	Climb	$h \rightarrow h + 2\,000$	M
10	Cruise	$h + 2\,000$	M
11	Descent	$h + 2\,000 \rightarrow h_{cm}$	M
12	Descent	$h_{cm} \rightarrow 10\,000$	$310\,\text{KIAS}$
13	Cruise	$10\,000$	$310\,\text{KIAS} \rightarrow 250\,\text{KIAS}$
14	Descent	$10\,000 \rightarrow 1\,500$	$250\,\text{KIAS} \rightarrow 150\,\text{KIAS}$
15	Landing	—	—

Table 2: Mission profile parameters for typical mission.

tons), fixed weight (200 tons), and fuel weight, which depends on the mission analysis. These component weights and moments gives an estimate of the entire aircraft's weight, as well as the nominal, forward, and aft center of gravity locations. During the mission analysis, the weight and center of gravity locations of these components can be individually updated, giving a more accurate picture of the aircraft's weight and balance as fuel is decremented in the integration.

5.5 Software Architecture

The computationally intensive part of our mission analysis module is implemented in Fortran, and then wrapped with Python. This combination has been proven to be effective. Fortran offers a significantly faster computational time as compared to Python, and the object-oriented Python provides the more practical user interface (scripting), ease of use of a class object, and plotting features. Using Python at the scripting level has also facilitated the integration of the different Fortran modules (e.g., aerodynamic solver, mission analysis, surrogate models, atmospheric module, and engine model). The aerodynamic and mission solvers are suitable for parallel implementation. In this work, the analyses are performed on a parallel computing platform [100].

6 Benchmarking Results

In this section, we first focus our discussion on the surrogate modeling performance comparison in approximating the C_L, C_D, and C_M in the four dimensional input space, for both the CRM and BWB configurations. We then present the surrogate model verification in the mission analysis context. The value ranges for the surrogate model input variables are listed in Table 3. The surrogate modeling techniques considered in this study are listed in Table 4. For simplicity,

Input variable	**Lower bound**	**Upper bound**
Mach number (M)	0.15	0.90
Angle of attack (α)	$-10.0°$	$20.0°$
Altitude (h)	0 ft	50 000 ft
Tail angle (η)	$-20.0°$ (CRM), $4.0°$ (BWB)	$20.0°$ (CRM), $40.0°$ (BWB)

Table 3: Value ranges for the surrogate model input variables.

"ordinary kriging" is referred to as "kriging", and "direct GEK" is "GEK" in the subsequent result presentation and discussion.

For illustration purposes, we first demonstrate the methods with a two-dimensional case, using data corresponding

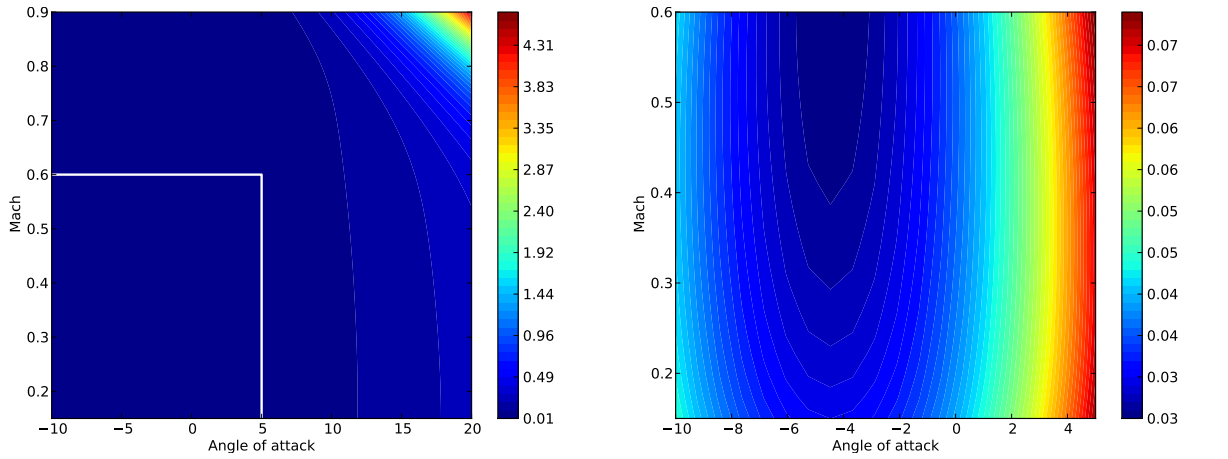
Model type	Kernel/correlation function	Sampling
Global models		
Kriging	Cubic (C)	Halton
	Gaussian (G)	Halton
	Gaussian (G)	Adaptive (maximum VMR)
Universal kriging	Gaussian (G)	Halton
GEK	Cubic (C)	Halton
	Gaussian (G)	Halton
	Gaussian (G)	Adaptive (maximum VMR)
RBF	Cubic (C)	Halton
	Gaussian (G)	Halton
	Thin plate splines (TPS)	Halton
Mixture of experts		
Kriging	Gaussian (G)	Adaptive (maximum VMR)
GEK	Gaussian (G)	Adaptive (maximum VMR)

Table 4: Surrogate models benchmarked in this study, and the corresponding model structures and sampling techniques.

to the BWB configuration. Next, the results corresponding to the four-dimensional cases are presented for both the BWB and CRM configurations. We then use the surrogate models corresponding to the BWB configuration for the mission analysis verification.

6.1 Two-dimensional Case with BWB Configuration

For the two-dimensional case, we fix the flight altitude to $38\,500\,\mathrm{ft}$ and the tail angle to $7.0°$, and we only discuss the surrogate model performance in approximating C_D. This is because the C_D profile exhibits a more complex profile than C_L or C_M, especially in the transonic region. Figure 4 shows the C_D contour using the truth set data. The entire input space is shown in Figure 4a, where we can observe the high drag gradient at the high M, high α region. Due to the large value range shown in this contour plot, the lower left corner seems flat. However, when we zoom into the region defined by the white rectangle, we see a quadratic profile, as shown in Figure 4b.



(a) C_D contour showing the high drag gradient

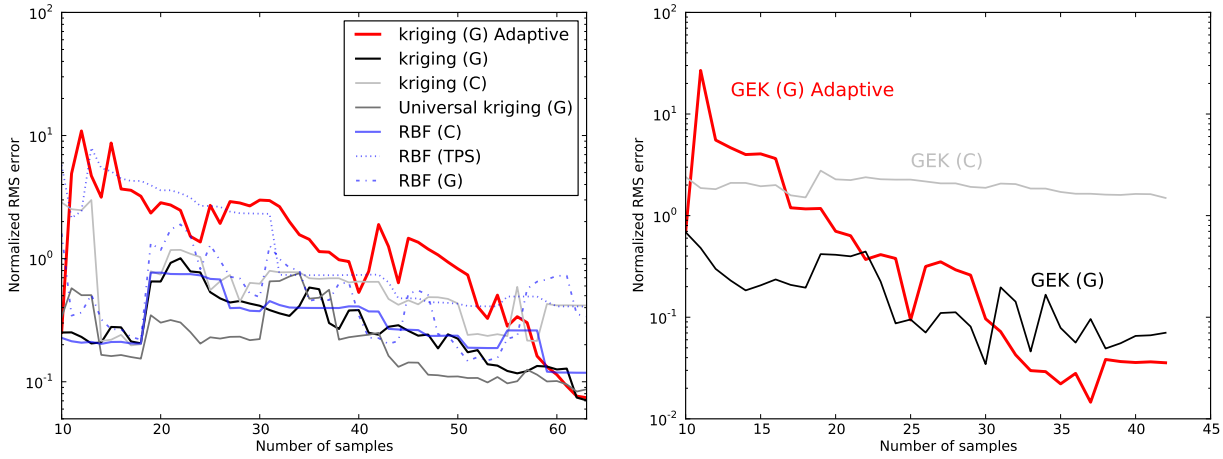
(b) C_D contour detail showing a much lower C_D value range

Figure 4: Drag coefficient contour exhibits different profiles in different input space regions. The white rectangle shown in the left hand side figure highlights the subset shown in the right hand side figure.

Now we look at the performance of global models in approximating this C_D contour. We first build kriging and GEK models with adaptive sampling, and then use the same sample size as the maximum number of Halton samples

considered. The convergence criterion used for the adaptive sampling in this case is maximum VMR $\leq 10^{-5}$. The convergence is achieved at $N_s = 63$ for the kriging model, whereas the GEK model converges with 42 samples.

The normalized RMSE for different surrogate models that use no gradient information are shown in Figure 5a, and those for GEK models are shown in Figure 5b. From the error plots we observe that the error trend is more monotonically decreasing when adaptive sampling is used. Although using more Halton samples in general decreases the approximation error, the convergence trend is more erratic than when we use adaptive sampling. Among the three RBF models, the one with the thin plate spline kernel function has the worst performance, especially with smaller sample size. The universal kriging, using the basis functions given in Section 5.3, shows the best performance when fewer samples are used, but is caught up by kriging models (with Gaussian correlation function) as more samples are added. This result shows that adding a known trend to the kriging model does improve the predictive performance, especially when we have a small sample budget. For the GEK models, using a cubic correlation function results in a poor predictive performance. In fact, its performance is worse than when no gradient information is used. GEK models require computing the second derivatives of the correlation function to assemble the extended correlation matrix (to include the correlation between function values and gradients, as well as between gradients). While the second derivatives of a cubic correlation function is continuous, it is only piecewise linear and thus not smooth. The Gaussian correlation function, on the other hand, has a smooth second derivative.



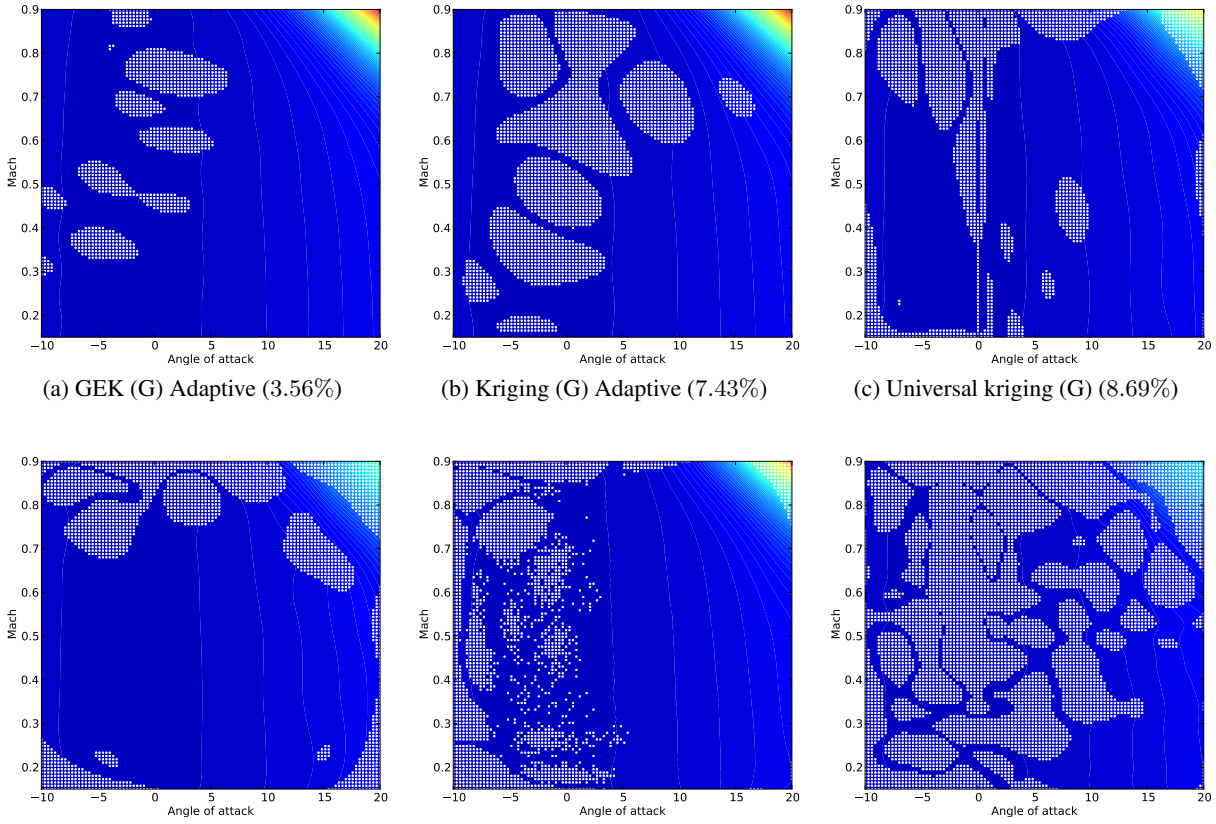
(a) No gradient information

(b) With gradient information

Figure 5: The normalized RMS error for the two-dimensional BWB case show decreasing trends when adaptive sampling is used, resulting in more accurate surrogate models. Halton sampling is used unless otherwise stated.

