Enhancing the handling qualities analysis by collaborative aerodynamics surrogate modelling and aero-data fusion

Zhang, Mengmeng; Bartoli, Nathalie; Jungo, Aidan; Lammen, Wim; Baalbergen, Erik; Voskuijl, Mark
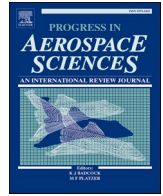
Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Enhancing the handling qualities analysis by collaborative aerodynamics surrogate modelling and aero-data fusion

Mengmeng Zhang [a],*, Nathalie Bartoli [b], Aidan Jungo [c], Wim Lammen [d], Erik Baalbergen [d], Mark Voskuijl [e]

[a] *Airinnova AB, Stockholm, Sweden*
[b] *ONERA/DTIS, Université de Toulouse, Toulouse, France*
[c] *CFS Engineering, Lausanne, Switzerland*
[d] *NLR, Netherlands Aerospace Centre, Collaborative Engineering Systems Department, Amsterdam, the Netherlands*
[e] *Faculty of Aerospace Engineering, Delft University of Technology, Delft, Netherlands*

ARTICLE INFO

ABSTRACT

In the modern aircraft design process numerical simulation is one of the key enablers. However, computational time increases exponentially with the level of fidelity of the simulation. In the EU Horizon2020 project AGILE different aircraft design analysis tools relative to different levels of fidelity are used. One of the challenges is to reduce the computational time - e.g. to facilitate an efficient optimization process - by processing the analysis data of various fidelity levels in a global surrogate model. This paper focuses on fusion of data sets via an automatic iterative process embedded in the collaborative multidisciplinary analysis (MDA) chains as applied in AGILE. Surrogate modeling techniques are applied, taking into account the optimal sampling and the corresponding fidelities of the samples. This paper will detail the different steps of the proposed collaborative approach. As a test case handling qualities analysis of the AGILE reference conventional aircraft is performed, by fusing the computed aerodynamic coefficients and derivatives. A full set of aerodynamic data computed either with different levels of fidelity or with only a low-fidelity tool has been derived and evaluated. The data set with multiple levels of fidelity significantly improved the accuracy of the flight performance analysis, especially for the transonic region in which the low fidelity aerodynamic method is not reliable. Moreover, the test case shows that by combining a collaborative surrogate modeling approach with fusion of the data sets, the fidelity of the analysis data can be significantly improved giving maximum relative prediction error less than 5% with minimal computing efforts.

1. Introduction

Today's latest large airliners in service, e.g., the A-380 and A-350, have over 6 million of parts and have 150000 design changes per year [1]. They show both outstanding performance and a reduced environmental impact. This is the result of a continuous optimization of the air-frame design and a progressive mastering by engineers of the interactions between systems and disciplines on the same configuration [2]. Aircraft design nowadays is a mature process based on an integrated approach, to handle the complexity of the product. However, given the level of maturity of the classical "tube and wing" configuration, evolving business models for developing new aircraft only by improving upon existing practices, has the consequence that engineers have small margins for improvements and future targets will be difficult to attain. Therefore, research centers and industry need to work together in the exploration of new integrated design concepts that can provide a disruptive approach and offer in this manner various possibilities for safer and greener vehicles. The current design of aircraft is an extremely interdisciplinary activity incorporating simultaneous consideration of complex, tightly coupled systems, functions and requirements. The design task is to achieve an optimal integration of all components into an efficient, robust and reliable aircraft with high performance that can be manufactured with low technical and financial risks, and has an affordable life-cycle cost.

To help achieve this goal, today's aeronautics market requires a

---

Nomenclature

AGILE	Aircraft 3rd Generation MDO for Innovative Collaboration of Heterogeneous Teams of Experts
API	Application Programming Interface
ACARE	Advisory Council for Aeronautics Research in Europe
SRA	Strategic Research Agenda
CAD	Computer Aided Design
CFD	Computational Fluid Dynamics
CSM	Computational Structural Mechanics
CPACS	Common Parametric Aircraft Configuration Schema
DOE	Design Of Experiments
DACE	Design and Analysis of Computer Experiments
EIF	Expected Improvement Function
FFD	Free-Form Deformation
FTP	File Transfer Protocol
IP	Intellectual Property
LES	Large-Eddy Simulation
MDA	Multidisciplinary Analysis
MDO	Multidisciplinary Design Optimization
MOE	Mixture of Experts
PHALANX	Performance, Handling Qualities and Loads Analysis Toolbox
PIDO	Process Integration and Design Optimization
RANS	Reynolds Average Navier-Stokes
(R)MSE	(Root) Mean Square Error
RSM	Response Surface Model
RCE	Remote Component Environment
SM	Surrogate Model
SMR	Surrogate Model Repository
S & C	Stability and Control
VLM	Vortex Lattice Method
XML	eXtensive Markup Language
csv	Comma Separated Values
lo-fi	Low fidelity (data)
hi-fi	High fidelity (data)
sp	Short Period mode
ph	Phugoid mode

Symbols

α or AoA	Angle of Attack [deg]
β	Sideslip Angle [deg]
δ	Elevator deflection angle [deg]
C_p	Pressure Coefficient [−]
ω	frequency [rad/s]
z	damping ratio [−]

manufacturer to develop more complex products by exploiting all opportunities provided, e.g., high fidelity multidisciplinary tools and frameworks and powerful High Performance Computing (HPC) capabilities [3]. Over the past 25 years the aeronautical industry has been transforming its operations to what can be called a *system-oriented approach,* instead of a functional-oriented approach. The primary focus becomes one of *integration of all disciplines.* To help meet the challenge, the core activities in the new paradigm, against the traditional analysis activities considered as "design-verification activities", now become to numerically "flight-test" a virtual aircraft with all its multi-disciplinary interactions in a computer environment as given by its different databases (aerodynamic, loads, stability and control) [1–3]. This new development paradigm has been identified in the ACARE Strategic Research and Innovation Agenda (SRIA) [4] as a major challenge to make the design of a complete virtual aircraft up to certification a reality. Several outstanding EU Framework RTD projects have addressed the notion of developing conceptual design methods with advanced close coupling of the aerodynamics, structures and flight control disciplines upon a series of evolutionary MDO systems. The FP6 STReP SimSAC (Simulating Aircraft Stability And Control Characteristics for Use in Conceptual Design) [5,6], VIVACE, CRESCENDO [7,8] and TOICA are some of those projects at both research and industry levels. Keep in mind, that current engineering analysis methods in designing complex systems rely very heavily on the knowledge and intuition of the individual designer. There still exist *information gaps* between the specialist component designs and the system level [2]. The challenge is to bridge the gaps and mitigate their effects so that the design process becomes smoother and more efficient.

The way to solve this is to establish a robust, collaborative and integrated framework with all disciplines involved to carry out the real MDO. Ciampa and Nagel [1] have identified that major obstacles in the current generation of MDO systems are related to the efforts required to setup complex collaborative frameworks. Up to 60–80% of the project time may be necessary to setup such a process. The 3rd generation MDO systems will rely on the integration of expertise in the collaborative optimization and knowledge formalization of processes and disciplinary domains. The EU funded H2020 research project AGILE, Aircraft 3rd Generation MDO for Innovative Collaboration of Heterogeneous Teams of Experts [9] builds upon the projects mentioned above. It addresses the

challenges by developing a next generation of aircraft MDO processes that target significant reduction in aircraft development costs and time to market. AGILE has formulated a novel design methodology, the so-called "AGILE Paradigm", which accelerates the deployment of collaborative, large-scale design and optimization frameworks.

The "AGILE Paradigm" emphasizes collaboration and integration by building a smooth design and analysis tool chain to carry out MDA (Multi-Disciplinary Analysis) processes. The MDA process shall includes:

(1) the collaboration and integration of the tools, data and outputs from different disciplines;
(2) the collaboration and integration of the analyses from one discipline from different tools of various fidelities. One major discipline - addressed in the paper - concerns the construction of the aerodynamic database for stability and control analysis.

Assessment of aircraft maneuverability and agility at the conceptual design stage brings great challenges in the design process regarding the stability and control analysis over the entire flight envelope. A large look-up table of forces and moments must be constructed by Computational Fluid Dynamics (CFD) while we have to address the computational cost. A useful look-up table for stability and control analysis, the so-called aerodynamic database, needs thousands of entries because of the high dimensionality of the parameter space.

"Brute-force" calculation would be far too costly. But there are ways to reduce the computational time. The first step is to use different CFD methods, from Large-Eddy Simulation via Reynolds-Averaged Navier-Stokes models, down to potential flow models, or even empirical methods from handbooks. Using the simplest method and geometry compatible with the level of accuracy required for each flight state can dramatically reduce the computational cost.

The whole spectrum of computational models is widely used in modern aerospace industry. In order not to lose too much accuracy while saving computational cost, the simplifying assumptions made to solve the standard Navier-Stokes equations should be acceptable for each single entry. For example, for low speed and non-accelerated small angle of attack flight, the in-compressible potential flow models can give acceptable predictions with significant time saving compared with Euler or RANS models.

The second step concerns all the data obtained by the different CFD methods over the entire flight envelope. In the database, dense low-cost and low-fidelity data indicating the trend must be *fused* with sparse high-cost and high-fidelity data correcting the low-fidelity data values [10].

The multi-fidelity modeling method is a means to navigate the flight state parameter space and cover the entire flight envelope in an effective way. A procedure is devised to integrate the results from different simulation tools based on different methods and having different complexity levels. Building the multi-fidelity model involves two main steps:

1. Populate the aerodynamic database over the whole flight envelope by dense low-fidelity data samples;
2. Correct the data using sparse high-fidelity samples.

A method is developed for building a reliable aerodynamic database using low-fidelity data and with a minimum number of high-fidelity samples including automatically chosen new sample points. The term *data fusion* refers to integrating data from multiple data sources to combine them into a single, comprehensive model [11]. It is a useful technique in many fields. This paper addresses *aero-data fusion* by using the aerodynamic data-sets obtained from different sources (e.g. CFD tools) to produce an efficient and effective surrogate model with minimal computational cost.

Data fusion builds surrogate models based on data fitting when data is not easily directly to obtain. The important issue is the accuracy of the model, namely, how well the surrogate model represents the real data/model. The surrogate model can be used for many cases, and the uncertainties analyzed to examine the accuracy of the surrogate model.

This paper focuses on the establishment of a collaborative architecture (CA) for aerodynamic data fusion by generating surrogate models and the handling qualities analysis by the fused data. Section 2 describes the AGILE collaborative framework and the different aerodynamic computations. Section 3 reviews the aero data fusion work done before through a number of previous EU projects, and their similarity and differences compared with the work in AGILE. The theoretical basis for constructing a surrogate model is reviewed. Section 4 details how to establish the data fusion technology through collaborative surrogates. Section 5 discusses the means of the choice of design of experiments (DOEs). Applications of data fusion in the AGILE framework are presented in Section 6, and Section 7 summarizes and concludes the work.

2. AGILE project overview

AGILE is an EU funded Horizon 2020 project coordinated by the Institute of Air Transportation Systems of the German Aerospace Center (DLR). Its objective is to implement the 3rd generation of multidisciplinary optimization through efficient international multi-site collaboration in overall design teams. The 19 partners bring different knowledge and competences regarding aircraft design and optimization. In this context the use of surrogate models is interesting to efficiently assemble results coming from various sources and levels of fidelity.

Data fusion aims to handle the large amount of data to be computed for each design cycle in the Multidisciplinary Design Optimization (MDO) process by constructing *surrogate models* from the data produced by variable fidelity analysis tools in the MDO framework. Methods like kriging, co-kriging, and adaptive modeling techniques are used for *fusing* the outputs of each tool. This technique is well established (see for instance the recent review [12]) and constitutes its own challenges in applying multi-fidelity MDA in AGILE.

This paper uses the nomenclature L* to indicate the fidelity Level of a simulation model:

L0: empirical or handbook methods;
L1: linearized-equation methods, e.g. linear aerodynamics, or the vortex lattice method (VLM);

L2: higher fidelity equation solvers with less simplifications, nonlinear-equation methods (e.g. Euler equation solvers);
L3: highest fidelity equation solvers (e.g. RANS equations).