The C_D contours given by the six global models are shown in Figure 6, using the maximum number of samples shown in Figure 5. We use white dots to indicate the test points with $> 5\%$ approximation errors, to visualize the input space regions where each surrogate model performs poorly. The GEK model with Gaussian correlation function and adaptive sampling (Figure 6a) shows the best performance, both in terms of the normalized RMSE and the error distribution. The gradient information seems to help fitting the different function characteristics in the different input space region significantly. Kriging (Figure 6b) and universal kriging (Figure 6c) models can both follow the trend in the high drag gradient region pretty well, but the performance in the low α region is still rather poor. A similar error distribution is observed when an RBF model with Gaussian kernel function is used (Figure 6e), though the overall normalized RMSE is significantly higher. Choosing a different kernel function affects the RBF model performance, as seen in Figure 6d where a cubic kernel function is used. This model shows an overall good performance, except in the regions that are close to the input space boundary. Kriging with cubic correlation function (Figure 6f) shows a poor predictive performance in the entire input space region, which is also reflected in its high normalized RMSE. We will now look into the ME results, and compare their performance to the global model performance.

Before generating the ME models, we need to determine the mixing proportions $\pi_k(\mathbf{x})$. As mentioned in Section 4, we use the modified cluster posterior probability as $\pi_k(\mathbf{x})$ (18), where we need to specify the weight ω. In Figure 7, we show the effect of changing ω on $\pi_k(\mathbf{x})$ (top row), and on the resulting C_D approximation contours (bottom row). For the $\pi_k(\mathbf{x})$ plots, we use different colors to indicate the different clusters. The color intensity within each cluster



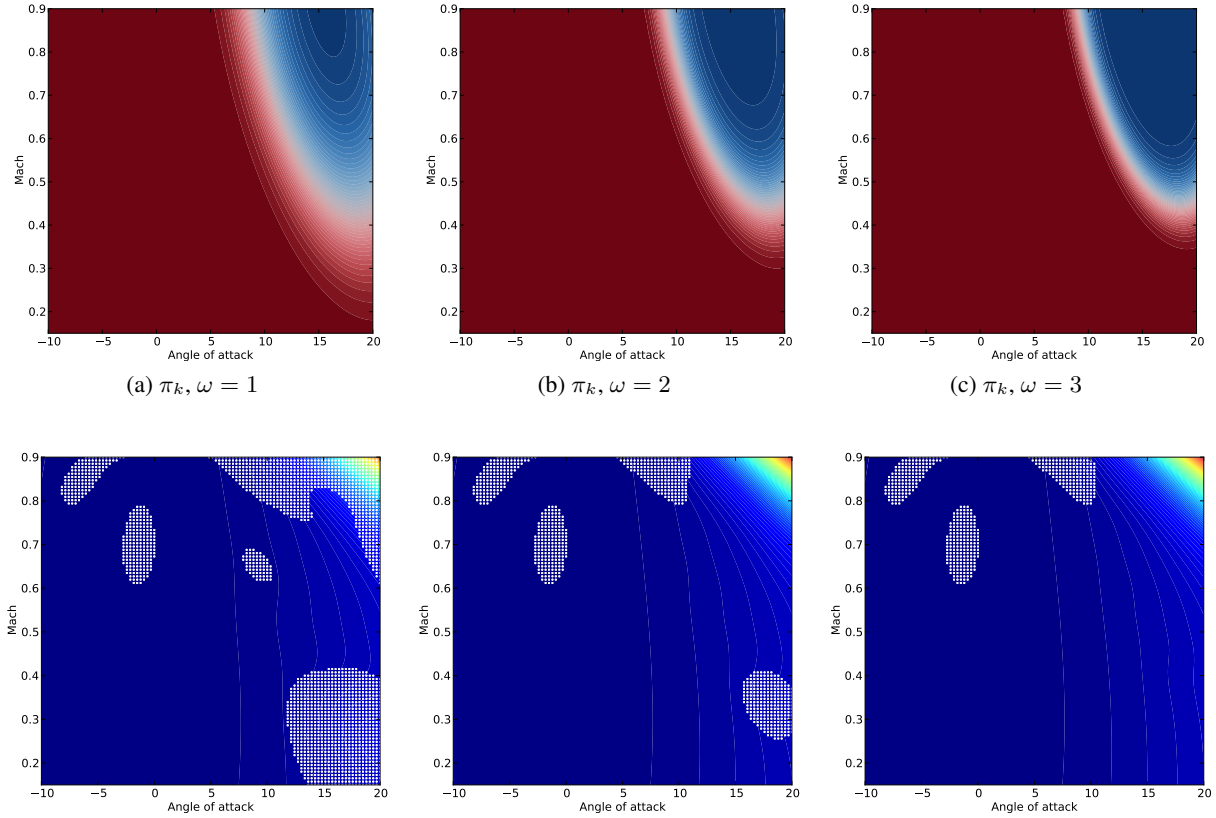
(a) GEK (G) Adaptive (3.56%) (b) Kriging (G) Adaptive (7.43%) (c) Universal kriging (G) (8.69%)

(d) RBF (C) (11.83%) (e) RBF (G) (33.94%) (f) Kriging (C) (41.68%)

Figure 6: The approximated C_D contours for the BWB configuration in a two-dimensional space from different global surrogate models, with the normalized RMS errors shown inside the brackets. The test points with $> 5\%$ approximation errors are indicated by the white dots.

represents the $\pi_k(\mathbf{x})$ value, where $0 \leq \pi_k(\mathbf{x}) \leq 1$. The lightest color corresponds to $\pi_k(\mathbf{x}) = 0$, whereas the darkest corresponds to $\pi_k(\mathbf{x}) = 1$. We show three ω values: 1 (the default value for the original posterior probability function), 2, and 3. For this benchmarking, we partition the input space based on the derivative ($\partial C_D/\partial M$) criterion, and use kriging models as the local experts.

As we can observe from these plots, the cluster boundary gets more clearly defined as ω is increased, which increases the sigmoid function slope. When $\omega = 1$, the region in the input space where both local experts "share responsibility" is larger. Consequently, each local expert needs to approximate the function value *beyond* its local area. Since kriging (including GEK) models are not good at extrapolation, this poor predictive performance is reflected in the overall approximation accuracy, as shown in Figure 7d. Increasing ω decreases the areas outside the local region that each expert needs to predict, resulting in the better predictive performance seen in both the error distribution plots and the overall normalized RMSE. Further increasing ω above 3 does not affect the predictive performance of the ME model, as shown in the error convergence plot displayed in Figure 8.



(a) π_k, $\omega = 1$ (b) π_k, $\omega = 2$ (c) π_k, $\omega = 3$

(d) $\hat{C}_D$, $\omega = 1$ (11.00%) (e) $\hat{C}_D$, $\omega = 2$ (3.12%) (f) $\hat{C}_D$, $\omega = 3$ (2.87%)

Figure 7: Effect of changing ω in computing $\pi_k(\mathbf{x})$ (18). For the $\pi_k(\mathbf{x})$ plots (top row), different colors correspond to different clusters. The highest color intensity within each cluster corresponds to $\pi_k(\mathbf{x}) = 1$ (maximum value). The overall normalized RMSE are shown inside the brackets.

In this two-dimensional study, we consider two clustering criteria: the function value (C_D) and the derivative of C_D with respect to Mach number ($\partial C_D/\partial M$). Both kriging and GEK models are considered as the local experts, and the samples are drawn adaptively. The results are summarized in Figure 9, showing the total number of samples N_s and the overall normalized RMSE for all cases considered here. When using multiple kriging models instead of one global kriging model, the ME models reduce the RMSE from 7.4% to $3-4\%$. When we use $\partial C_D/\partial M$ as the clustering criterion, the improved performance is achieved with fewer samples. When using GEK models as the local experts, we can achieve $< 1\%$ overall performance error with similar sample sizes as compared to when a global model is used. Comparing the two clustering criteria, the overall approximation errors are of the same order for each number of

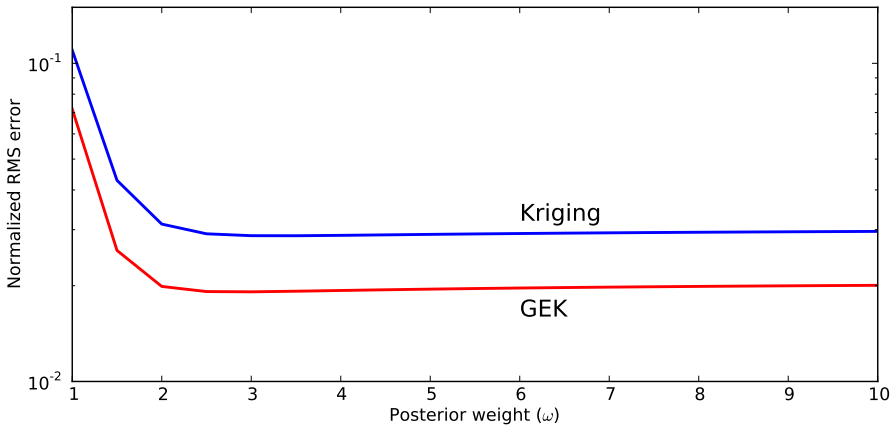


Figure 8: Normalized RMSE converges upon increasing ω in $\pi_k(\mathbf{x})$.

clusters considered. However, fewer samples are required when we use $\partial C_D/\partial M$ as the clustering criterion, showing that this derivative value is a better indicator for the heterogeneity in the function profile.

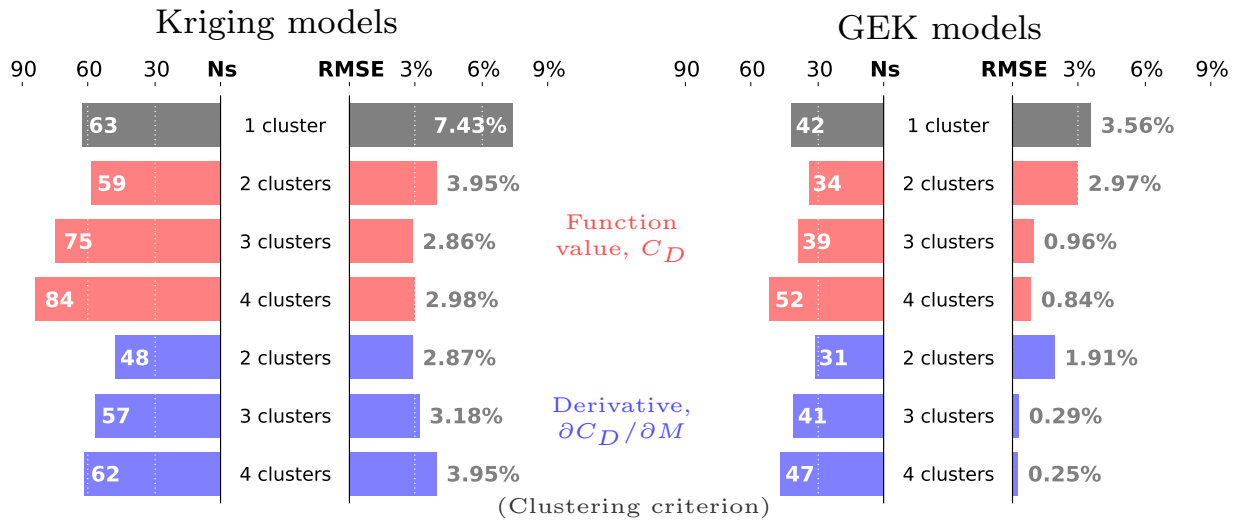


Figure 9: Mixture of experts result summary with two clustering criteria to approximate C_D contour in a two-dimensional space for the BWB configuration.

We now compare the ME performance for different numbers of clusters (local experts). The top row of Figure 10 shows the partitioning of input space (shown as the mixing proportion contour plots) with 2, 3, and 4 clusters when using $\partial C_D/\partial M$ as the clustering criterion. The C_D approximation contours with kriging and GEK models as the local experts are also shown, with the distribution of test points with $> 5\%$ approximation errors shown as white dots. These plots show that the mixtures of GEK models offer a notably better performance than their kriging counterparts.

We now look at the optimum length scales (kriging hyperparameters, $\boldsymbol{\theta}$) for the various local kriging and GEK models as obtained via the MLE procedure. Table 5 shows the different optimum $\boldsymbol{\theta}$ obtained for each local expert. Each square-bracket corresponds to the $\boldsymbol{\theta}$ of one local expert. The first number refers to the correlation parameter in the M dimension, whereas the second number explains the correlation in the α dimension. A smaller number indicates a stronger correlation. This correlation can be interpreted as how much the knowledge of function value at one point helps to deduce the function value at another point. Therefore, a simple linear function has a strong correlation, whereas a highly nonlinear function (e.g., a function which exhibits pronounced oscillations) has a weak correlation.

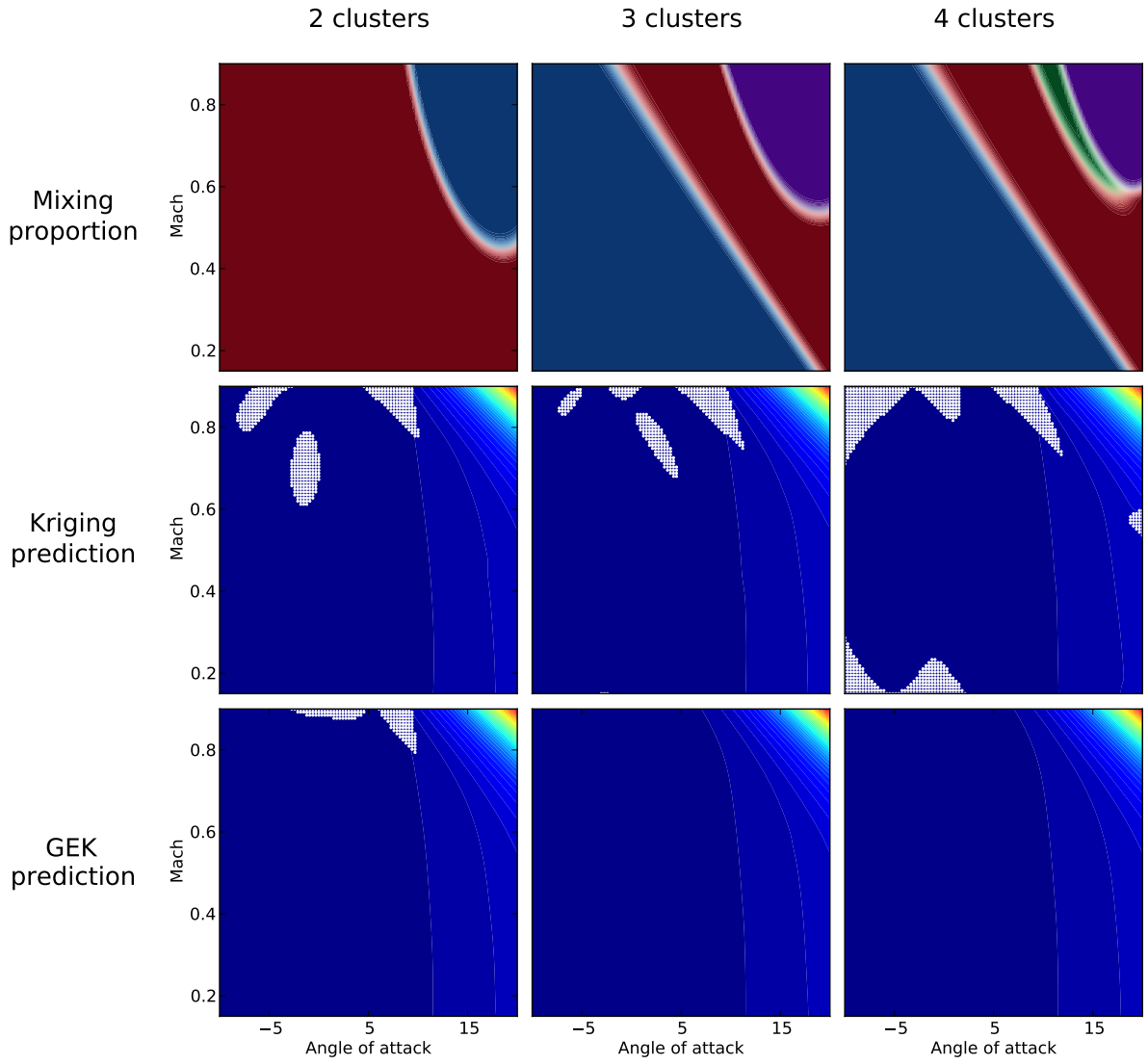


Figure 10: Mixing proportion and C_D approximation contours from the mixture of experts, using $\partial C_D/\partial M$ value as the clustering criterion for the BWB configuration.

Figure 11 displays the optimum $\boldsymbol{\theta}$ in the different input space partitions corresponding to the ME with kriging models. For the low M, low α region (blue), we see a stronger correlation in the M dimension than in the α dimension. This outcome is not surprising, as we could see in Figure 4b that C_D values do not vary much in Mach (stronger correlation, lower θ), whereas it exhibits a quadratic profile in α (weaker correlation, higher θ). In the middle region (red), we find almost equal length scales in both the M and α dimensions. In the high M, high α region (purple), the correlations are weak but it is stronger in the α dimension. The optimum length scales in the latter partition are the closest to the ones obtained when we use a single global kriging model (see Table 5), suggesting that this is the most dominant profile when fitting a global surrogate model. These observations suggest that partitioning the input space lets each local expert to model the dependence between function value and inputs separately, thus results in a better approximation model overall.

Number of clusters	**Length scales ($\boldsymbol{\theta}$)**
Local experts: kriging models	
1	$[6.04, 3.74]$
2	$[0.29, 1.90]$, $[7.03, 2.44]$
3	$[0.08, 0.35]$, $[0.01, 3.89]$, $[6.18, 2.95]$
4	$[0.08, 0.35]$, $[0.01, 3.89]$, $[7.69, 4.15]$, $[11.47, 12.05]$
Local experts: GEK models	
1	$[7.94, 2.51]$
2	$[2.70, 0.73]$, $[7.48, 2.77]$
3	$[2.63, 2.90]$, $[0.90, 2.19]$, $[8.35, 3.33]$
4	$[3.17, 3.63]$, $[0.90, 2.19]$, $[7.57, 2.60]$, $[7.28, 7.13]$

Table 5: Local kriging and GEK models have different optimum model parameters (length scales $\boldsymbol{\theta} = [\theta_M, \theta_\alpha]$) in the partitioned input space, suggesting that the divide-and-conquer approach is better in modeling the different characteristics in the function profile.

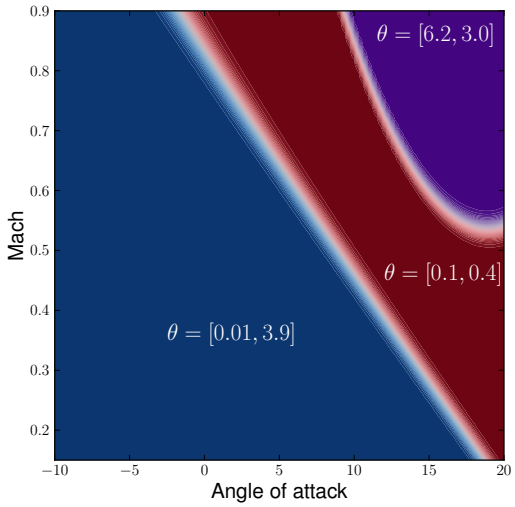


Figure 11: Different local kriging models have notable differences in the optimum model parameters (length scales $\boldsymbol{\theta} = [\theta_M, \theta_\alpha]$).

6.2 Four-dimensional Case with BWB Configuration

Modeling the C_D profile in the four-dimensional space is significantly more complicated than in the two-dimensional case. The BWB configuration, in particular, exhibits a more complex correlation between drag and trim, and thus the drag profile becomes more nonlinear in the tail angle dimension. As we will see, this complex profile imposes challenges in fitting surrogate models that accurately predict the C_D profile in the entire input space.

The adaptive sampling procedures performed for the global kriging and GEK models converge slowly. We thus set the maximum N_s to be 600 for kriging and 200 for GEK model. The convergence (or the lack thereof) of the maximum VMR, which is the criterion used for the adaptive sampling procedure, and the normalized RMSE are shown in Figures 12 and 13 for the kriging and GEK model, respectively. Drawing 600 samples adaptively for the kriging model takes around 22 hours on a single 2.00GHz processor, yet the approximation accuracy is still really poor—the normalized RMSE is 27.13%. The adaptive sampling procedure for the GEK model takes approximately 32 hours to complete (using the same processor), and yet the resulting approximation accuracy is only 43.68%. In both cases (kriging and GEK), the maximum VMR converges erratically, though the kriging model starts showing a smoother convergence at $N_s > 250$. However, looking at the convergence slope, adding more samples does not seem to provide any meaningful improvement in the approximation accuracy.

Another issue we identify from the convergence plots are the sudden spikes of maximum VMR at around 100 samples for both kriging and GEK, and also at 220 for kriging, which are also reflected in the increasing normalized RMSE. This phenomenon is common when performing an adaptive sampling procedure based on the maximum variance criterion. The procedure tries to converge to a certain kriging shape by adding more samples. However, it reaches a point where adding a sample changes the shape it needs to converge to, thus the spike occurs. The kriging shape after the spike is typically more complicated than the one before. In other words, there is a certain profile characteristic that is only captured by the model with enough samples. We have often observed this phenomenon, even in simpler analytical functions, and they typically converge in the end. Having multiple spikes in the convergence plot is not uncommon either. However, when the function profile is too complex, the procedure converges too slowly, just as observed here.

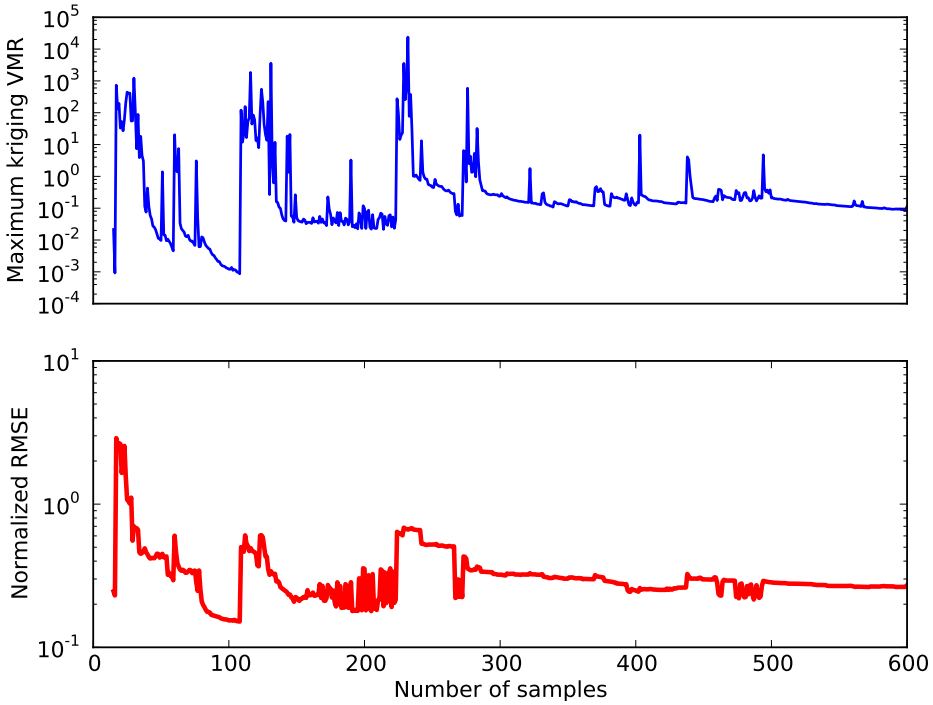


Figure 12: The slow convergence of using a global kriging model (with adaptive sampling) to approximate the complex C_D profile of the BWB configuration in a four-dimensional space. The convergence criterion for the adaptive sampling is not achieved.