The following part of this section will show the tools and mechanisms to be used to establish the collaborative surrogate models within the AGILE framework.

2.1. Tools for establishing the collaborative workflow

This section introduces the collaboration tools that were developed and applied and that support the AGILE collaborative surrogate modeling.

2.1.1. Collaborative work language CPACS
During the collaborative and integrated design process, data need to be exchanged. Managing the interconnections is complex and prone to errors. Adoption of a standardized, data-centric scheme for storage of all data improves consistency and reduces risks of misconceptions and errors in the process. It however requires an initial effort to make interfaces between analysis modules and the data archive. The CPACS (Common Parametric Aircraft Configuration Schema) [13–15] is widely used in the frame of AGILE.

CPACS is an XML-based representation of an aircraft design. In AGILE the aircraft design is contained in a single CPACS file. The analysis tools available from the various partners must read the input from the CPACS file and produce the output in the form of aggregating the data in the updated CPACS file. Making a tool available for sharing in AGILE requires wrapping the tool to map CPACS onto the tool's legacy input and output. The CPACS tools TIGL and TIXI [15] support integration, use and inspection of CPACS information.

2.1.2. RCE
The RCE integration environment and workflow manager [16] implements the sequence of analysis modules and manages the local data exchange and translation as well as logging the process. RCE makes it easy to set up and run a local workflow also using modules in which the engineers are not discipline-experts. Creating collaborative workflows as collections of local RCE workflows is done via Brics [17], which allows to "call" a module or a tool (and so produce results) remotely on "the specialist" computer. The remote specialist receives a request for some calculation or analysis. The input is generally a CPACS file containing all the information required. The new data generated are added to the CPACS file and sent back to the requester.

2.1.3. Brics
The interconnection mechanism available in AGILE is Brics [17,18], developed by NLR. Brics provides technology for interconnecting PIDO environments and for defining and streamlining workflows that cross organizational borders, while complying with the security constraints and dealing with the security measurements of the collaborating partners. Brics comprises protocols and middleware that facilitate remote execution of sub-processes from within a process, independent of the local PIDO environment (i.e., workflow manager) being used. Brics is based on a "single-task" protocol that arranges the execution and data flow between an orchestrating ("master") process in one organization and a remote ("slave") sub-process in another organization under control of a specialist who is notified to start the sub-processes. To cater for iterations, Brics supports the notion of a "multi-task" protocol, enabling a remote specialist to easily deal with series of similar sub-processes. Brics also supports easy experimentation with different set-ups of collaborative scenarios to support the Design Campaigns and configuration of services involved. Its nonintrusive character facilitates easy integration with existing COTS as well as in-house developed tools and solutions. It enables the AGILE partners to experience collaborative scenarios.
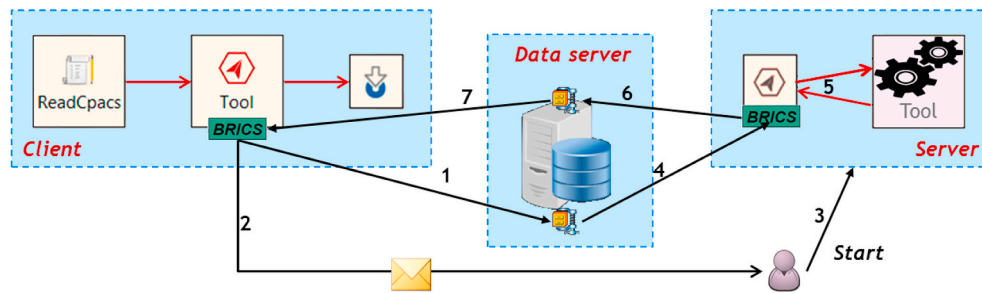
Fig. 1. Schematic overview of Brics application in AGILE from within a client workflow and a server workflow, to accomplish remote tool execution [20].
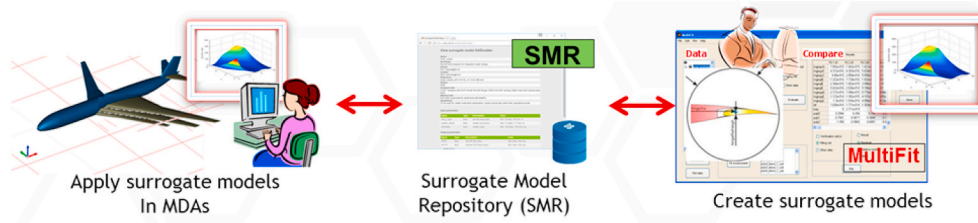

Fig. 2. The Surrogate Model Repository as library for sharing surrogate models and as broker for creating surrogate models on demand.

2.1.4. AGILE collaborative architecture (CA) steps

Ciampa et al. [18] reviewed the whole collaborative architecture used to support AGILE framework, which is also supporting the collaborative data fusion technology. The Collaborative Architecture deployment steps are as follows [18]:

1. CPACS compliance for all tools;
2. Process integration by RCE;
3. Provide accessibility by Brics.

All the tools used in CAmust be made CPACS compatible, in other words, the integration framework (RCE) must support different fidelity tools per discipline and tools for different disciplines. Zhang [19] shows the study of CPACS compliance for the variable fidelity aerodynamic tools, with one of the applications being data fusion. Within AGILE frameworks CPACS is used as common data format for tools interaction. Usually the workflow (including tools operated by different specialists) is integrated into some workflow environment, such as RCE, with tools callable using Brics.

Fig. 1 shows the Brics application from a client workflow and a server workflow, to enable a specialist to respond to a request to run the "tool", and to accomplish remote tool execution. The (input) CPACS file is downloaded (from a server of e.g., a customer) using Brics, and it is sent to the "Tool" operated by a local specialist. After execution, the (output) CPACS file is sent back and uploaded (to the customer who calls for this service) using Brics. Since the data fusion service involves many tools those are not operated by a single site, the communication between the tools is made by Brics calls as illustrated in Fig. 1. The numbered arrows indicated the actions. First, the input files for the remote service are uploaded to the central data server in a neutral domain (1). Next, the remote specialist gets notified (2), who in response may start the service (3). The service first retrieves the input files from the data server (4), runs the tool that implements the service (5), and uploads the output files to the data server (6). Finally, the output files are downloaded to the client's side (7), and the client workflow continues. For proof of concept, partners in AGILE have formed teams to apply data fusion as part of RCE workflows via Brics, which will provide data fusion solutions for the whole AGILE MDA system. More details about the AGILE collaborative approach can be found in Refs. [18,20,21].

2.2. Surrogate Model Repository

Many surrogate models of various types have been created in the AGILE project, e.g. to support reuse of knowledge and models, efficient optimizations and partner collaborations. The question then arises how to manage, share and deploy these models which must be applied with care. Bounds on the allowed input space for the surrogate model need to be clearly specified, e.g. to avoid extrapolation. Furthermore the prediction accuracy of the outputs must be specified, so that the user has a clearer idea of its applicability, quality and limitations. To address these aspects a specific Surrogate Model Repository (SMR) has been developed as part of the AGILE Development Framework. The SMR is a central broker for registration, storage, deployment and usage of surrogate models so that these may be shared and reused in collaborations in a managed way. SMR development details can be found in Refs. [20,22].

With respect to the sharing of surrogate models two cases have been considered in the frame of AGILE and are supported by the SMR.

- Full share: Share all of the compiled binary code of a surrogate model (SM) to support its use by others. In this case the complete SM implementation, e.g., an executable program such as Mixture of Experts (MOE) is uploaded to the SMR with meta-information describing its usage. The SM may be used by downloading its code and running it.
- Partial share: Share only the usage of a SM, others may use the SM 'as a service' while the code remains at the owner's or developer's site. In this case only the meta-information of the SM is uploaded to the repository. The SMR provides a user interface to directly use this SM by calling the remote service.

In addition to the sharing capabilities, the SMR can export a neutral XML format (i.e., CMDOWS [21]) that supports usage of the SM as part of a workflow system without further intervention of the SMR. As such the SMR can interface to other components in the AGILE Development Framework. An illustration of this export facility is described in Ref. [23]. The SMR may also function as a broker with respect to the creation of surrogate models. The user of the SMR may indicate that a specific surrogate model instead of a high-fidelity model must be used within an analysis. If the desired surrogate model does not exist already, the SMR initiates a surrogate model creation process behind the scenes. The role of the SMR both for sharing and for triggering the creation of
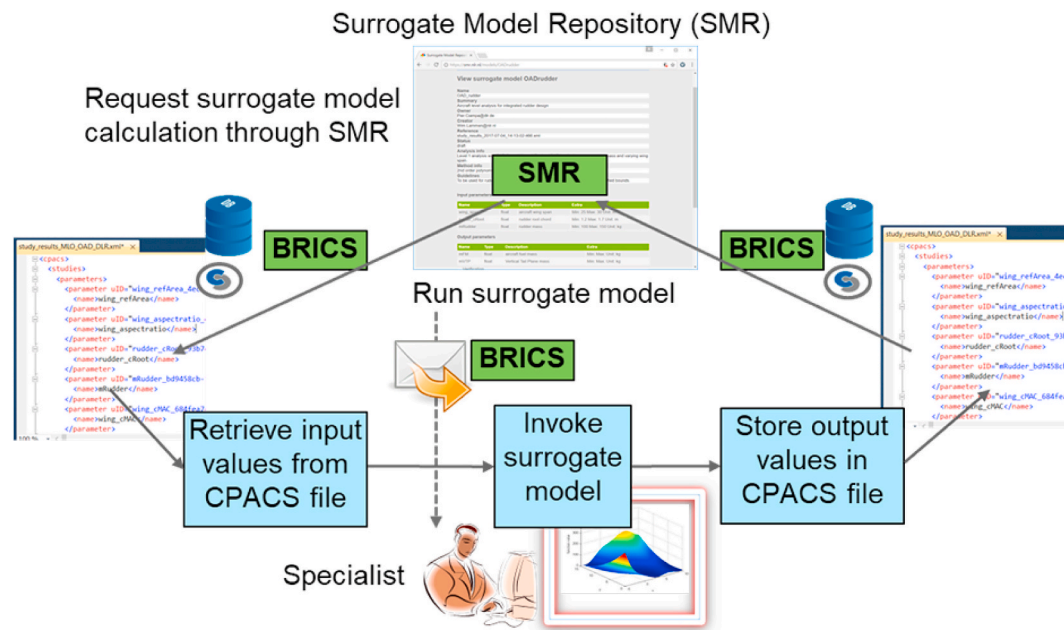
Surrogate Model Repository (SMR)

Fig. 3. Depiction of the "under the hood" process for an "as a service" calculation.

surrogate models is depicted in Fig. 2.

Section 6.2.2 will describe an example of surrogate model deployment through the SMR, using a surrogate model that was developed with the *data fusion* service. A partial share is applied: the SM is provided "as a service". This process can be semi-automatic, which is depicted in Fig. 3, and facilitated by Brics, a common building block for the realization of collaborative workflows and part of the AGILE development framework [20].

3. Review of the aero *data fusion* applications

Using CFD frameworks for aero-table generation and model assessment is not new. In this section the authors will review the work have been done before through a number of EU projects for aero-table generation using surrogate models, or aero-*data fusion*. The similarities and differences compared with the current work (within AGILE framework) will also be discussed.

Zhang [24] presented a method for efficient creation of the aero-dynamic database for the X-31 experimental aircraft, from low fidelity (Euler) and high fidelity (RANS) CFD sources. The challenge is to make reasonable predictions at extremes of the flight envelope. A co-Kriging interpolation model is used for aerodynamic moments, forces and span loads, with a "decision support system" based on the root mean squared error (RMSE) and expected improvement function (EIF). Proper Orthogonal Decomposition for data reduction is used to predict aero-loads. This work was done within FP7 project ALEF, Aerodynamic Loads Estimation at Extremes of the Flight Envelope [25].