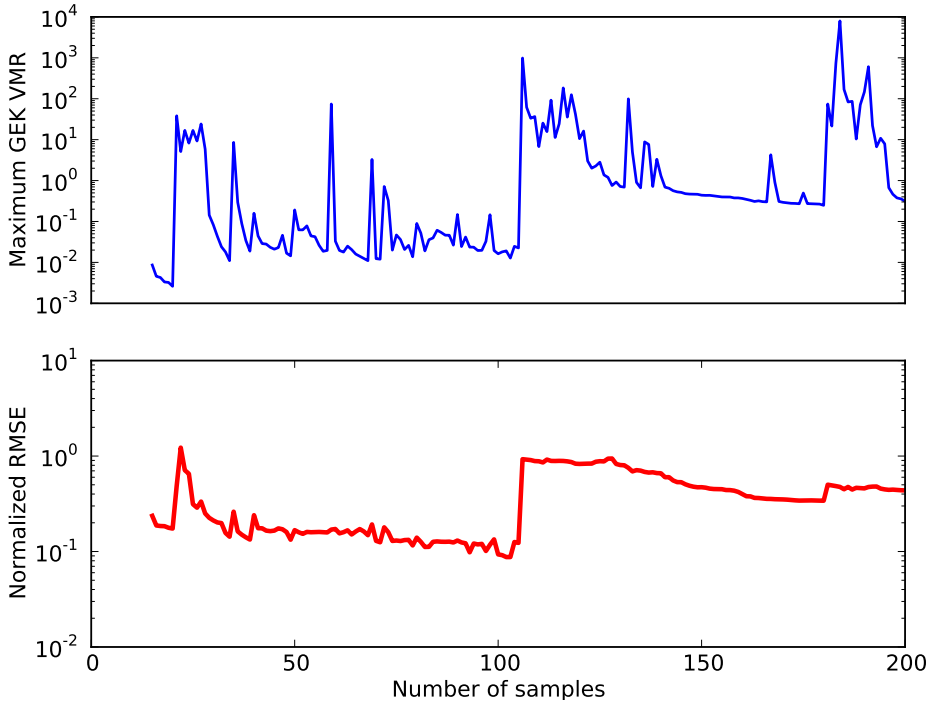
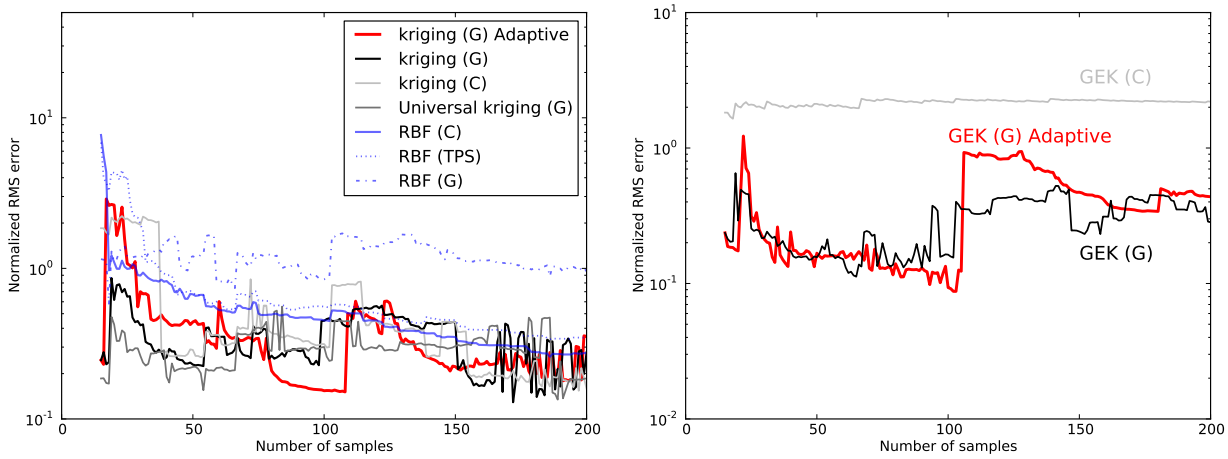


Figure 13: The slow convergence of using a global GEK model (with adaptive sampling) to approximate the complex C_D profile of the BWB configuration in a four-dimensional space. The convergence criterion for the adaptive sampling is not achieved.

The other global surrogate models are tested with up to 200 Halton samples, since we do not have converged numbers of samples with the adaptive sampling procedure. The error convergence plots are shown in Figure 14. For all the eight surrogate models tested, only two yield approximation errors that are less than 20%: the universal kriging (18.59%), and the ordinary kriging with cubic correlation function (15.85%). GEK with cubic correlation function has the worst performance, as also observed in the two-dimensional case. Thus, using any of these surrogate models in any analyses or optimizations will not yield meaningful results.

Due to the poor predictive performance of all the global surrogate models considered in this study, we now look into using ME models and see how they can improve the accuracy. The training samples to build each local expert are selected through the adaptive sampling procedure. For this problem, the convergence is considered achieved when the maximum VMR $< 10^{-3}$, and the maximum number of samples is set to 50 for each local expert. This adaptive sampling procedure starts with the first 15 clustering training data assigned to the local region, $\mathcal{T}_k$. Similarly to the two-dimensional case, using the derivative, $\partial C_D/\partial M$, as the clustering criterion yields better performance overall with fewer samples than when C_D value is used. Thus, we only show the results from the former clustering criterion here, which are summarized in Figure 15. Here we try partitioning the input space to up to seven clusters. Using kriging models as the local experts result in normalized RMSE ranging between 7% to 11%, whereas using GEK models further improves it to 4–5%. The total number of samples increases as we increase the number of clusters. When GEK models are used as the local experts, we achieve a good compromise between the number of samples and approximation error even with only 2 clusters ($N_s = 74$, with 4.73% RMSE). These results show that applying the divide-and-conquer approach in approximating a complex function profile notably improves the predictive performance. Compared to the global model training with adaptive sampling procedure that takes more than 20 hours without achieving convergence, the training time for the ME models (including clustering, adaptive sampling, and constructing the local experts) are more than 200 times faster than the global model training, using the same processor. When using kriging models as the local experts, the training is done in less than 3 minutes, whereas when GEK models are used, it is less than 8 minutes. This further supports our earlier argument that the adopted distributed approach can help reduce the computational cost to build and use the surrogate models.



(a) No gradient information

(b) With gradient information

Figure 14: The normalized RMSE plots show poor predictive performance by all the global surrogate models considered in approximating the C_D profile of the BWB configuration in a four-dimensional space. Halton sampling is used unless stated otherwise.

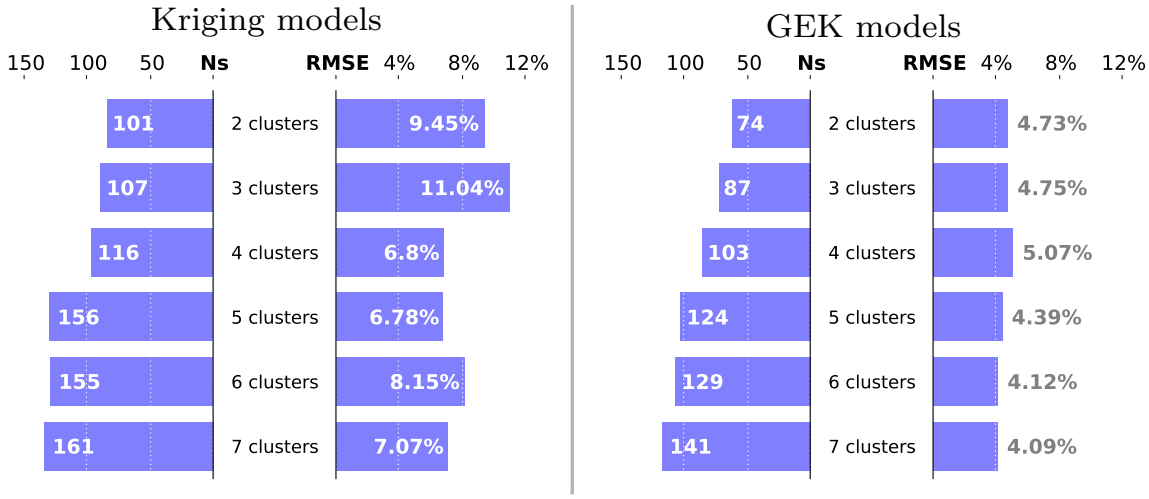


Figure 15: Mixture of experts result summary with $\partial C_D/\partial M$ clustering criterion to approximate the C_D profile of the BWB configuration.

Figure 16 shows the convergence plots for the maximum VMR and normalized RMSE for each local expert in a ME with five clusters. Each local expert is built using a GEK model. The convergence displayed in this plot shows a stark difference from those of the global models (Figures 12 and 13). Here, the adaptive sampling procedure within each local expert converges nicely until the convergence criterion is achieved, which translates to a smooth convergence of the normalized RMSE. From these results, we can see that the adopted divide-and-conquer approach overcomes the challenges of modeling a highly nonlinear function by partitioning the input space into smaller subregions that are much easier to tackle.

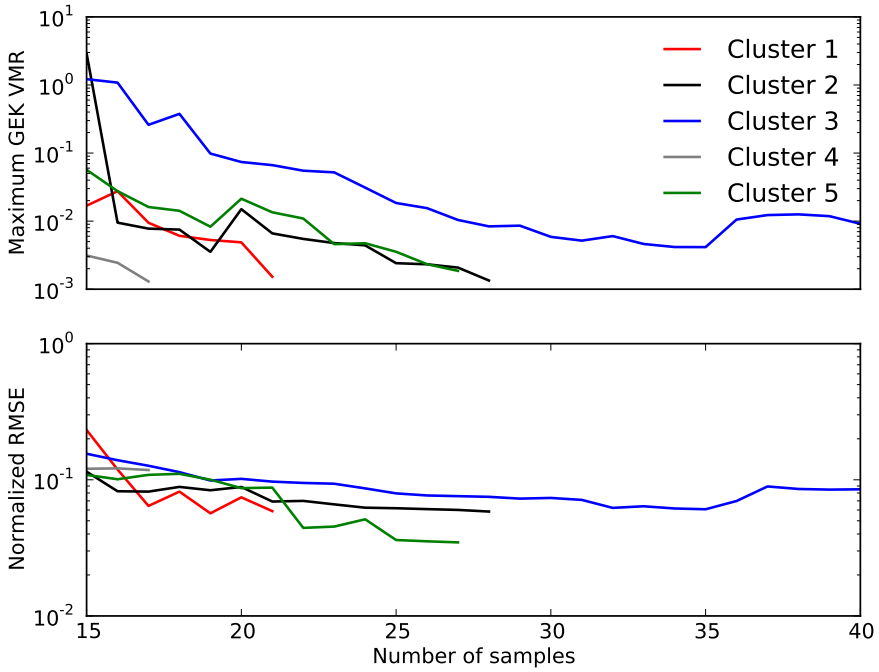
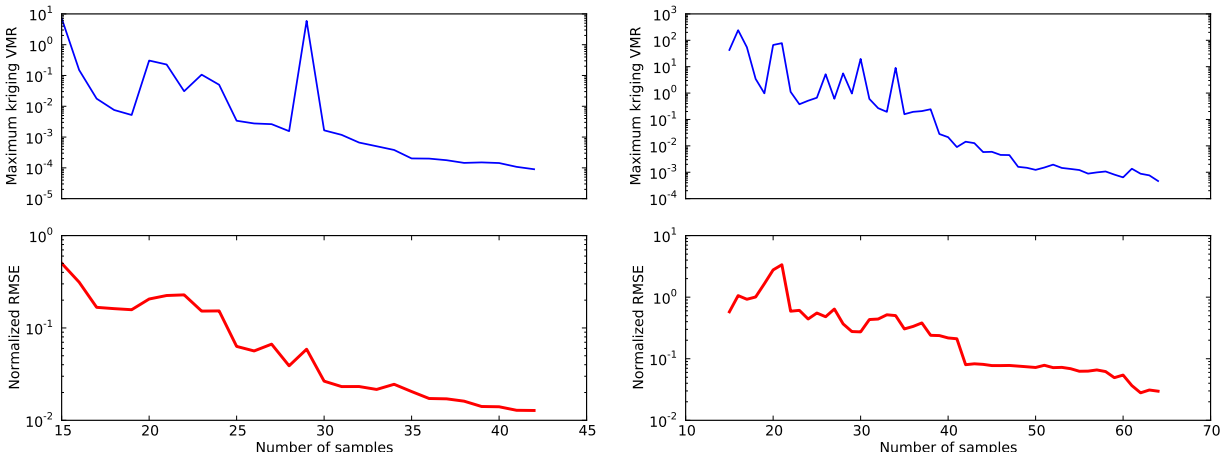


Figure 16: The convergence plots for the maximum VMR and normalized RMSE for each local expert (GEK), with the input space partitioned into 5 clusters, when approximating the C_D profile of the BWB configuration in a four-dimensional space.

Building surrogate models for C_L and C_M is much easier than for C_D, owing to the much simpler function profiles. Using our aerodynamic solver, C_L and C_M values are independent of the flight altitude, and thus their derivatives in the altitude dimension are zero. These zero derivatives impose difficulties when fitting a GEK model, thus we restrict the following discussion to surrogate models with no derivative information, which will prove to be sufficient in approximating C_L and C_M.

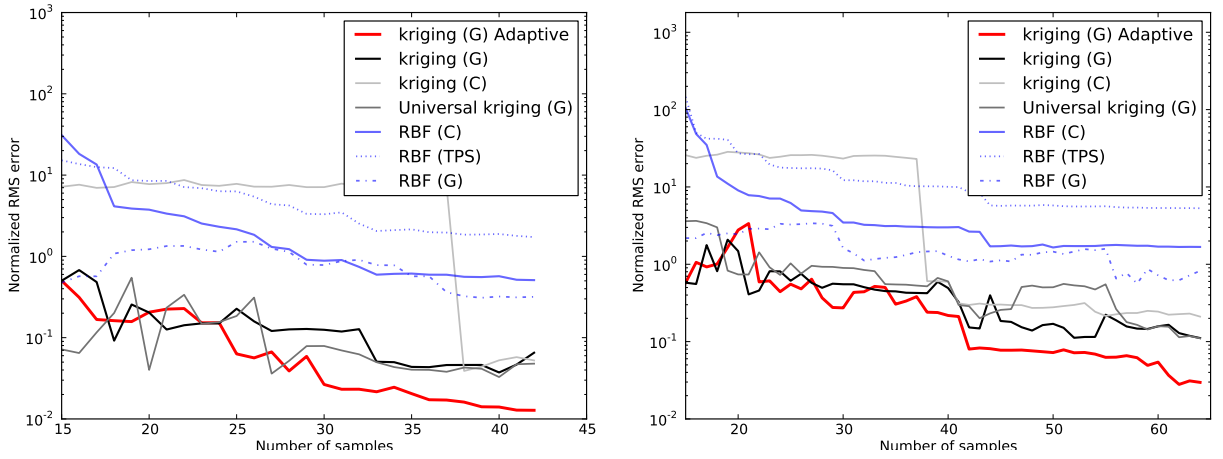
Unlike C_D, performing the adaptive sampling procedure in building a global model results in good convergence characteristics, for both the C_L and C_M kriging models, as shown in Figure 17. For the C_L kriging model, the maximum VMR decreases to below 10^{-4} with 42 samples. The resulting surrogate model gives an overall approximation error of 1.28%. With the same convergence criterion for the adaptive sampling procedure, the C_M kriging model requires 64 samples and achieves an overall approximation error of 2.97%. Using these converged numbers of samples, we now run other global models using Halton samples, as shown in Figure 18. To approximate both C_L and C_D, kriging with adaptive sampling offers the best performance. Since C_L and C_M profiles do not exhibit any strong quadratic trends in the M and α dimensions, the universal kriging (which is set to have the same basis functions as the ones for C_D) does not have any advantage over the ordinary kriging models. The kriging models with a cubic correlation function show rather poor performance and convergence at $N_s < 35$, but catch up with other kriging models for larger number of samples. The approximation accuracy of the three RBF models converge slowly, and thus at the selected N_s their approximation errors are still high.

For modeling simple profiles such as C_L and C_M, the ME models do not offer much advantage. In fact, going



(a) $\hat{C}_L$ (b) $\hat{C}_M$

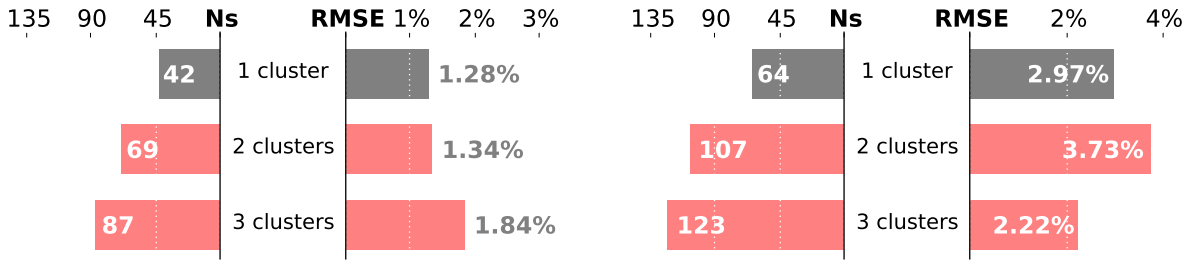
Figure 17: The convergence plots for the maximum VMR and normalized RMSE for kriging model with adaptive sampling to approximate C_L and C_M profiles of the BWB configuration in a four-dimensional space.



(a) $\hat{C}_L$ (b) $\hat{C}_M$

Figure 18: The convergence plots for the normalized RMSE for global models (with no gradient information) to approximate C_L and C_M profiles of the BWB configuration in a four-dimensional space.

from using two clusters to three clusters does not show much difference in the input space partitioning. The ME results summary for these two function profiles is shown in Figure 19, using C_L and C_M values as the clustering criterion, respectively. The adaptive sampling procedure is performed for each local expert (kriging model). From these results we can see that adding more clusters requires more total samples to build the surrogate models, with no improvement in the approximation accuracy. Therefore, when we deal with simple function profiles, global surrogate models are sufficient and the computational complexity associated with implementing the ME model is not necessary.



(a) $\hat{C}_L$ (b) $\hat{C}_M$

Figure 19: Summary for mixtures of experts applied to C_L and C_M profiles of the BWB configuration in a four-dimensional space.

The function profile is expected to increase in complexity when high-fidelity models are used to generate samples, as they can model more features not captured in low-fidelity models. Based on the results discussed in this section, we could expect that the ME models would offer even more advantages than the conventional surrogate models, though consequently additional local experts might be required to model the more complex profiles accurately.

6.3 Four-dimensional Case with CRM Configuration

The conventional CRM configuration has a simpler C_D profile than that of the BWB in the four-dimensional input space considered here, since the drag is not as strongly coupled to trim. We found that surrogate models with adaptive sampling performs much better than those with Halton sampling. Thus, we only discuss adaptive sampling in this section. Figures 20 and 21 show the convergence of maximum VMR and normalized RMSE for both kriging and GEK models. We set the maximum number of samples to be 600 for kriging and 200 for GEK. As the plots show us, the maximum numbers of samples are reached before the adaptive sampling procedures converge (maximum VMR $< 10^{-3}$). At termination, the normalized RMS errors are 20.18% and 29.02% for kriging and GEK, respectively.

We now implement the ME approach to approximate this C_D profile. Both kriging and GEK models are considered as the local experts. The clustering criterion is based on the values of $\partial C_D/\partial M$, which was previously used in the BWB case. An adaptive sampling procedure is performed for each local expert, where convergence is achieved when $\max(\text{VMR}) < 10^{-3}$. The results are summarized in Figure 22. When the ME model uses kriging as the local expert, we see an improvement in the overall approximation errors, with the lowest, 6.85%, is achieved when six local experts are used. The adaptive sampling procedures converge for all local experts, with a maximum total N_s of 135 (with seven clusters). Using GEK models as the local experts shows a notable improvement in terms of the approximation accuracy, achieving overall normalized RMS errors of approximately 5%. The total N_s increases with more clusters. As in the BWB case discussed previously, a good compromise between the sample size and accuracy is achieved even with two clusters ($N_s = 43$, with an error of 5.12%)

Similarly to the BWB case, simple global kriging models with adaptive sampling offer good predictive performance to approximate C_L and C_M of the CRM configuration. The convergence plots of the maximum VMR and normalized RMS error are shown in Figure 23. We achieve a normalized RMS error of 2.13% with 50 samples in the C_L approximation, and 1.49% error with 76 samples in the C_M approximation.

6.4 Surrogate-based Mission Analysis Benchmarking

Now we compare the performance of the different surrogate models in predicting the mission performance, which is the ultimate goal of this work. We benchmark the reference mission analysis for the BWB configuration, as described in Section 5.4. In particular, we will compare the computed mission ranges for a given fuel weight.

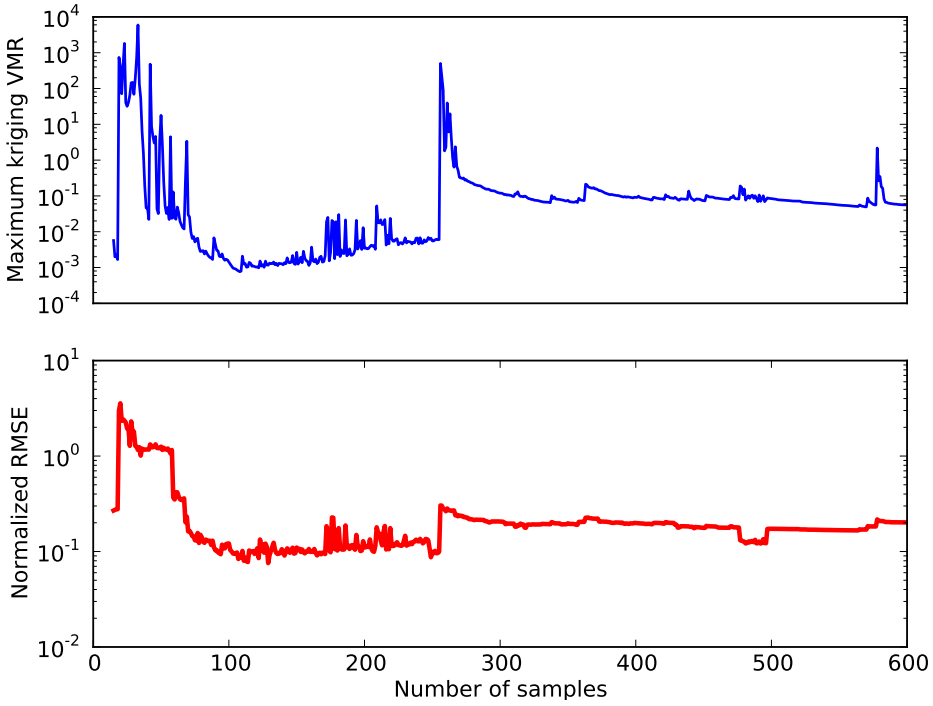


Figure 20: Using a global kriging model with adaptive sampling to approximate the C_D profile of the CRM configuration in a four-dimensional space results in 20.18% approximation error, and the convergence criterion for the adaptive sampling is not achieved.

The comparison procedure is as follows. We first run the mission analysis procedure using the selected "best" surrogate models: ordinary kriging models with adaptive sampling for C_L and C_M, and a mixture of experts with two GEK models for C_D. For this mission analysis, we start with arbitrary initial states (by assuming equal fuel weights for all segments), as well as an arbitrary fuel weight. Once converged, we find the corresponding mission ranges for the same fuel weight using an aerodynamic model in the mission analysis procedure. The mission range obtained using each aerodynamic surrogate model is then compared to that of the reference mission analysis (which uses TriPan), and the relative approximation error can then be assessed.

By comparing the computed mission ranges for a given fuel weight, we eliminate the need to iterate over the different fuel weights to achieve the desired mission range. Therefore, the residual equations for the mission analysis are only solved once. The computational cost saving matters significantly when running the reference mission analysis, since using the aerodynamic solver takes approximately $1\,700$ times longer than using surrogate models. To further accelerate the residual equation convergence, we use the final states (for the residual equations) of the previously converged mission analysis as the initial states, instead of setting them arbitrarily. We repeat this procedure for the selected 10 mission profiles listed in Table 1.

The results for this surrogate-based mission analysis benchmarking is summarized in Figure 24. The eight surrogate models used in this study are listed in the left hand side column, with each ME model has two local experts. The sampling procedure ('A' for the adaptive and 'H' for Halton) and the total number of samples (N_s) are indicated as well. The last column shows the global RMSE, which is computed based on the $10\,000$ validation points described in Section 5.3. The error bars shown in this figure summarize the performance for each surrogate model, by showing the minimum and maximum absolute range approximation errors. The absolute approximation error corresponding to each benchmark mission profile is indicated by the black circle shown on the error bar.