Da Ronch et al. [26] have constructed the aero-table used for flight dynamics by kriging-based surrogate model based on CFD computations. The aerodynamic models are of various fidelities from semi-empirical prediction method to CFD by different solvers. The sampling space is enriched by iterative sampling based on the root mean squared error (RMSE) and expected improvement function (EIF) which is discussed as the "fill-in strategy" in Ref. [27]. Five test cases are shown for aero-table generation and the flight handling qualities analysis. The work provides a quite complete reference for the aero-table generation technology by data fusion of various fidelity aerodynamics. However, the aerodynamic models with different levels of fidelities are prone to data-loss by communicating with different tools/solvers.

Ghoeryshi et al. [28] solved this mis-matching by addressing the

geometry and mesh treatment using a high level conceptual aircraft description, i.e., a so-called XML-aircraft description used in software CEASIOM (the Computerized Environment for Aircraft Synthesis and Integrated Optimization Methods) [6,29]. Kriging was used to construct all of the aero-table entries and co-kriging was used to update the aero-table with additional samples calculated. The sampling method used is to identify non-linearities in the force and moment tables. The flight handling qualities are thus predicted and analyzed by the Simulation and Dynamic Stability Analyser (SDSA) [30].

Similar work was done in Ref. [31] for the Standard Dynamic Model (SDM), with filling up the tables by co-kriging [27]. For example, to construct the lateral coefficients in the baseline table, the handbook code DATCOM [32] is used to generate the initial static tables as low fidelity data, and co-kriging with a few Euler results are used to update the tables. The prediction of the aerodynamic forces from the static tabular model and the slow motion replay show close agreement as expected. The slow motions used are trimmed level flight, pull-ups with constant and varying angle of attack, wingover and 90-deg turns. Disagreement appears at fast maneuvers with the presence of unsteady terms. The work shows the usefulness of the limits of the tabular aerodynamic models by co-kriging surrogates.

The work has been extended at the University of Liverpool, and applied to 3 test cases, namely, the Standard Dynamic Model, the Ranger 2000 jet trainer, and the Stability and Control Unmanned Combat Air Vehicle [33]. For each case, a tabular aerodynamic model is constructed based on CFD predictions nd a kriging interpolator to populate the tables, with validation against wind tunnel experiments and flight test measurements.

Da Ronch et al. [34] have investigated the methodology for aero-loads generation by surrogates and applied to the Transonic Cruiser aircraft (TCR). The basic rule is to predict an effective model with a minimum number of samples (namely, minimal computing time). Kriging is used to predict the surrogate model. Two design of experiments strategies were investigated. The first one is a traditional Latin Hypercube approach, and the second strategy is the adaptive design of experiments (ADOE) technique. The result shows that ADOE provides better predictions especially for complex and non-linear engineering phenomena, such as pitch moment coefficient and aerodynamics at transonic speeds. The ADOE approach is adapted into the DOE technology for the current work, details can be seen in Section 5.

Table 1
Projects.

	AGILE (2015–2018)	ALEF (2009–2012)	SimSAC (2006–2009)
Aero. Models	CPACS	Standard CAD	CEASIOM-based XML
CFD Tools	Multi-Fidelity CPACS-compliance	Mid-to-High Fidelity	Multi-Fidelity CEASIOM-compliance
Surrogate Models	various surrogates co-kriging/MoE/etc.	External co-kriging	External co-kriging
CA platform	RCE & Brics can across firewall	–	CEASIOM
Collaborative Process features	Brics, SMR PIDO at diff. sites/locations	via Emails & FTP etc	via Emails & FTP etc.

The work by Da Ronch [26], Ghoeryshi [28,31] and Vallespin [33] within the FP6 project SimSAC started in year 2006, Simulating Aircraft Stability And Control Characteristics for Use in Conceptual Design [5,6]. It was applied to various aircraft models, the Transonic Cruiser model (TCR), an Asymmetric aircraft model, the DLR-F12 model, a B747-like model, the Ranger 2000 aircraft, the Standard Dynamic Model (SDM), and the Stability and Control Unmanned Combat Air Vehicle (SACCON). Table 1 summarizes the data fusion from some of the previous EU projects mentioned above as well as the current AGILE projects. As AGILE is built upon the previous projects, the data fusion technology is improved in a number of aspects with the outcomes used for handling qualities analyses. The outstanding features are to bring it into the collaborative architecture, with all the tools loosely coupled and integrated by RCE and accessible by Brics (see Section 2.1). The DOEs strategies (see Section 5) have been improved, and data can be stored and retrieved through SMR, see Section 2.2. The application will be discussed in Section 6.

4. Data fusion through collaborative surrogates

4.1. Surrogate models overview

A surrogate model is a cheap-to-evaluate function $\hat{y} = g(x)$ approximation to the exact but expensive-to-evaluate function $y = f(x)$. Another name is *Response Surface Model* (RSM). A used-to-be-well-known example is the table of logarithms which reduces arduous exact manual calculation to much quicker approximate calculations by repeated table lookup, interpolation, and addition.

In this paper data fusion means the integration of results from different simulation models into a single surrogate model. In AGILE the surrogate models are constructed by different methods provided by partners. In the context of this paper the RSM function is specified as:

$$[\text{force and moment } coefficients] = \text{RSM}(\alpha, \beta, Alt, Mach, \ldots, D) \tag{1}$$

where the aerodynamic variables are α the angle of attack, β the sideslip angle, Alt the flight altitude and the Mach number. The data set is denoted by D.

Surrogate modeling has two distinct steps, first the *training* to produce the $g-$ function, and second, the use of the generated g-function. The data set D is the training set. In this paper y represents the computations of forces, moments, structural deformations and stresses, etc., associated with a particular flight state of a defined aircraft configuration which is the x. Evaluation of y requires at least a flow solution, and possibly a complete aero-servo-elastic simulation. The training should produce a RSM with maximal accuracy and minimal cost is required for producing the training set. This is done by the choice of f (multi-fidelity modeling characterized by fidelity level, L0–L3) and by the choice of a training set. Much effort has been devoted to algorithms for the choice of an optimal training set, as described in Ref. [35].

Evaluation of a surrogate model requires the input x and the corresponding results of the training, e.g. a set of polynomial coefficients, or a look-up table with the whole flight envelope (flight states, or x), processed in to the function g:

$$\hat{y}(x) = g(x, D) \tag{2}$$

It is assumed that the evaluation will be made for many (thousands) values of x for a single D. The training, including the choice of the training set, is an iterative process. In a single iteration k, an algorithm M increments/fuses the training set with a set $\{x, y\}_k$ and updates D:

$$D_{k+1} = M(D_k, \{x, y\}_k) \tag{3}$$

For the new D, another algorithm S suggests new samples for testing the accuracy requirements:

$$\{x\}_{k+1} = S(D_{k+1}). \tag{4}$$

Both the M and S algorithms make frequent use of g.

A typical RSM kit contains computer codes for an RSM evaluator g, data fuser algorithm M and sampling algorithm S, plus a proviso (see
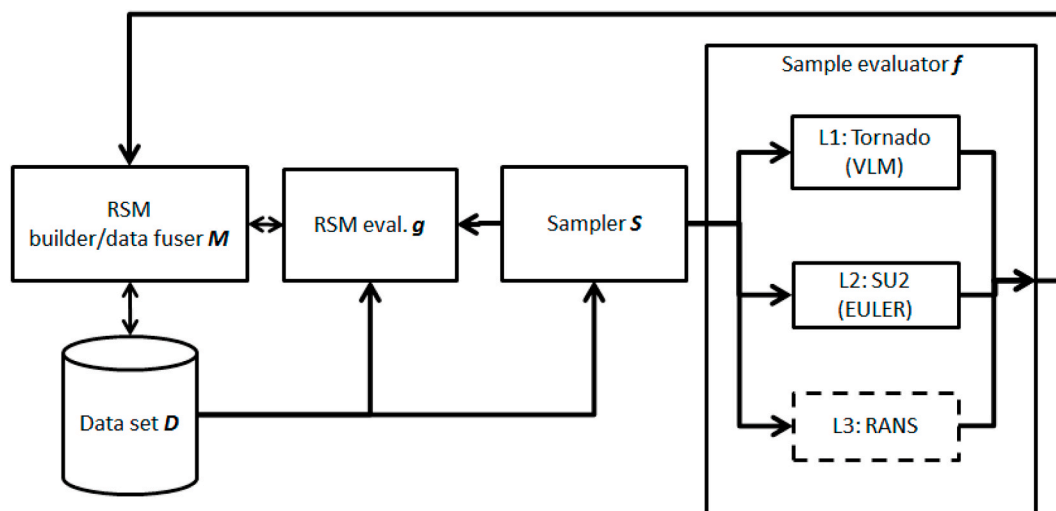

Fig. 4. Workflow for data fusion in the MDA chain, interchangeable modules exchanging data using Brics.
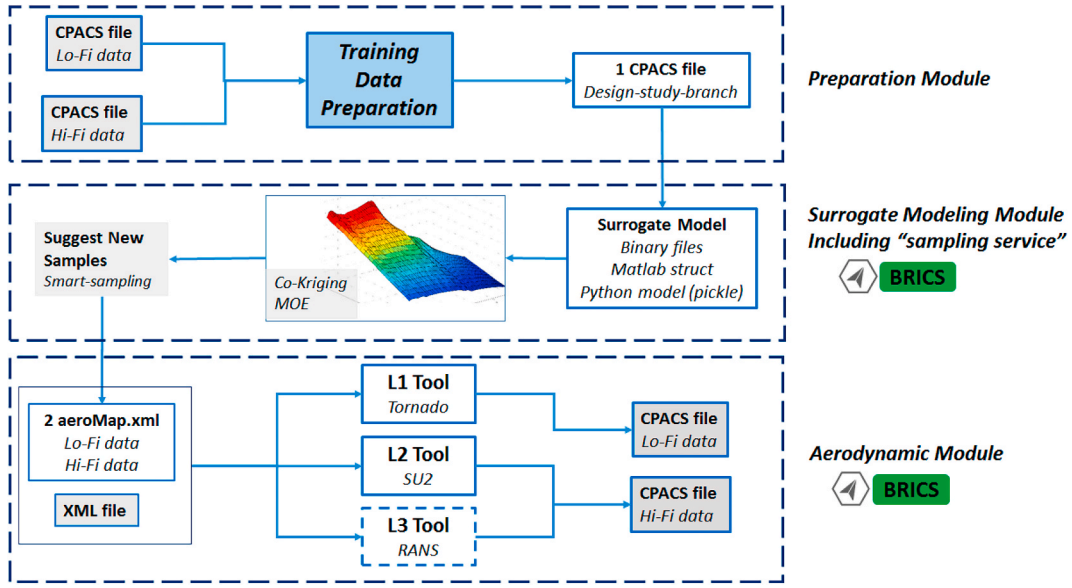
Fig. 5. Data fusion service in the MDA chain with 3 stages.

Section 5.1) to evaluate *f*. A developed instance, the data fusion process, which is ready to run, and improved on further, will contain *g*, *M*, *S* and the data set *D*. Fig. 4 shows an example of a module-independent data fusion workflow exchanging data using Brics. It should be noted that exchanging *D* for each call is impractical in general, due to the size of *D*. Therefore, in the standard use cases, *D* is built on one partner's network/machine, and then made available. The *f* − samples however can easily be exchanged using Brics since the large discretization and solution files (CFD mesh, finite element mesh for CSM) are usually only stored locally. However, for the purpose of tracking possible bugs, etc., the parameters used in mesh generation and simulation should be accessible also "after the fact" to the network.

Use of an existing developed instance requires (access to) computer code *g* and data set *D*. The *g* and *M* functions are in principle well known and open. *S* may involve some proprietary tools which raises IP and access issues, and as *D* embodies information about a design, it also has distribution and access restrictions. Within AGILE a specific Surrogate Model Repository (SMR) was developed that deals with theses issues, see Ref. [22] and Section 2.2.

4.2. Surrogate models description

There are various ways to build surrogate models, the popular ones are for example, polynomial interpolations, the Radial Basis Function (RBF) and Kriging interpolation. In this section it describes two of the mostly used surrogate models in AGILE. Note that, in AGILE framework, the various surrogate models are free to choose to build the CA as long as they follow the steps described in Section 2.1.4.

4.2.1. Kriging and co-kriging

Kriging is a method for scattered data interpolation which sees the data to be predicted as a stationary stochastic field with correction of the predictions depending only on their relative location. The *co-kriging* process provides a technology for the multi-fidelity results where a great quantity of low-fidelity data is coupled with a small amount of high-fidelity data to enhance the accuracy of a surrogate model. The approximation is obtained by updating the correlation ψ and the correlation matrix $\mathbf{\Psi}$ to the co-variance $\mathbf{c}$ and co-variance matrix $\mathbf{C}$ between the low (**c**heap) and high-fidelity (**e**xpensive) observations. The co-kriging estimation $\widehat{y}$ at x^* is:

$$\widehat{y}_{co}\left(x^*\right) = \widetilde{\mathscr{F}}\left(\beta, x^*\right) + \mathbf{c}^T\mathbf{C}^{-1}\left(\mathbf{y} - \widetilde{\mathscr{F}}\right) \tag{5}$$

where $\widetilde{\mathscr{F}}$ is the mean value obtained by the regression model predicted in the same pattern as kriging, by considering both lo-fi and hi-fi samples. If the prediction is made at the *i*th high-fidelity points, and the lo-fi data have m_c design sites, then **c** is the $m_c + i$th column of **C**. Eq. (5) is an interpolator of the hi-fi data just like ordinary kriging, but the lo-fi data will regress in a well defined sense unless it coincides with the expensive observations $\mathbf{y}_e$. Details about kriging, co-kriging and Gaussian Process can be found in Refs. [27,36].

4.2.2. Mixture of experts

The Mixture of Experts (MOE) is a technique which combines local surrogate models in order to approximate heterogeneous functions (flat and steep regions, first and zero order discontinuities) dividing the problem space into homogeneous regions. In the context of this paper, the mixture of experts techniques is used to mix multi-fidelity models as co-kriging. The main idea of mixture of experts is to provide a clustering of the training basis into regions where the function to be approximated is expected to be continuous or at least more simple. In this specific case, the local experts $\widehat{y_k}$ dedicated to each cluster *k*, which are the co-kriging models defined in Eq. (5). As explained in Refs. [37,38], the Gaussian mixture model provides a way to create a global model and predict its value at a new point x_{new} of the design space as a recombination of the local models:

$$\widehat{y}(x_{\text{new}}) = \sum_{k=1}^{K} \mathbb{P}(\kappa = k|\mathbf{x} = x_{\text{new}})\widehat{y_k}(x_{\text{new}}) \tag{6}$$

In this Eq. (6), *K* is the number of clusters, $\mathbb{P}(\kappa = k|\mathbf{x} = x^*)$ denoted by gating network, is the probability to lie in cluster *k* knowing that $\mathbf{x} = x_{\text{new}}$ and $\widehat{y_k}$ is the local expert built on cluster *k*. The number of clusters *K* is chosen automatically to minimize the generalization error on a validation data set [39]. MOE has been made available to AGILE partners for different applications as a remote service [40,41]. MOE has also been wrapped using the Brics connection protocols to enable the transfer of models (constructed on data provided by the user) and not only the results of the evaluation of the models.
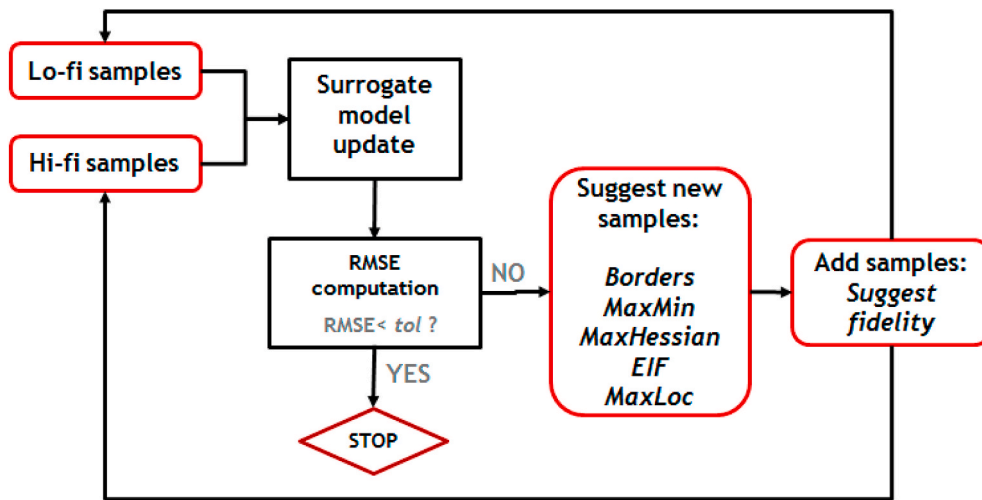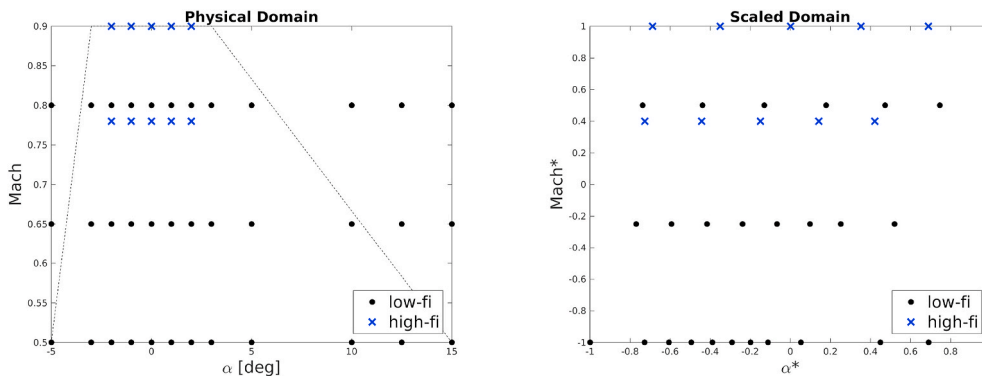
Fig. 6. The "smart sampling" technology workflow.

(a) The physical domain.

(b) The scaled domain.

Fig. 7. The initial samples of the domain for the Test Case and its transformation, at altitude 10 km.

4.3. Data fusion within the collaborative MDA chain

As described at the beginning of the paper, in overall MDO, some tools may need large amounts of data, for instance, the flight simulation tool used for Stability and Control (S&C) analysis needs a complete aerodynamic-database as its input. Some entries in the database can be computed by high-fidelity analysis modules, but not all of them, therefore we need data fusion between various fidelity levels. The fused database, thereafter, will be delivered to the e.g. flight simulation tool.

The data fusion used for AGILE project is implemented as a *process*, and it calls for other tools to run the process. The whole fusion procedure is iterative.

This section shows how the surrogate-based data fusion is implemented in the MDA chain, and how deployment and application of the surrogate model is facilitated by the Surrogate Model Repository.

Fig. 5 shows how the datafusion workflow is executed, the three modules (described later in the paragraph) can be execute at different locations by specialists. This workflow is implemented with RCE which allows to use the Brics plug-in to share information between the main computer and specialist's computer, in the way as Fig. 1 describes.

The data fusion package includes the development/delivery of: (1) the surrogate model builder/fuser M and model evaluator g; (2) the sampling module S; (3) the samples computation module f (aerodynamic module); (4) the graphic feedback module.

This package can be used for fusing computed tables of forces and moments for the purpose of aero-dataset construction. Namely, it can be

used for fusing the **aerodynamic coefficients data** from different tools (sources) and fidelities, and storing the surrogate models.

Fig. 5 spells out the data fusion service in the MDA chain. It has three core modules shown below, the graphic feedback module can be used separately and is not included in the MDA chain.

- **Preparation Module:** It prepares the training data storing in the CPACS file *design study branch*, or in two separate column-based csv files
- **Surrogate Modeling Module:** The surrogate model builder is provided by different tools/partners. This module also includes the "sampling service". The surrogate modeling techniques are
  (1) co-kriging [27] is provided by Airinnova AB in *Matlab* based on the DACE toolbox [42], or its alternative Python's built-in persistence *Model* (scikit-learn), containing the co-kriging model parameters corresponding to the co-kriging surrogate trained with incoming training data. The resulting *Model* file (both Matlab-based and Python built-in) can be read by Python and used by the Python *scikit-learn* toolbox. So that it can be stored, transferred, shared and re-produced by partners via SMR without Matlab license.
  (2) Mixture of Experts (MOE) method for surrogate modeling is provided by ONERA [39] as an executable file. It will output one (binary) file containing the MOE model parameters corresponding to the MOE surrogate trained with incoming training data. The MOE strategy [37,38] recalled in Section 4.2.2 has
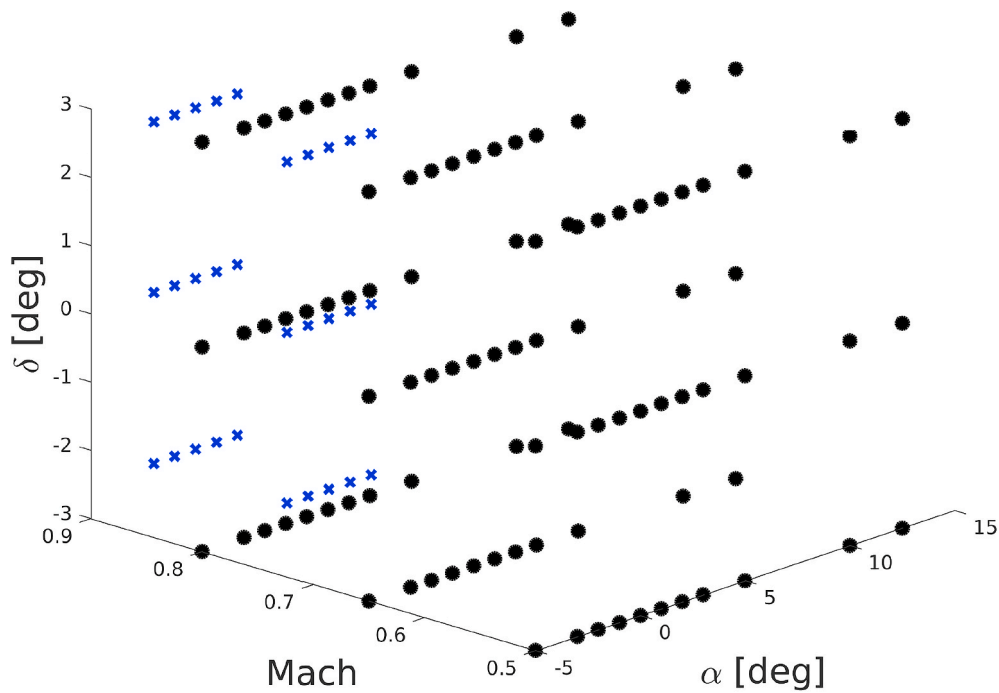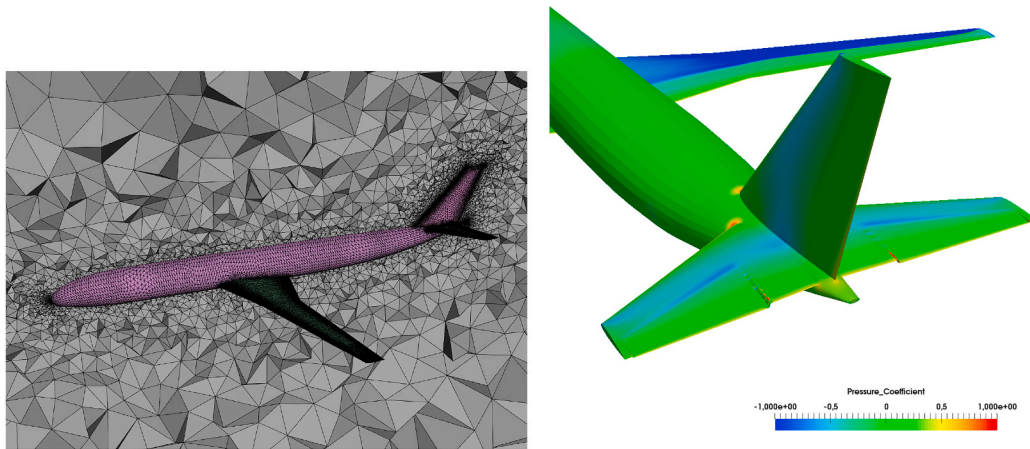
Fig. 8. Initial DOE defined in the CPACS files. Black dot: low-fidelity data; blue cross: high-fidelity data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(a) The volume mesh of the test case.