From these error bars, we observe that the ME approach offers a better predictive capability than other surrogate models. When using GEK models as the local experts, nine out of ten mission profiles have approximation errors of less than 1%, and one of 1.82%. When using kriging models as the local experts, seven out of ten mission profiles have approximation errors of less than 2%, and the rests are below 3.1%. The global GEK model with adaptive sampling

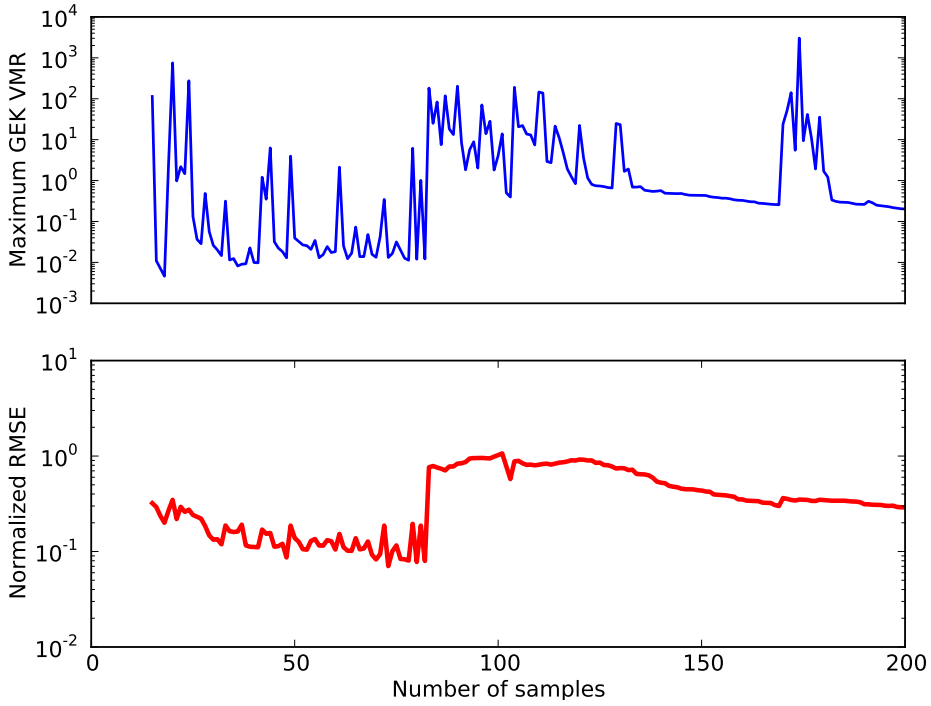


Figure 21: Using a global GEK model with adaptive sampling to approximate the C_D profile of the CRM configuration in a four-dimensional space results in 29.02% approximation error, and the convergence criterion for the adaptive sampling is not achieved.

offers the next best performance, with a maximum approximation error of 7.3%, and four of them are below 2%.

Looking at the global RMSE, which indicates the surrogate modeling accuracy in the input space, and the accuracy of mission analysis performance prediction, we can see that some surrogate models perform worse than others in predicting the mission performance, despite their better overall accuracy. For example, the ordinary kriging with Halton sampling has a global RMSE of 26.32%, which is better than that of universal kriging (36.65%). However, in predicting mission ranges, the ordinary kriging model yields absolute approximation errors of 10.78–78.33%, whereas the universal kriging model performs better with 6.29%–19.35% approximation errors. This observation suggests that the knowledge of the distribution of predictive error of a surrogate model in the input space is more valuable than knowing the global RMSE alone. Some surrogate models might have poorer accuracy in the input space subregions that are used more in the mission analysis, despite the lower global RMSE. These surrogate models thus perform more poorly in approximating the mission ranges, as compared to other surrogate models with higher global RMSE's, but with lower errors in those subregions. However, it is challenging, if not impossible, to know *a priori* the error distributions of the surrogate models, as well as the input space subregions that are most used in the mission analysis. This problem, however, is eliminated when we use the mixture of experts approach, as we can observe from the results. With the divide-and-conquer strategy, we now tackle much simpler problems locally. The consistency of good approximation accuracy is thus much easier to maintain.

The computational time required to complete the mission analysis varies for each benchmark mission. When starting from arbitrary fuel weight and initial states, the surrogate-based mission analysis procedures are completed in 6–18 minutes, using 16 processors. This computational time includes the search algorithm to find the fuel weight for the specified mission range. When the TriPan aerodynamic solver is used, the computational time ranges from 7 to 45 hours, also using 16 processors. This computational time will be even longer when we need to perform a search algorithm to solve for range, and if we start with arbitrary initial states.

The computational cost of running mission analyses without relying on surrogate models would be further exacerbated if we used finer grids or higher-fidelity models, and this would be prohibitively expensive the when performing design optimization. This finding emphasizes the importance of developing an efficient surrogate-based mission anal-

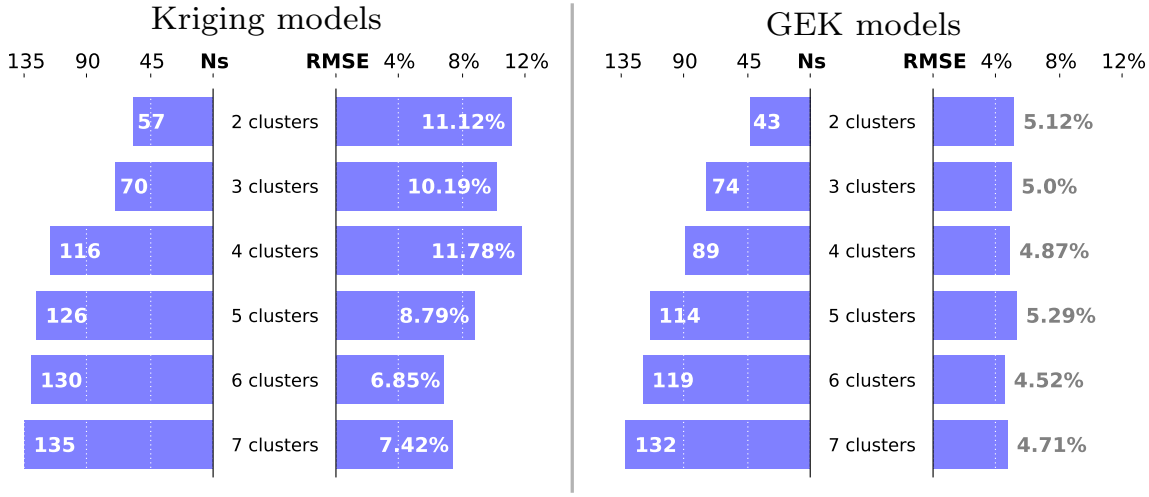
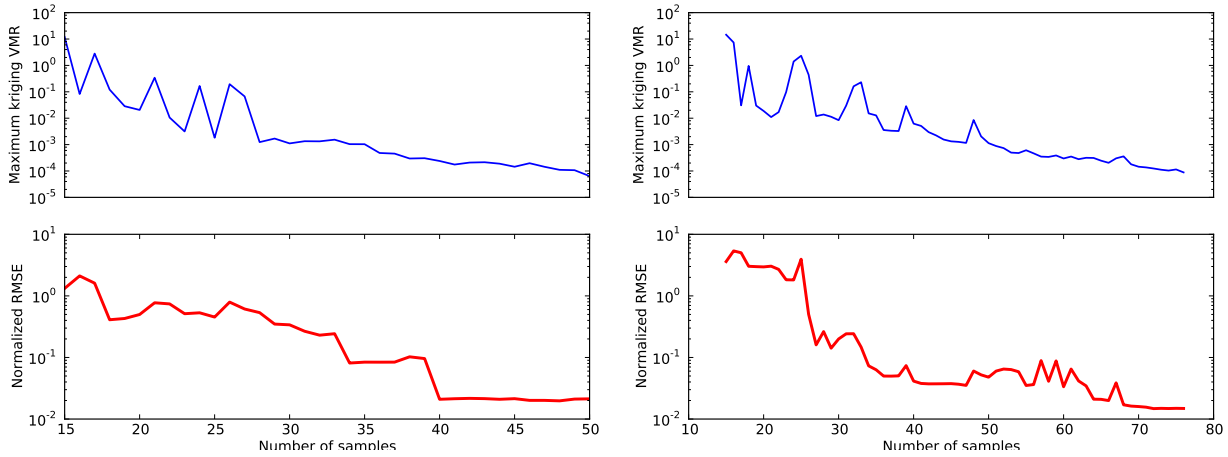


Figure 22: Mixture of experts result summary with $\partial C_D/\partial M$ clustering criterion to approximate the C_D profile of the CRM configuration in a four-dimensional space.



(a) $\hat{C}_L$ (b) $\hat{C}_M$

Figure 23: Convergence plots for the maximum VMR and normalized RMS error for kriging model with adaptive sampling to approximate C_L and C_M profiles of the CRM configuration in a four-dimensional space.

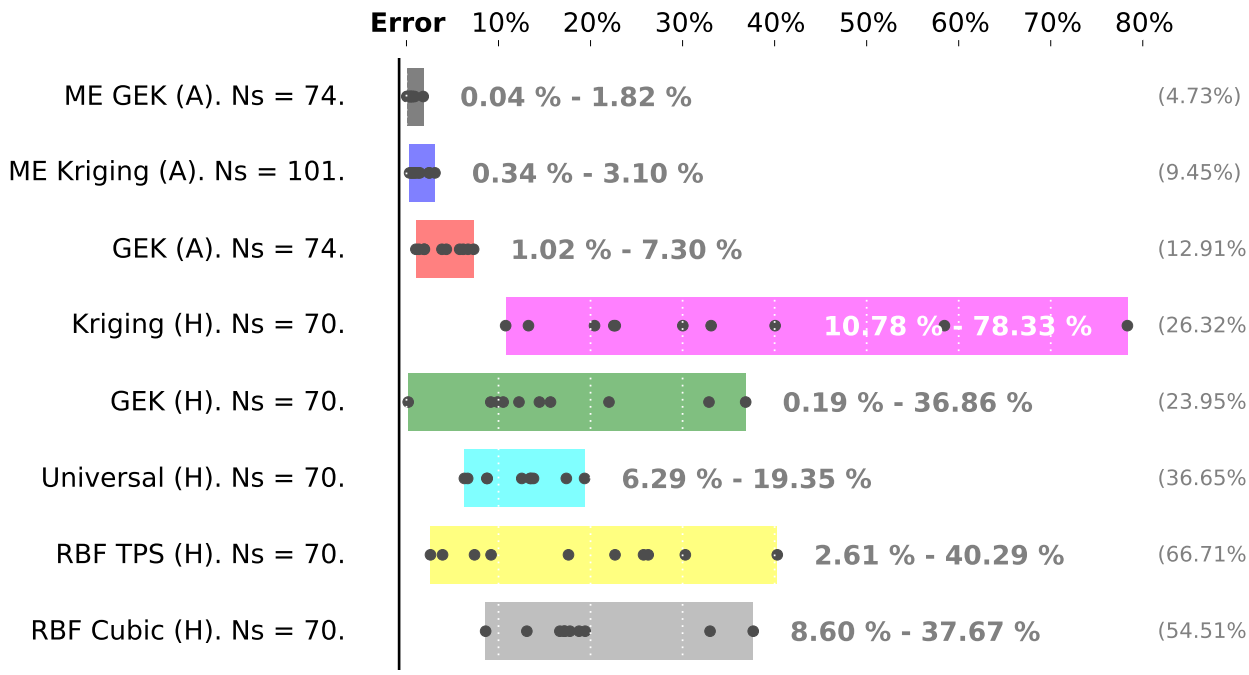


Figure 24: Error bars summarizing the range errors for each metamodel. The black circles are the errors for the ten benchmark mission profiles.

ysis, which can accurately predict the mission performance at a much lower computational cost. Our results show that the proposed ME approach is superior to global surrogate models and can offer better predictive accuracy (where it brings the range estimation error to $< 2\%$) and efficiency (more than 70 times faster) in the mission analysis context.

7 Conclusion

Motivated by the need to accurately compute aircraft fuel burn in high-fidelity mission and aerostructural optimizations, we developed a surrogate-based mission analysis procedure. The success of this procedure relies on the approximation accuracy of the surrogate models. For this purpose, we explored the use of several interpolating surrogate modeling techniques for modeling aerodynamic coefficients, for both the conventional (CRM) and unconventional (BWB) aircraft configurations. In addition to using the well-established kriging and RBF techniques, we proposed a means to combine surrogate models by adopting the ME approach. We complemented this method with an adaptive sampling procedure. While the adaptive sampling is shown to improve the accuracy of surrogate models, the convergence could be slow in some cases, particularly when modeling complex profiles. Therefore, with smaller sample budget, a simple space-filling sampling technique is a better option.