(b) The C_p contour from Euler solutions.

Fig. 9. (a) The unstructured volume mesh with 5.9 million cells of the test case configuration, created by sumo [48] and TetGen [49]; (b) the C_p contour of the horizontal tail of the test case aircraft from Euler solutions computed by SU2, Mach $= 0.78$, $\alpha = 0°$ with elevator deflection $\delta = 4°$. The elevator deflection is modelled by deforming the mesh defining by FFD.

been made available to AGILE partners for different applications [41,43,44].

- **Aerodynamic Module:** In this module the new samples are computed by the aerodynamic tools. In this paper, the L1 aerodynamic data are provided by Tornado, which is a Vortex Lattice Method code [45]; L2 aerodynamic data are provided by SU2 [46] which is a computational fluid dynamics simulation software. Both of the tools are operated by different specialists at different sites. Other aerodynamic tools can be integrated into this MDA chain by modifying their API to the data fusion package, through the 3 steps described in Section 2.1.4.

5. Design of experiments

This section discusses the issues about the design of experiments, including the sampling algorithm and the design domain validity used in the collaborative data fusion described in Section 4.3.

5.1. Smart sampling algorithm

The "smart sampling" is employed to determine new sampling locations *and* to choose the tools (variable fidelity). The smart sampling scheme uses a mixture of sampling methods with the stopping criteria by examining the MSE or the RMSE (Root Mean Square Error). The sampling identifies a number of sampling locations. The response at the sampling points can be evaluated in parallel, with a constraint to ensure
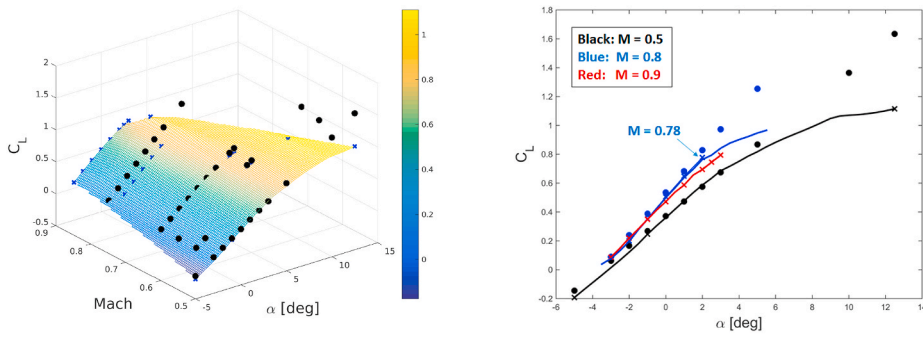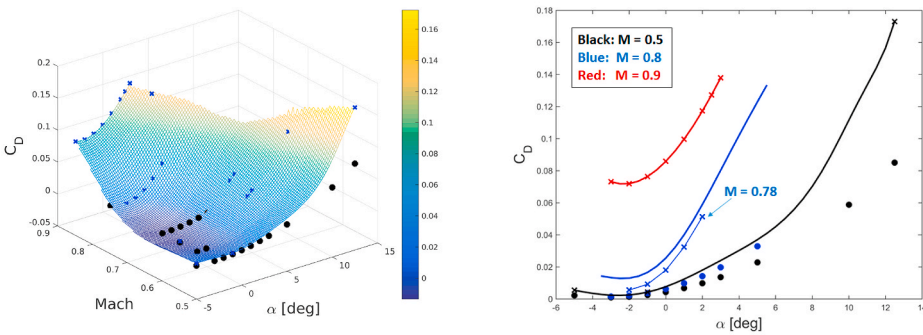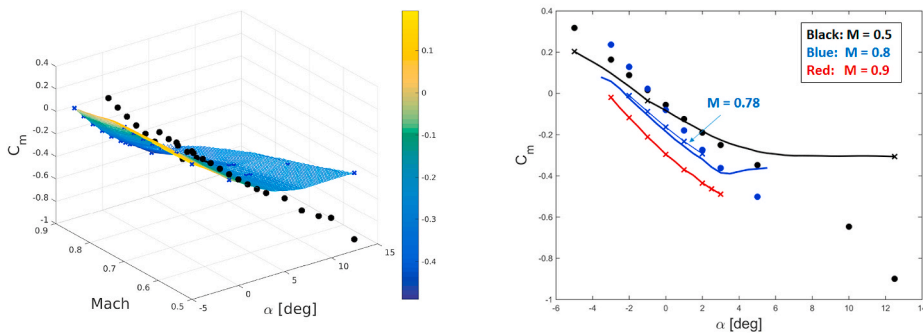
(a) Lift coefficient C_L surface and the sampled data.

(b) Fused lift coefficient C_L for Mach number 0.5, 0.8 and 0.9.

(c) Drag coefficient C_D surface and the sampled data.

(d) Fused drag coefficient C_D for Mach number 0.5, 0.8 and 0.9.

(e) Pitch moment coefficient C_m surface and the sampled data.

(f) Fused pitch moment coefficient C_m for Mach number 0.5, 0.8 and 0.9.

Fig. 10. The co-kriging surrogate model results of AGILE reference aircraft for C_L, C_D and C, with elevator deflection $\delta = 0$ deg. Notations: dot: lo-fi samples; cross: hi-fi samples; line: the response surfaces. Figures (a), (c) and (e): the response surfaces and sampled data over the flight envelope. Figure (b), (d) and (f): the cuts for Mach number 0.5 (black), 0.8 (blue) and 0.9 (red) from the response surfaces, and their corresponding sampled data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

that the sample locations found having a significant distance between each other to avoid overemphasis of the interpolation on a small portion of the predicted function. Due to the inherent nature of the sampling algorithm S (shown in Fig. 6), it is strongly coupled with the RSM evaluator g, namely, kriging or co-kriging surrogate models.

Fig. 6 shows the workflow of the smart sampling technology. For each iteration, the new samples (low-fidelity, or lo-fi & high-fidelity, or hi-fi) are merged so that the surrogate model is updated and the RMSE is computed and compared with the defined tolerance, or *tol*, chosen by the user. If the stopping criteria are met, the iteration ends. Otherwise new samples are suggested according to the listed methods in Fig. 6 and described in the following. The new suggested samples will be computed by the corresponding tools and will be added to the samples to update

the surrogate model. The computational tool fidelity is also suggested.

The methods used to suggest new samples are described below, and the rules are understood as a hierarchical sequence (from 1 to 5):

1. *Borders*: only applies when loading the initial samples. Check the "borders" of the DOE domain, if the hi-fi samples are missing at any of the border locations, add them (computed by hi-fi tools). This must be done at the first step of the sampling, because Kriging relies on the assumption of constant mean and variance within the domain of interest, and is not performing well at extrapolation. The low-fidelity data tends predicts the wrong trends, that the high-fidelity samples must be added to correct the wrong trend especially at the borders of the design domain to avoid extrapolation.
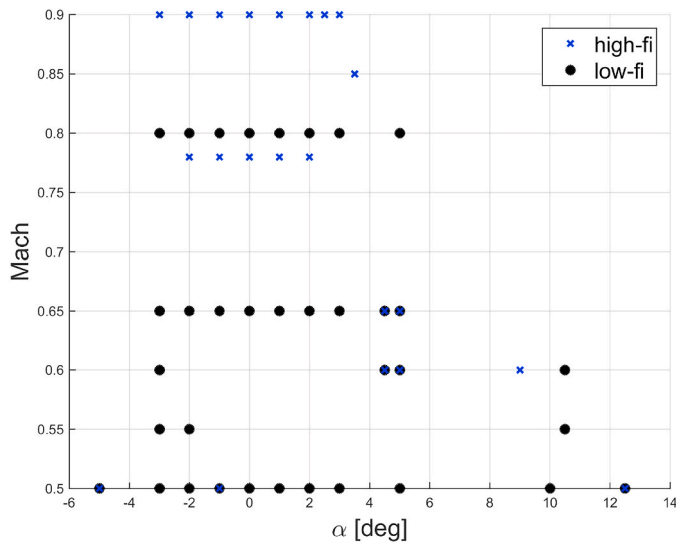
Fig. 11. Final DOEs for building the surrogate models, viewed in a 2D space.

2. *MaxMin*: finds the local maximum or minimum of the surrogate model and the new samples will be added there. The position of the local maxima or minima is computed by considering the full surrogate model, comparing any function values with all the points inside a sphere of radius previously computed and centered in it. The sphere radius is initially computed as minimum of the Euclidean norms of any two points with all non-equal coordinates. If the function value is bigger or smaller than all the others in the sphere, the point is marked as local maximum or minimum respectively [47]. The "MaxMin" method can at a time recommend n-set of samples (defined by the user). The new samples shall be computed by L1 tool (lo-fi) by default if tool fidelity is not specified.

3. *MaxHessian*: finds/computes the maximum curvature of the surrogate model and the new samples will be added there. The new samples shall be computed by L1 tool (lo-fi) by default. Computation details can be found in Appendix in Ref. [40]. This method only becomes active if *MaxMin* fails.

4. *EIF*: finds the Expected Improvement Function (EIF) location [27] and the new samples will be added there *only* when the *MaxMin* and *MaxHessian* fail to add the effective samples, i.e., the new samples suggested by previous methods are already or very close to the existing samples via a distance criterion threshold chosen by the user. The new samples shall be computed by L1 tool (lo-fi) by default.

5. *MaxLoc*: finds the maximum RMSE locations and new samples will be added there. This method is used *only* when the maximum RMSE is not significantly improved compared with the previous iterations:

$$\text{RMSE}(k+1) > \nu \cdot \text{RMSE}(k) \tag{7}$$

where $0 < \nu < 1$ is the improvement factor chosen by the user, thus the sampling method shall be "switched" to *MaxLoc*.

As stated above, except *Borders*, the new samples suggested by all other methods will be computed by L1 tools (lo-fi) by default. There is a hierarchy of models and the highest fidelity one is considered as the *truth*, i.e., we have no error estimate between the highest fidelity models and the "reality". We wish to use lower fidelity (lo-fi) models where they give results close to the hi-fi models, so first use the lo-fi until error estimate satisfies the lo-fi criterion. Then the hi-fi points are filled in until the error estimate between response surface and the hi-fi is small enough (i.e., fulfills the hi-fi criterion). Note that this may "waste" the lo-fi calculations in regions where the lo-fi is bad so the hi-fi is necessary.

The hierarchical sequence of choosing samples is based on an investigation and combination of methods stated in Refs. [27,34,47], the goal being to improve the surrogates using a minimum number of samples, i.e., minimal computing efforts. First check the *borders* to avoid extrapolation, then enrich the surrogate models by infilling the points suggested by *MaxMin* and *MaxHessian*, which allows the predicted models quickly converge upon a maximum/minimum value, or a peak/dip of the true function [27,47]. This is particularly useful when the model to be predicted is highly non-linear, such as C_L, C_m aero coefficient curves especially for transonic flight conditions. If the above two methods fail, which means the search only finds a local optimum, or does not even find a local minimum/maximum (for example, the large portions of the surrogate function landscape are flat), the surrogate model does not approximate the whole function well, a sampling strategy that can search away from the current minimum and explore other regions is required. The EIF is error based exploration which is suitable for the situation stated above. In many situations, the EIF function would give the same recommended sampling locations as the *Max/Min* and *MaxHessian*. However, the EIF may be not so efficient when the prediction "thinks" that the function is very smooth [27] that EIF function diminishes to a very small value, with a very small estimated error RMSE. The small error leads to an overemphasis on exploitation of the prediction and the sampling approach falls into the trap because the unknown Kriging model based parameters is assumed to be estimated correctly. *MaxLoc* is used to improve the surrogates at the locations with the largest MSE values if the maximum RMSE is not significantly improved by *Find EIF*. The source fidelity for the samples also needs to be upgraded accordingly. The infilling samples suggested by the maximum MSE locations tends to make a uniform distribution over the design domain, which may "waste" samples. The ideal scenario is to find the best fitted model as quickly as possible, that the sampling points get "clustered" at the non-linear parts, and get "scattered" at the linear parts.

It shall be possible to choose freely between source tools with different fidelity levels. The criteria for going to the next fidelity level and to switch methods are:

- If the new suggested samples are already in the hi-fi sampled domain, switch to the next method level. For example, if the *MaxMin* suggests a new sample that is already in the hi-fi sampled domain, we then use the *MaxHessian*.
- If the new suggested samples are already in the lo-fi sampled domain but not in the hi-fi sampled domain, upgrade the fidelity.
- If a low fidelity tool fulfills the lo-fi accuracy criteria, for example, if the maximum RMSE is small enough, go to the next fidelity level, namely, L_k "upgrades" to L_{k+1}, where k is the current fidelity level.
- Another indication is that the maximum RMSE is *not* significantly improved compared with the previous iteration as described with Eq. (7). This means either the methods shall be "switched" or the fidelity shall be improved.

This "smart" sampling algorithm needs now to be associated with a re-scaling of the design domain in order to ensure that points are added throughout the domain.

5.2. Domain validity issues

The problem of fusing the aerodynamic characteristics, it relates the flight envelope identification, or the identification of the domain validity of the surrogate models. Usually, the DOE techniques are designed to work on rectangular (cubic for 3D) domains. Moreover, the inherent characteristics of the CPACS file definition *only* supports the uniform distribution of the samples. Therefore, the initial samples, are computed with regular sampling. However, the physical flight envelope is just part of the "CPACS envelope" [40]. For instance, the test case which will be used in this paper, the AGILE reference aircraft, as a conventional transport airliner, its physical flight envelope, will not cover, the high angles of attack at high Mach numbers. The parameter space should be chosen as the physical flight envelope, otherwise the sampling rules will

Fig. 12. Meta information of an example surrogate model available through the SMR. In the text, its is mentioned that this is a SM resulting from the test/experiments.

fail and the new suggested samples will always locate around the edges of the (rectangular) domains since they are highly non-linear (but not realistic). This section will spell out how to cope with the domain validity issues.

In this paper, the physical domain is defined by the flight mechanics specialist co-author, as the dashed line bounds shown in Fig. 7a for a two-dimensional parameter space. The sample points on the physical domain will be re-scaled to fit the new domain within the interval [-1,1] in both dimensions, see Fig. 7b. All samples which are left out of the physical domain will be excluded. During the surrogate modeling process, the design domain will firstly be re-scaled and exclude the parameter space which are left out of the physical domain, before building a surrogate model and iterate the sampling procedures. This will provide a better "coverage" of the area of the domains that are typically difficult to model (extreme flight conditions).

6. Application to the collaborative aerodynamic surrogate modeling

6.1. Aerodynamic model and handling quality tool descriptions

The test case aircraft configuration is the AGILE reference aircraft, a regional jet-liner, which was analyzed and simulated using the AGILE MDA system, without experimental data being available. This aircraft does not correspond to an existing one, but it is in the range of an Airbus 320 or Boeing 737. The reference aircraft is defined in CPACS [14] format.

Some previous numerical simulations have been performed for this aircraft to test the data fusion tool [40], however the primary control surfaces were not modelled. In the test case of this paper, the aerodynamic coefficients and derivatives for the longitudinal analysis are computed by L1 and L2 tools, including the elevator deflections. Those computational results are fused as an aerodynamic-database for the
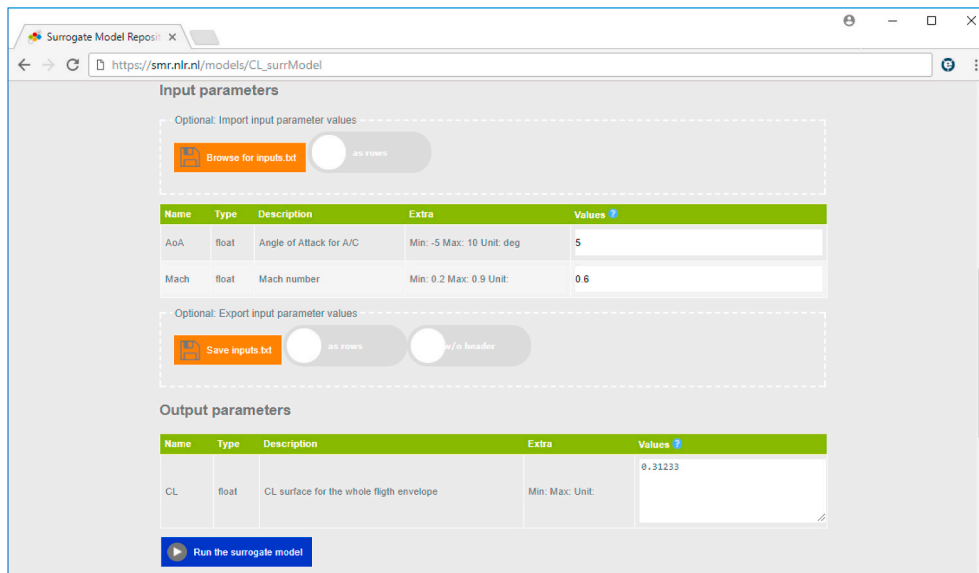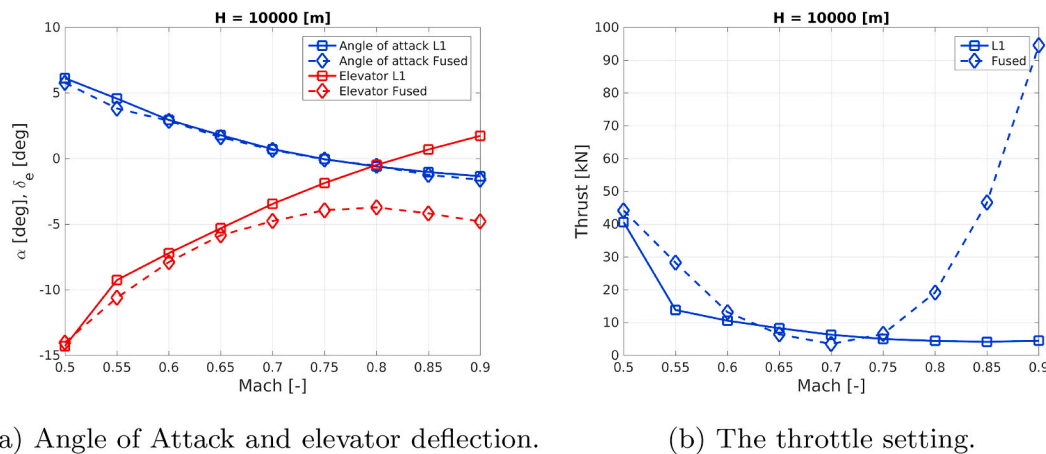
Fig. 13. SMR user interface for the "as a service" calculations with the surrogate model.

(a) Angle of Attack and elevator deflection.

(b) The throttle setting.

Fig. 14. Angle of attack, elevator deflection and the throttle setting between the L1 data and the fused data in the horizontal trimmed flight at 10 km altitude, as a function of Mach number.

longitudinal flight simulation using the flight simulator PHALANX, which will be described in the last part of this section.

The L1 aerodynamic tool used for the data fusion workflow is the Vortex Lattice Method (VLM) code Tornado [45], which solves the linearized equations and considers the flow compressibility by taking the Prandtl-Glaurent rule [?]. The L2 aerodynamic tool used is the open-source code SU2 developed by the Stanford University, which is a fluid dynamics solver for solving the incompressible/compressible and inviscid/viscous flows. In this paper the SU2 is used as L2 level, namely, as an Euler equation solver for solving the inviscid compressible flows. Both of L1 and L2 tools have been fully adapted to the CPACS format, integrated onto RCE and are callable using Brics. Details can be found in Ref. [19].

As stated in Section 5.2, the initial DOEs from the CPACS files are uniformly distributed, then the valid domain is selected according to the physical flight envelope. Fig. 8 spells out the initial DOEs defined in the CPACS files by regular sampling, which consist of two sets of data from both the low-fidelity (L1) and the high-fidelity (L2) tools. The low-fidelity data are symmetrically distributed in the range of angle of attack [-5, 5], and sparser at higher angle of attack. This is because the L1 solver is based on linearized equations and its prediction is questionable at higher angles of attack where nonlinear aerodynamic is

dominant, so that the L1 sample locations at high angles of attack are trivial. The parameter space is three-dimensional: the angle of attack α, Mach number and the elevator deflection δ. The flight altitude is fixed at 10 km and the sideslip angle is $0°$. Again, the initial DOEs which are defined/stored in the CPACS files exceed the bounds of the design domain, which means some of the computations (data) are "wasted" and will be ignored in the future fusion process.

The CPACS file of the AGILE reference aircraft is converted into another type of XML definition geometry to be opened with the mesh generator software sumo [48]. A surface mesh is created automatically by sumo and it calls TetGen [49] to create an unstructured Euler mesh. The mesh used for this study has been chosen following a mesh sensitivity analysis performed in Ref. [50]. It is an unstructured mesh with 5.9 million tetrahedrons, see Fig. 9a.

In order to compute the elevator deflections, the SU2 built-in mesh deformation function SU2_DEF is used to deform the mesh around the elevator locations on the horizontal tail. A Free-form deformation (FFD) [51] box is defined at the elevator locations. With the hinge line location specified, the mesh in the FFD box can be deformed around the hinge line within a certain angle. To avoid high aspect ratio cells (or even negative volume) usually small deflection angle is preferred. According to authors experience, with a deflection angle less than $8°$ the deformed
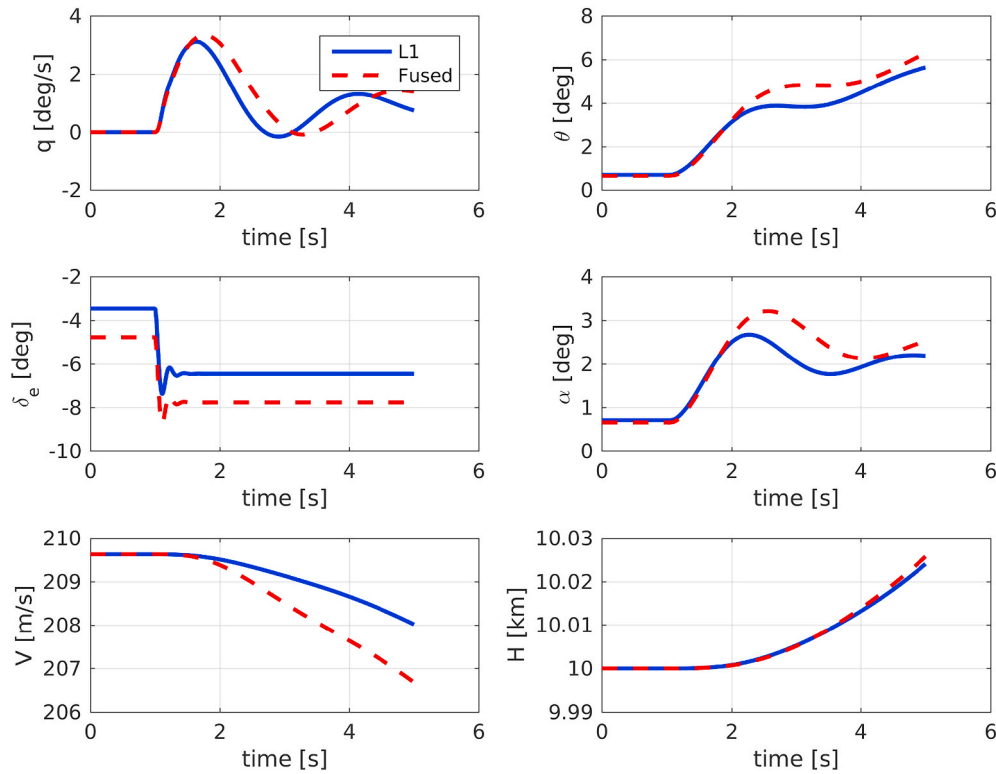
Fig. 15. A step input on the elevator and its dynamic motion for a single Mach number M = 0.7.