The performance of the surrogate models was assessed by computing the normalized RMSE using $10\,000$ validation points. The traditional surrogate models performed well to model the simpler C_L and C_M profiles. However, they proved to be insufficient to model the complex profile of C_D, especially in the transonic drag regime. Significant improvements were observed when we used the proposed ME approach. The divide-and-conquer approach overcame the challenges of modeling a complex terrain by partitioning the input space into smaller subregions, each with a simpler profile to model. For the four-dimensional case with BWB configuration, a good compromise between the required number of samples and predictive accuracy could be achieved with two clusters. We obtained an approximation error of 4.73% with a mixture of two GEK models (74 samples), and 9.45% when kriging models were used as the local experts (101 samples). On the other hand, the adaptive sampling procedures for the global kriging and GEK models failed to converge, yielding 27.63% and 43.68% approximation errors at termination. Moreover, the distributed approach in the ME model notably helped reduce the computational cost to build and use the surrogate models. The training times for the global kriging and GEK models (with adaptive sampling) were 22 and 32 hours (and yet they still fail to converge); these numbers were reduced to 3 and 8 minutes when using the ME models. In other words, training the ME models was more than 200 times faster than training the global models. Each local expert in the mixtures

found different optimum model parameters. This observation shows that by partitioning the input space, each local expert models the dependence between function value and inputs separately, yielding a better approximation overall.

We also assessed the surrogate modeling performance in the context of mission performance evaluation, by evaluating the corresponding range estimation errors when they were used in the mission analysis procedure. ME models again proved to be superior than the conventional global models. In particular, the range estimation errors corresponding to the ME models were less than 1.82% for the ten benchmark mission profiles when GEK models were used as the local experts, and less than 3.10% when kriging models were used. On the other hand, the several global surrogate models that were implemented proved unable to provide sufficient accuracy to produce meaningful results with RMSE values ranging from 12.91% to 66.71% and predicted mission range errors ranging from 0.19% to 78.33%. Moreover, completing the surrogate-based mission analysis was more than 70 times faster than when an aerodynamic model was used in the analysis. Based on these results we conclude that the mixture of experts technique is both necessary and sufficient to model the aerodynamic coefficients for surrogate-based mission analysis.

In this study, we used the same model type for all local experts in the ME model. However, the proposed ME approach offers the flexibility of using different models types, e.g., using RBF models in simpler subregions and GEK models to model more complex profiles in other subregions. The derived mixing proportions can still be used in such a case. Moreover, this ME approach is generally applicable to other highly nonlinear functions. The advantages of the ME approach comes with added computational complexity and more parameters to tune, such as the number of clusters and the clustering criterion. Applying the principle of parsimony, it is wise to use simple global models whenever sufficient, to avoid the unnecessary complexity that is inherent in the ME approach. The ME model would also have a broader applicability to other nonlinear problems, where the model structure (e.g., the required number of local experts) would vary with the varying complexity of the problems.

Acknowledgments

The authors are grateful for the funding provided by the Vanier Canada Graduate Scholarships. The computations were performed on the GPC supercomputer at the SciNet HPC Consortium. SciNet is funded by the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; the Ontario Research Fund—Research Excellence; and the University of Toronto. The authors would like to recognize the other members of our research group, especially Gaetan Kenway, Graeme Kennedy, Edmund Lee, and Peter Lyu for their contributions to the solvers and framework used in this work.

References

References

[1] Lee, J. J., "Can we accelerate the improvement of energy efficiency in aircraft systems?" *Energy Conversion and Management*, Vol. 51, 2010, pp. 189–196.

[2] Nidumolu, R., Prahalad, C. K., and Rangaswami, M. R., "Why Sustainability Is Now the Key Driver of Innovation," *Harvard Business Review*, Vol. 87, No. 9, 2009, pp. 56–64.

[3] Martins, J. R. R. A. and Lambe, A. B., "Multidisciplinary Design Optimization: A Survey of Architectures," *AIAA Journal*, 2013. doi:10.2514/1.J051895, (In press).

[4] Lee, J. J., *Historical and Future Trends in Aircraft Performance, Cost, and Emissions*, Master's thesis, Aeronautics & Astronautics Department and Technology & Policy Program, Massachusetts Institute of Technology, September 2000.

[5] Randle, W. E., Hall, C. A., and Vera-Morales, M., "Improved Range Equation Based on Aircraft Flight Data," *Journal of Aircraft*, Vol. 48, No. 4, July–August 2011, pp. 1291–1298. doi:10.2514/1.C031262.

[6] Roskam, J., *Airplane Design Part I: Preliminary Sizing of Airplanes*, Roskam Aviation and Engineering Corporations, Ottawa, KS, 1985.

[7] Kenway, G. K. W. and Martins, J. R. R. A., "Multi-point High-fidelity Aerostructural Optimization of a Transport Aircraft Configuration," *Journal of Aircraft*, Vol. 51, 2014, pp. 144–160. doi:10.2514/1.C032150.

[8] Yan, B., Jansen, P. W., and Perez, R. E., "Multidisciplinary Design Optimization of Airframe and Trajectory Considering Cost and Emissions," *14th AIAA/ISSMO Multidisciplinary Analysis and Optimization (MAO) Conference*, Indianapolis, IN, September 2012. doi:10.2514/6.2012-5494, AIAA 2012-5494.

[9] Simpson, T. W., Booker, A. J., Ghosh, D., Giunta, A. A., Koch, P. N., and Yang, R. J., "Approximation methods in multidisciplinary analysis and optimization: a panel discussion," *Struct Multidisc Optim*, Vol. 27, 2004, pp. 302–313.

[10] Simpson, T. W., Toropov, V., Balabanov, V., and Viana, F. A. C., "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come—or Not," *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Victoria, BC, Canada, September 2008. doi:10.2514/6.2008-5802, AIAA 2008-5802.

[11] Sobieszczanski-Sobieski, J. and Haftka, R. T., "Multidisciplinary aerospace design optimization: survey of recent developments," *Structural Optimization*, Vol. 14, 1997, pp. 1–23. doi:10.1007/BF01197554.

[12] Chung, H. S. and Alonso, J. J., "Design of a Low-Boom Supersonic Business Jet Using Cokriging Approximation Models," *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta, GA, September 2002, AIAA Paper 2002–5598.

[13] Chung, H. S. and Alonso, J. J., "Using Gradients to Construct Cokriging Approximation Models for High-Dimensional Design Optimization Problems," *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Reno, NV, January 2002, AIAA Paper 2002–0317.

[14] Toal, D. J. J. and Keane, A. J., "Efficient Multipoint Aerodynamic Design Optimization via Cokriging," *Journal of Aircraft*, Vol. 48, No. 5, September–October 2011, pp. 1685–1695. doi:10.2514/1.C031342.

[15] Zimmermann, R. and Görtz, S., "Non-linear reduced order models for steady aerodynamics," *Procedia Computer Science*, Vol. 1, 2010, pp. 165–174. doi:10.1016/j.procs.2010.04.019.

[16] Amsallem, D., Cortial, J., and Farhat, C., "Toward Real-Time Computational-Fluid-Dynamics-Based Aeroelastic Computations Using a Database of Reduced-Order Information," *AIAA Journal*, Vol. 48, No. 9, September 2010, pp. 2029–2037.

[17] Fossati, M. and Habashi, W. G., "Multiparameter Analysis of Aero-Icing Problems Using Proper Orthogonal Decomposition and Multidimensional Interpolation," *AIAA Journal*, Vol. 51, No. 4, 2013, pp. 946–960. doi:10.2514/1.J051877.

[18] Koko, F., *Aerostructural and Trajectory Optimization of Morphing Wingtip Devices*, Master's thesis, Faculty of Aerospace Engineering, Delft University of Technology, October 2011.

[19] Liem, R. P., Kenway, G. K. W., and Martins, J. R. R. A., "Multimission Aircraft Fuel Burn Minimization via Multipoint Aerostructural Optimization," *AIAA Journal*, 2014. doi:10.2514/1.J052940, (In press).

[20] Liem, R. P., Mader, C. A., Lee, E., and Martins, J. R. R. A., "Aerostructural design optimization of a 100-passenger regional jet with surrogate-based mission analysis," *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, Los Angeles, CA, August 2013. doi:10.2514/6.2013-4372.

[21] Lyu, Z. and Martins, J. R. R. A., "Aerodynamic Design Optimization Studies of a Blended-Wing-Body Aircraft," *Journal of Aircraft*, 2014. doi:10.2514/1.C032491.

[22] Laurenceau, J. and Sagaut, P., "Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging," *AIAA Journal*, Vol. 46, No. 2, 2008, pp. 498–507.

[23] Laurenceau, J. and Meaux, M., "Comparison of gradient and response surface based optimization frameworks using adjoint method," *4th AIAA Multidisciplinary Design Optimization Specialist Conference*, Schaumburg, IL, 2008.

[24] Laurenceau, J., Meaux, M., Montagnac, M., and Sagaut, P., "Comparison of gradient-based and gradient-enhanced response-surface-based optimizers," *AIAA Journal*, Vol. 48, No. 5, 2010, pp. 981–994.

[25] Journel, A. G. and Rossi, M. E., "When Do We Need a Trend Model in Kriging?" *Mathematical Geology*, Vol. 21, No. 7, 1989, pp. 715–739. doi:10.1007/BF00893318.

[26] Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E., "Adaptive Mixtures of Local Experts," *Neural Computation*, Vol. 3, 1991, pp. 79–87.

[27] Tang, B., Heywood, M. I., and Shepherd, M., "Input partitioning to mixture of experts," *Proceedings of the 2002 International Joint Conference on Neural Networks*, Vol. 1, IEEE, 2002, pp. 227–232. doi:10.1109/IJCNN.2002.1005474.

[28] Liem, R. P. and Martins, J. R. R. A., "Surrogate Models and Mixtures of Experts in Aerodynamic Performance Prediction for Mission Analysis," *15th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Atlanta, GA, June 2014, AIAA-2014-2301.

[29] Coffin, J. G., "A Study of Airplane Range and Useful Loads," NACA-TR-69, NACA, 1920.

[30] Breguet, L., "Calcul du Poids de Combustible Consummé par un Avion en Vol Ascendant," *Comptes Rendus Hebdomodaires des Séances de l'Académie des Sciences*, Vol. 177, July 1923, pp. 870–872.

[31] McCormick, B. W., *Aerodynamics, Aeronautics, and Flight Mechanics*, John Wiley & Sons, New York, US, 1979.

[32] Lee, H. and Chatterji, G. B., "Closed-Form Takeoff Weight Estimation Model for Air Transportation Simulation," *10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, Fort Worth, TX, Sept 13–15 2010. doi:10.2514/6.2010-9156, AIAA 2010-9156.

[33] Kroo, I. M., *Aircraft Design: Synthesis and Analysis*, Desktop Aeronautics, Palo Alto, CA, 1st ed., Sept 2006.

[34] Henderson, R. P., Martins, J. R. R. A., and Perez, R. E., "Aircraft Conceptual Design for Optimal Environmental Performance," *The Aeronautical Journal*, Vol. 116, 2012, pp. 1—22.

[35] PASS, "Program for Aircraft Synthesis Studies Software Package," Desktop Aeronautics, Inc., Palo Alto, CA, 2005.

[36] Roskam, J. and Lan, C. T. E., *Airplane Aerodynamics and Performance*, DARcorporation, Lawrence, KS, 1997.

[37] Martins, J. R. R. A., Alonso, J. J., and Reuther, J. J., "Aero-Structural Wing Design Optimization Using High-Fidelity Sensitivity Analysis," *Proceedings of the CEAS Conference on Multidisciplinary Aircraft Design and Optimization*, edited by H. Hölinger, DGLR, Bonn, June 2001, pp. 211–226.

[38] Maute, K., Nikbay, M., and Farhat, C., "Sensitivity Analysis and Design Optimization of Three-Dimensional Non-Linear Aeroelastic Systems by the Adjoint Method," *International Journal for Numerical Methods in Engineering*, Vol. 56, No. 6, 2003, pp. 911–933. doi:10.1002/nme.599.

[39] Martins, J. R. R. A., Alonso, J. J., and Reuther, J. J., "A Coupled-Adjoint Sensitivity Analysis Method for High-Fidelity Aero-Structural Design," *Optimization and Engineering*, Vol. 6, No. 1, March 2005, pp. 33–62. doi:10.1023/B:OPTE.0000048536.47956.62.

[40] Martins, J. R. R. A., Alonso, J. J., and Reuther, J. J., "High-Fidelity Aerostructural Design Optimization of a Supersonic Business Jet," *Journal of Aircraft*, Vol. 41, No. 3, May–June 2004, pp. 523–530. doi:10.2514/1.11478.

[41] Brezillon, J., Ronzheimer, A., Haar, D., Abu-Zurayk, M., Lummer, K., Krugër, W., and Nattere, F. J., "Development and application of multi-disciplinary optimization capabilities based on high-fidelity methods," *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Honolulu, HI, April 2012, AIAA 2012-1757.