Table 2
The characteristics of short period and phugoid modes comparison between L1 and the fused data for Mach = 0.7 at altitude 10 km.

-	L1	Fused
ω_{sp} [rad/s]	2.7063	2.3476
z_{sp} [−]	0.2319	0.2656
ω_{ph} [rad/s]	0.0766	0.0632
z_{ph} [−]	0.0084	0.0591

mesh can give well-converged solutions. In this paper we limit the deflection angle within 4° for the high speed flight at the high altitude.

Tornado usually takes less than 1 min to make one single VLM analysis on a modern laptop, and the Euler solver takes around 4 min on a 32-cores work station for the reference aircraft. The computational cost ratio is at least 128 (provided that the SU2 parallel computing speed-up is linear). In this test case, L1 solver is chosen as low-fidelity data source and L2 solver is chosen as high-fidelity data source, so that the low-fidelity samples are very cheap to generate. Even the "expensive" samples are fast to obtain because of the relatively coarse mesh. The reason why the cheap samples are not generated at all locations before co-Kriging begins is, there are not always tools to provide cheap samples as fast as the L1 tool (e.g. Tornado). For example, in the study of surrogate model generation for aero-data and aero-loads of X-31 aircraft [24], the cheap samples are provided by the Euler solver (L2) and the expensive samples are provided by the RANS solver (L3) on the very fine grids, with computational time substantially increased. Similar case was also found in Ref. [28]. In these situations, it is not possible to generate even the low-fidelity samples at all locations in the design domain, because the computation cost for the low-fidelity samples is still not cheap. The generation of surrogate models also includes to populate the aerodynamic database over the whole flight envelope by the relatively "dense" low-fidelity data samples, as stated in Section 1. The method and technology described in this paper show a generic way of

data fusion including sampling strategies, with the goal of data fusion by surrogate modeling with minimal computing efforts keeping in mind.

The Performance, Handling Qualities and Loads Analysis Toolbox (PHALANX) is a selective fidelity flight mechanics modeling and analysis tool. It is specifically designed to be used in a multidisciplinary design optimization framework [52] and to support the analysis of future novel aircraft designs [53,54]. PHALANX has the capability to automatically construct and analyze aircraft models within an MDO environment [55,56]. This makes it possible to analyze the flight mechanics of many different variants of novel aircraft and configurations without a user in the loop. The aircraft models are nonlinear simulation models which serve as virtual flight test vehicles. An extensive analysis suite is available to evaluate aircraft performance characteristics, to perform handling assessments and to simulate loads resulting from both intentional manoeuvres and atmospheric conditions.

6.2. Results

The section shows the results obtained from the collaborative aerodynamic data fusion framework, including the fused aerodynamic coefficients, and flying qualities prediction from the fused data, and the application of SMR.

6.2.1. Fused aerodynamic characteristics
Fig. 10 shows the fused C_L, C_D and C_m aero-coefficient results of the AGILE reference aircraft model from the both fidelities with the elevator deflection $\delta = 0°$ over the flight envelope. The dot sign (·) represents the lo-fi samples and the cross sign (×) represents the hi-fi samples. Fig. 10a,c and e show the response surfaces from the surrogate models as well as the sampled data over the flight envelope in the three-dimensional space. Fig. 10b,d and f represent the two-dimensional cuts for Mach number 0.5 (black), 0.8 (blue) and 0.9 (red) from the response surfaces, and their corresponding sampled data. Note that for $M = 0.8$ there are no hi-fi samples computed, instead there are hi-fi samples computed at $M = 0.78$, which are shown and marked in the

figures.

Fig. 10a and b shows the surrogate models (response surfaces) for C_L. The co-kriging predicts the non-linear behaviors at higher angles of attack, as the hi-fi samples indicate.

Fig. 10c and d shows the prediction for C_D. The surrogate model predicts higher drag than the lo-fi samples show, since they cannot predict wave drag. It is a promising sign that the surrogate model picks up the compressible phenomena from the hi-fi samples.

The surrogate model for C_m is shown in Fig. 10e and f. Note again that the surrogate model predicts the non-linear trends at high AoA, as expected. The coarse hi-fi samples correct the response surfaces significantly.

The co-kriging predicts 238×2 cases in the physical domain which are selected from a rectangular domain with 324×2 cases with α from $-5°$ to $12.5°$, and Mach from 0.5 to 0.9 for elevator deflection $\delta = 0°$. The full surrogate prediction has 238×3 cases with elevator deflection from $-3°$ to $+3°$. The computation time of the surrogate model is ≈ 0.05 s on a desktop computer with four CPUs. The surrogate model is reliable with $max(\text{RMSE}) = 0.048 < 5\%$. The final DOEs for building the surrogate models have been shown in Fig. 11, with 22×3 hi-fi samples and 35×3 lo-fi samples for elevator deflection from $-3°$ to $+3°$. It is viewed in a 2D space since for each elevator deflection, the samples are at the same AoA and Mach locations.

The static coefficients are computed by L1 and L2 tools, and the database is generated by co-kriging, as stated above. However, the dynamic stability derivatives are only computed based on L1 tool Tornado in order to reduce the computing efforts while demonstrating the benefits of the fused data.

6.2.2. Example of surrogate model deployment with SMR

Using the AGILE Data Fusion service a surrogate model has been derived that predicts the aerodynamic lift coefficient C_L as a function of angle of attack (AoA) and Mach number. The meta-information of this surrogate model has been stored at the SMR, see Fig. 12. The meta-information includes a description of the purpose and the background of this surrogate model. Furthermore, the input and output variables are specified. For the input variables the allowable range is specified. For the output variables the verification result is given, as such providing information on the accuracy of the surrogate model.

By setting the toggle in the bottom (see Fig. 12) the surrogate model can be invoked. Input values for AoA and Mach are provided. Under the hood, the SMR translates these values into a CPACS file, uploads the file to a shared location and sends a notification (e.g. by email) to the specialist that provides the surrogate model calculation service. The specialist downloads and reads the CPACS file with input values, performs the calculation and writes and uploads the output in the same CPACS format to the shared location. From here the SMR retrieves the output value and projects it in the output value field, see Fig. 13.

6.2.3. Prediction of flying qualities from fused data

The flight performance and flight dynamics are analyzed by PHALANX and compared between the fused data (by co-kriging from L1 and L2 tools) and the L1 data alone. PHALANX is written in Matlab and makes extensive use of the Simulink©platform, its toolboxes and the Simscape©environment for modeling and simulating physical systems. For the current test case, the equations of motion are modelled using multi-body dynamics. Thereby, the mass and inertia of the empty aircraft can be modelled separately from the payload and fuel. The aerodynamics are modelled using a database in the form of look-up tables. These tables are a function of angle of attack, pitch rate and elevator deflection. The propulsion system (thrust and fuel consumption) is modelled based on engine performance maps which are a function of Mach number, altitude and throttle setting. The resulting nonlinear simulation model is used to assess the trim condition (e.g. the prediction of flight envelope limits and power required as function of Mach number) and a handling qualities assessment. The analysis results

for various flight conditions and aircraft configurations are written in the CPACS file.

Fig. 14a shows the angle of attack and elevator deflection in trimmed flight for the whole range of Mach numbers at 10 km altitude. One can clearly see that the elevator deflection results start to deviate at higher Mach numbers. This is caused by the differences in the static moment coefficient between the two databases. The required lift coefficient is relatively small and for the this range of lift coefficients, the fused database and L1 database give similar predictions for the angle of attack. As a consequence, the control authority of the elevator at high speed flight is better predicted by the fused database.

Fig. 14b shows the throttle setting in trimmed flight as a function of Mach number. The fused data shows a shape which is to be expected. The L1 results are completely wrong at high Mach numbers since it is rather obvious that Tornado cannot predict C_D properly at high Mach numbers due to the presence of the wave drag.

For a single Mach number M = 0.70, a step input of $3°$ on the elevator is performed. Actuator dynamics are included in the simulation, see Fig. 15. As a result one can see the dynamic motion. Again there are differences between the L1 and the fused model. The first observation that can be made is a difference in control power. The results based on the fused data show a larger attitude change for the same elevator change. This is caused by the combination of a lower static moment coefficient and a slightly different control derivative.

Table 2 shows the characteristics of the Short Period mode and the Phugoid mode comparison between L1 and the fused data. The Phugoid mode is largely determined by the lift over drag ratio for conventional aircraft. The main reason that the Phugoid frequency is reduced and the damping is increased is because the fused database with L2 correction taking the compressibility/nonlinearity into account, thus it has a decreased L/D compared to the L1-Tornado results.

The short period depends to a large extent on the dynamic derivative M_q, which is identical for both databases since they are produced by L1 tools for both. The small differences of the results are due to the differences between the static lift coefficient and the static moment coefficient.

7. Conclusion and perspectives

The paper has presented a collaborative and surrogate-based data fusion technology for generating the aerodynamic database for the handling qualities analysis. This data fusion technology is implemented in a collaborative MDA workflow, utilizing the existing tools within AGILE, establishing an iterative and collaborative, distributed process. The surrogate models which are built from the data fusion service can be stored and deployed for reuse with the Surrogate Model Repository (SMR). An example was given for running a surrogate model "as-a-service" through the SMR. A regional jet defined within AGILE project as a reference aircraft is used as the test case for the data fusion tool. A full spreadsheet of aerodynamic data computed either with different levels of fidelity or with only a low-fidelity tool has been derived using the data fusion package. It has been shown that the quality of the flight performance simulation was significantly improved especially for the transonic region in which the low fidelity aerodynamic method is not reliable. The test case shows that by using a surrogate model based the data fusion technique, the fidelity of the analysis data can be significantly improved with minimal computing efforts.

The data fusion process, integrated in the AGILE collaborative MDA chain with different tools or modules interconnected via Brics and RCE, can be used as a well-established and ready-to-use service to be applied to any other aircraft for generation of aerodynamic databases and flying qualities analyses, in order to reduce the computational time and increase the overall prediction accuracy.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

References

[1] P. Ciampa, B. Nagel, The agile paradigm: the next generation of collaborative mdo, in: AIAA Paper 2017-4137, AIAA Aviation 2017, 2017. Denver, USA, https://www.agile-project.eu/cloud/index.php/s/6xxsl5yBO1cCw7d.

[2] M. Zhang, Contributions to Variable Fidelity MDO Framework for Collaborative and Integrated Aircraft Design, Department of Aeronautics, Royal Institute of Technology, KTH, Stockholm, Sweden, 2015. Ph.D. thesis.

[3] A. Abbas-Bayoumi, K. Becker, An industrial view on numerical simulation for aircraft aerodynamic design, J. Math. Industry 1 (10)..

[4] The strategic research and innovation agenda (sria), 2018-09-10, https://www.acare4europe.org/sria.

[5] The simsac project homepage, 2018-09-01, http://simsacdesign.org.

[6] A. Rizzi, et al., CEASIOM Validation and its Use in Design - Status of SimSAC Project, SAAB Flygtekniksseminarium, Komården, Sweden, 2008.

[7] E.H. Baalbergen, J. Kos, W.F. Lammen, Collaborative multi-partner modelling & simulation processes to improve aeronautical product design, in: 4th CEAS Air & Space Conference, 2013. Linköping, Sweden.

[8] P. Coleman, O. Tabaste, The Behavioural Digital Aircraft Vision for Simulation and Collaborative Product Development, 2013.

[9] AGILE EU project portal. http://www.agile-project.eu.

[10] M. Ghoreyshi, K. Badcock, M. Woodgate, Accelerating the numerical generation of aerodynamic models for flight simulation, J. Aircraft 46 (3) (2009) 972.

[11] M. Haghighat, M. Abdel-Mottaleb, W. Alhalabi, Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition, IEEE Trans. Inf. Forensics Secur. 11 (9) (2016) 1984–1996, https://doi.org/10.1109/TIFS.2016.2569061.

[12] M. G. Fernández-Godino, C. Park, N.-H. Kim, R. T. Haftka, Review of Multi-Fidelity Models, arXiv preprint arXiv:1609.07196.

[13] D. Böhnke, B. Nagel, M. Zhang, A. Rizzi, Towards a collaborative and integrated set of open tools for aircraft design, in: 51st AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, Grapevine (Dallas/Ft. Worth Region), Texas, USA, 2013, pp. 7–10.

[14] CPACS, A common language for aircraft design. https://software.dlr.de/p/cpacs/home/. (Accessed 28 August 2017).

[15] B. Nagel, D. Böhnke, V. Gollnick, P. Schmollgruber, A. Rizzi, G.L. Rocca, J. Alonso, Communication in aircraft design: can we establish a common language?, in: 28th International Congress of the Aeronautical Sciences, 2012.

[16] D. Seider, P. Fischer, M. Litz, A. Schreiber, A. Gerndt, Open source software framework for applications in aeronautics and space, in: IEEE Aerospace Conference, 2012. http://elib.dlr.de/77442/.

[17] E. Baalbergen, J. Kos, C. Louriou, C. Campguilhem, J. Barron, Streamlining cross-organisation product design in aeronautics, Proc. IME G J. Aero. Eng. 213 (12) (2017).

[18] P. Ciampa, E. Moerland, D. Seider, E. Baalbergen, R. Lombardi, R. D'Ippolito, A collaborative architecture supporting agile design of complex aeronautics products., Vol. AIAA Paper 2017-4138, in: AIAA Aviation 2017, 2017. Denver, USA, https://www.agile-project.eu/cloud/index.php/s/FxMh77oRF5nzKds.

[19] M. Zhang, A. Jungo, A. A. Gastaldi, T. Melin, Aircraft geometry and meshing with common language schema cpacs for variable-fidelity mdo applications, Aerospace 5 (2).

[20] E. Moerland, P.D. Ciampa, S. Zur, E. Baalbergen, R. D'Ippolito, R. Lombardi, Collaborative Architecture Supporting the Next Generation of MDO within the AGILE Paradigm, Submitted to Progress in Aerospace Sciences, Special Issue TBD (TBD), TBD, 2018.

[21] I. van Gent, B. Aigner, B. Beijer, J. Jepsen, G.L. Rocca, Knowledge Architecture Supporting the Next Generation of Mdo in the AGILE Paradigm, Submitted to Progress in Aerospace Sciences, Special Issue TBD (TBD), TBD, 2018.

[22] E. Baalbergen, E. Moerland, W. Lammen, P.D. Ciampa, Methods supporting the efficient collaborative design of future aircraft, in: 6th CEAS Conference, 2017. Bucharest, Romania.

[23] A. de Wit, W. Lammen, W. Vankan, H. Timmermans, T. van der Laan, P. Ciampa, Aircraft Rudder Optimization - a Multi-Level and Knowledge - Enabled Approach, Submitted to Progress in Aerospace Sciences, Special Issue TBD (TBD), TBD, 2018.

[24] M. Zhang, M. Tomac, C. Wang, A. Rizzi, Variable fidelity methods and surrogate modeling of critical loads on x-31 aircraft, in: 51st AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition vol. 1081, Grapevine, 2013, pp. 7–10. Texas, USA.

[25] ALEF Consortium, ALEF aerodynamic loads estimation at extremes of the flight envelope, 2012 final publishable report, Tech. Rep., D0.1-1, May 6.

[26] A. Da Ronch, M. Ghoreyshi, K. Badcock, On the generation of flight dynamics aerodynamic tables by computational fluid dynamics, Prog. Aero. Sci. 47 (8) (2011) 597–620.

[27] A. Forrester, A. Keane, et al., Engineering Design via Surrogate Modelling: a Practical Guide, John Wiley & Sons, 2008.

[28] M. Ghoreyshi, K. Badcock, A. Da Ronch, D. Vallespin, A. Rizzi, Automated cfd analysis for the investigation of flight handling qualities, Math. Model Nat. Phenom. 6 (3) (2011) 166–188.

[29] R. von Kaenel, A. Rizzi, J. Oppelstrup, T. Goetzendorf-Grabowski, M. Ghoeryshi, L. Cavagna, A. Berard, CEASIOM: Simulating Stability & Control with CFD/CSM in Aircraft Conceptual Design, 26th International Congress of the Aeronautical Sciences, ICAS., 2008.

[30] Sdsa, Simulation and dynamic stability analyser, 2018-09-10, https://www.meil.pw.edu.pl//add/ADD/Teaching/Software/SDSA.

[31] M. Ghoreyshi, K. Badcock, A.D. Ronch, S. Marques, A. Swift, N. Ames, Framework for establishing limits of tabular aerodynamic models for flight dynamics analysis, J. Aircraft 48 (1) (2011) 42–55.

[32] J. Williams, S.R. Vukelich, The USAF Stability and Control Digital DATCOM, Tech. Rep, mcDonnel Douglas Astona UTICS Company, St. Louis, MO, 1979. AFFDL TR-79-3032.

[33] D. Vallespin, K. Badcock, A. Da Ronch, M. White, P. Perfect, M. Ghoreyshi, Computational fluid dynamics framework for aerodynamic model assessment, Prog. Aero. Sci. 52 (2012) 2–18.

[34] A. Da Ronch, M. Panzeri, M.A. Abd Bari, R. d'Ippolito, M. Franciolini, Adaptive design of experiments for efficient and accurate estimation of aerodynamic loads, Aircraft Eng. Aero. Technol. 89 (4) (2017) 558–569.

[35] T. Lefebvre, N. Bartoli, S. Dubreuil, M. Panzeri, R. Lombardi, P. Della Vecchia, F. Nicolosi, P. Ciampa, K. Anisimov, A. Savelyev, Methodological enhancements in MDO process investigated in the AGILE European project, in: 18st AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2017. Denver, Colorado, USA, AIAA 2017–4140.

[36] C.E. Rasmussen, C.K. Williams, Gaussian Processes for Machine Learning, vol. 1, MIT press Cambridge, 2006.

[37] D. Bettebghor, N. Bartoli, S. Grihon, J. Morlier, M. Samuelides, Surrogate modeling approximation using a mixture of experts based on EM joint estimation, Struct. Multidiscip. Optim. 43 (2) (2011) 243–259, https://doi.org/10.1007/s00158-010-0554-2.

[38] R.P. Liem, C.A. Mader, J.R.R.A. Martins, Surrogate models and mixtures of experts in aerodynamic performance prediction for mission analysis, Aero. Sci. Technol. 43 (2015) 126–151, https://doi.org/10.1016/j.ast.2015.02.019.

[39] N. Bartoli, I. Kurek, R. Lafage, T. Lefebvre, R. Priem, M.-A. Bouhlel, J. Morlier, V. Stilz, R. Regis, Improvement of efficient global optimization with mixture of experts: methodology developments and preliminary results in aircraft wing design, in: 17th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2016, https://doi.org/10.2514/6.2016-4001. Washington D.C., USA.

[40] M. Zhang, A. Jungo, N. Bartoli, Disciplinary data fusion for multi-fidelity aerodynamic application, in: 6th CEAS Aerospace Europe Conference, Paper no.953, 2017. Bucharest, Romania.

[41] T. Lefebvre, N. Bartoli, S. Dubreuil, M. Panzeri, R. Lombardi, W. Lammen, B. de Wit, M. Zhang, I. van Gent, P.D. Ciampa, Overview of mdo enhancement in AGILE project: a clustered and surrogate-based mda use case, in: 6th CEAS Aerospace Europe Conference, Paper no.956, 2017. Bucharest, Romania.

[42] N. Lophaven, H. Nielsen, J. Sondergaard, Dace: A Matlab Kriging Toolbox, Technical University of Denmark, 2002 version 2.0, august 1.

[43] T. Lefebvre, N. Bartoli, S. Dubreuil, M. Panzeri, R. Lombardi, R. D'Ippolito, P. D. Vecchia, F. Nicolosi, P.D. Ciampa, K. Anisomov, A. Savelyev, Methodological enhancements in MDO process investigated in the AGILE european project, in: AVIATION/AIAA 2017, 18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2017, https://doi.org/10.2514/6.2017-4140. Denver, United States.

[44] N. Bartoli, T. Lefebvre, S. Dubreuil, M. Panzeri, R. D'Ippolito, K. Anisomov, A. Savelyev, Robust nacelle optimization design investigated in the AGILE european project, in: Submitted to AVIATION/AIAA 2018, 19th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2018. Atlanta, United States.

[45] T. Melin, Using Internet Interactions in Developing Vortex Lattice Software for Conceptual Design, Ph.D. thesis, Department of Aeronautics, Royal Institute of Technology, KTH, Stockholm, Sweden, 2003. software available from: http://redhammer.se/tornado/. July 2009.

[46] T. D. Economon, F. Palacios, S. R. Copeland, T. W. Lukaczyk, J. J. Alonso, SU2: an open-source suite for multiphysics simulation and design, AIAA Journal, March, Vol. 54, No. 3 : pp. 828-846.

[47] L. Pronzato, Minimax and maximin space-filling designs: some properties and methods for construction, J. Soc. Fr. Stat. 158 (1) (2017) 7–36.

[48] M. Tomac, D. Eller, From geometry to CFD grids: an automated approach for conceptual design, Prog. Aero. Sci. 47 (11) (2011) 589–596, https://doi.org/10.1016/j.paerosci.2011.08.005.

[49] H. Si, TetGen: a quality tetrahedral mesh generator and 3D delaunay triangulator, Tech. rep., User's Manual, wIAS Technical Report No. 13 (2013).

[50] A. Jungo, J. Vos, M. Zhang, A. Rizzi, Benchmarkig new CEASIOM with CPACS adoption for aerodynamic analysis and flight simulation, in: 6th SCAD & READ Workshop, 2016. Warsaw, Poland.

[51] T.W. Sederberg, S.R. Parry, Free-form deformation of solid geometric models, in: SIGGRAPH '86 Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques vol. 20, 1986, pp. 151–160, https://doi.org/10.1145/15922.15903.

[52] T. Pfeiffer, B. Nagel, D. Böhnke, M. Voskuijl, A. Rizzi, Implementation of a Heterogeneous, Variable-Fidelity Framework for Flight Mechanics Analysis in Preliminary Aircraft Design, in: 60, Deutscher Luft- und Raumfahrtkongress, Bremen, Germany, 2011.

[53] M. Voskuijl, J. de Klerk, D. van Ginneken, Flight Mechanics Modelling of the Prandtl Plane for Conceptual and Preliminary Design, vol. 66, Springer-Verlag London Limited, 2012.

[54] M. Voskuijl, G. La Rocca, F. Dircken, Controllability of blended wing body aircraft, in: 26th Congress of International Council of the Aeronautical Sciences, Anchorage, Alaska, 2008.

[55] T. Fengnian, M. Voskuijl, Automated generation of multiphysics simulation models to support multidisciplinary design optimization, Adv. Eng. Inf. 29 (4) (2015) 1110–1125.

[56] M.J. Foeken, M. Voskuijl, Knowledge-based simulation model generation for control law design applied to a quadrotor UAV, Math. Comput. Model. Dyn. Syst. 16 (4) (2010) 241–256.