[42] Kenway, G. K. W., Kennedy, G. J., and Martins, J. R. R. A., "Scalable Parallel Approach for High-Fidelity Steady-State Aeroelastic Analysis and Derivative Computations," *AIAA Journal*, Vol. 52, No. 5, 2014, pp. 935–951. doi:10.2514/1.J052255.

[43] Giannakoglou, K. C., Papadimitriou, D. I., and Kampolis, I. C., "Aerodynamic shape design using evolutionary algorithms and new gradient-assisted metamodels," *Computer Methods in Applied Mechanics and Engineering*, Vol. 195, 2006, pp. 6312–6329. doi:10.1016/j.cma.2005.12.008.

[44] Jin, R., Chen, W., and Simpson, T. W., "Comparative studies of metamodelling techniques under multiple modelling criteria," *Structural and Multidisciplinary Optimization*, Vol. 23, 2001, pp. 1–13. doi:10.1007/S00158-001-0160-4.

[45] Meckesheimer, M., Booker, A. J., Barton, R. R., and Simpson, T. W., "Computationally Inexpensive Metamodel Assessment Strategies," *AIAA Journal*, Vol. 40, No. 10, 2002, pp. 2053–2060. doi:10.2514/2.1538.

[46] Eldred, M. S., Giunta, A. A., Collis, S. S., Alexandrov, N. A., and Lewis, R. M., "Second-Order Corrections for Surrogate-Based Optimization with Model Hierarchies," *11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany, NY, August 30–September 1 2004. doi:10.2514/6.2004-4457, AIAA 2004-4457.

[47] Antoulas, A. C., *Approximation of large-scale dynamical systems*, SIAM, Philadelphia, 2005.

[48] Robinson, T. D., Eldred, M. S., Willcox, K. E., and Haimes, R., "Strategies for Multifidelity Optimization with Variable Dimensional Hierarchical Models," *47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Newport, RI, 1–4 May 2006, AIAA Paper 2006-1819.

[49] Robinson, T. D., *Surrogate-Based Optimization using Multifidelity Models with Variable Parameterization*, Ph.D. thesis, Massachusetts Institute of Technology, May 2007.

[50] Ahmed, M. Y. M. and Qin, N., "Surrogate-Based Aerodynamics Design Optimization: Use of Surrogates in Aerodynamics Design Optimization," *13th International Conference on Aerospace Science & Aviation Technology*, Cairo, Egypt, 26–28 May 2009.

[51] Lovison, A. and Rigoni, E., "Adaptive sampling with a Lipschitz criterion for accurate metamodeling," *Communications in Applied and Industrial Mathematics*, Vol. 1, No. 2, 2010, pp. 110–126. doi:10.1685/2010CAIM545.

[52] Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F., "Kriging Metamodels for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal*, Vol. 39, No. 12, 2001, pp. 2233–2241.

[53] Wang, G. G. and Shan, S., "Review of Metamodeling Techniques in Support of Engineering Design Optimization," *Journal of Mechanical Design*, Vol. 129, No. 4, April 2007, pp. 370–380. doi:10.1115/1.2429697.

[54] Forrester, A. I. J. and Keane, A. J., "Recent advances in surrogate-based optimization," *Progress in Aerospace Sciences*, Vol. 45, 2009, pp. 50–79.

[55] McKay, M. D., Conover, W. J., and Beckman, R. J., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, Vol. 21, No. 2, 1979, pp. 239–245.

[56] Sobol', I. M., "On the Systematic Search in a Hypercube," *SIAM Journal of Numerical Analysis*, Vol. 16, No. 5, October 1979, pp. 190–193.

[57] Halton, J. H., "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numerische Mathematik*, Vol. 2, 1960, pp. 84–90. doi:10.1007/BF01386213.

[58] Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D., "Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments," *Journal of the American Statistical Association*, Vol. 86, No. 416, December 1991, pp. 953–963.

[59] Mitchell, T. J. and Morris, M. D., "Bayesian design and analysis of computer experiments: Two examples," *Statistica Sinica*, Vol. 2, No. 2, July 1992, pp. 359–379.

[60] Krige, D. G., "A statistical approach to some basic mine valuation problems on the Witwatersrand," *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, Vol. 52, 1951, pp. 119–139.

[61] Matheron, G., "Principles of geostatistics," *Economic Geology*, Vol. 58, 1963, pp. 1246–1266.

[62] Olea, R. A., "Sampling Design Optimization for Spatial Functions," *Mathematical Geology*, Vol. 16, No. 4, 1984, pp. 369–392. doi:10.1007/BF01029887.

[63] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, 1989, pp. 409–423. doi:10.1214/ss/1177012413.

[64] Koehler, J. R. and Owen, A. B., "Computer Experiments," *Handbook of Statistics*, edited by S. Ghosh and C. Rao, Vol. 13, Elsevier Science, New York, 1996.

[65] Viana, F. A. C., Simpson, T. W., Balabanov, V., and Toropov, V., "Metamodeling in Multidisciplinary Design Optimization: How Far Have We Really Come?" *AIAA Journal*, Vol. 52, No. 4, 2014, pp. 670–690. doi:10.2514/1.J052375.

[66] Zimmermann, R., "Asymptotic Behavior of the Likelihood Function of Covariance Matrices of Spatial Gaussian Processes," *Journal of Applied Mathematics*, 2010. doi:10.115/2010/494070.

[67] Lophaven, S. N., Nielsen, H. B., and Søndergaard, J., "DACE – A Matlab Kriging Toolbox, Version 2.0." Tech. Rep. IMM-REP-2002-12, Informatics and Mathematical Modeling, Technical University of Denmark, 2002.

[68] Cressie, N., "The Origins of Kriging," *Mathematical Geology*, Vol. 22, No. 3, 1990, pp. 239–253.

[69] O'Hagan, A. and Kingman, J. F. C., "Curve Fitting and Optimal Design for Prediction," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 40, No. 1, 1978, pp. 1–42.

[70] Omre, H., "Bayesian kriging—Merging observations and qualified guesses in kriging," *Mathematical Geology*, Vol. 19, No. 1, 1987, pp. 25–39. doi:10.1007/BF01275432.

[71] Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D., "A Bayesian Approach to the Design and Analysis of Computer Experiments," Tech. Rep. ORNL-6498, National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161, 1988.

[72] Handcock, M. S. and Stein, M. L., "A Bayesian Analysis of Kriging," *Technometrics*, Vol. 35, No. 4, November 1993, pp. 403–410.

[73] Han, Z.-H., Görtz, S., and Zimmermann, R., "On Improving Efficiency and Accuracy of Variable-Fidelity Surrogate Modeling in Aero-data for Loads Context," *Proceedings of CEAS 2009 European Air and Space Conference*, London, U.K. : Royal Aeronautical Society, Manchester, UK, October 26–29 2009.

[74] Rijpkema, J. J. M., Etman, L. F. P., and Schoofs, A. J. G., "Use of Design Sensitivity Information in Response Surface and Kriging Metamodels," *Optimization and Engineering*, Vol. 2, 2001, pp. 469–484.

[75] Costa, J.-P., Rostaing, P., and Pitarque, T., "A comparison between kriging and radial basis function networks for nonlinear prediction." *International Workshop on Nonlinear Signal and Image Processing, NSIP'99.*, Antalya, Turkey., 1999.

[76] Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, 2006.

[77] Yuksel, S. E., Wilson, J. N., and Gader, P. D., "Twenty Years of Mixture of Experts," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 23, No. 8, 2012, pp. 1177–1193. doi:10.1109/TNNLS.2012.2200299.

[78] Shi, J. Q., Murray-Smith, R., and Titterington, D. M., "Hierarchical Gaussian process mixtures for regression," *Statistics and Computing*, Vol. 15, 2005, pp. 31–41.

[79] Magerman, D. M., "Statistical Decision-tree Models for Parsing," *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics*, ACL '95, Association for Computational Linguistics, Stroudsburg, PA, USA, 1995, pp. 276–283. doi:10.3115/981658.981695.

[80] Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T., "Bayesian Model Averaging: A Tutorial," *Statistical Science*, Vol. 14, No. 4, 1999, pp. 382–401.

[81] Jordan, M. I. and Jacobs, R. A., "Hierarchical mixtures of experts and the EM algorithm," *Neural Computation*, Vol. 6, 1994, pp. 181–214.

[82] Breiman, L., Friedman, J. H., Olsen, R. A., and Stone, C. J., *Classification and Regression Trees*, Wadsworth International Group, Belmont, CA, 1984.

[83] Friedman, J. H., "Multivariate adaptive regression splines," *The Annals of Statistics*, Vol. 19, 1991, pp. 1–141.

[84] Quinlan, J. R., "Induction of decision trees," *Machine Learning*, Vol. 1, 1986, pp. 81–106.

[85] Bettebghor, D., Bartoli, N., Grihon, S., Morlier, J., and Samuelides, M., "Surrogate modeling approximation using a mixture of experts based on EM joint estimation," *Structural ad Multidisciplinary Optimization*, Vol. 43, No. 2, February 2011, pp. 243–259. doi:10.1007/s00158-010-0554-2.

[86] Xing, H.-J. and Hu, B.-G., "An adaptive fuzzy c-means clustering based mixtures of experts model for unlabeled data classification," *Neurocomputing*, Vol. 71, No. 4–6, January 2008, pp. 1008–1021. doi:10.1016/j.neucom.2007.02.010.

[87] Nguyen-Tuong, D., Peters, J., and Seeger, M., "Local Gaussian process regression for real time online model learning and control," *In Advances in Neural Information Processing Systems 22 (NIPS)*, 2008.

[88] Masoudnia, S. and Ebrahimpour, R., "Mixture of experts: a literature survey," *Artificial Intelligence Review*, 2012, pp. 1–19. doi:10.1007/s10462-012-9338-y.

[89] Jain, A. K., "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, Vol. 31, 2010, pp. 651–666. doi:10.1016/j.patrec.2009.09.011.

[90] Ueda, N. and Ghahramani, Z., "Optimal model inference for Bayesian mixture of experts," *Proceedings of the 2000 IEEE Signal Processing Society Workshop*, Vol. 1, IEEE, December 2000, pp. 145–154. doi:10.1109/NNSP.2000.889405.

[91] Ueda, N. and Ghahramani, Z., "Bayesian model search for mixture models based on optimizing variational bounds," *Neural Networks*, Vol. 15, No. 10, 2002, pp. 1223–1241. doi:10.1016/S0893-6080(02)00040-0.

[92] Rasmussen, C. E. and Ghahramani, Z., "Infinite Mixtures of Gaussian Process Experts," *Advances in Neural Information Processing Systems 14*, edited by T. Diettrich, S. Becker, and Z. Ghahramani, MIT Press, 2002.

[93] Tresp, V., "Mixtures of Gaussian Processes," *Advances in Neural Information Processing Systems*, Vol. 13, 2000, pp. 654–660.

[94] Reynolds, D., "Gaussian Mixture Models," *Encyclopedia of Biometric Recognition*, Springer, February 2008.

[95] Vassberg, J. C., DeHaan, M. A., Rivers, S. M., and Wahls, R. A., "Development of a *Common Research Model* for Applied CFD Validation Studies," *26th AIAA Applied Aerodynamics Conference*, AIAA, Honolulu, HI, August 2008. doi:10.2514/6.2008-6919, AIAA 2008-6919.

[96] Kennedy, G. J. and Martins, J. R. R. A., "Parallel Solution Methods for Aerostructural Analysis and Design Optimization," *13th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Fort Worth, TX, September 2010. doi:10.2514/6.2010-9308, AIAA 2010-9308.

[97] Kennedy, G. J. and Martins, J. R. R. A., "A parallel aerostructural optimization framework for aircraft design studies," *Structural and Multidisciplinary Optimization*, 2014, (accepted for publication).

[98] Raymer, D. P., *Aircraft Design: A Conceptual Approach*, Education Series, AIAA, Washington, DC, 1992.

[99] Sadraey, M. H., *Aircraft Design: A Systems Engineering Approach*, John Wiley & Sons, Chichester, West Sussex, 2012.

[100] Loken, C., Gruner, D., Groer, L., Peltier, R., Bunn, N., Craig, M., Henriques, T., Dempsey, J., Yu, C.-H., Chen, J., Dursi, L. J., Chong, J., Northrup, S., Pinto, J., Knecht, N., and Zon, R. V., "SciNet: Lessons Learned from Building a Power-efficient Top-20 System and Data Centre," *Journal of Physics: Conference Series*, Vol. 256, No. 1, 2010, pp. 012026. doi:10.1088/1742-6596/256/1/012026.