

## Disciplinary Data Fusion For Multi-Fidelity Aerodynamic Application

*Mengmeng Zhang  
Airinnova AB (AIRI)  
Research Scientist  
Stockholm, Sweden*

*Aidan Jungo  
CFS Engineering (CFSE)  
Research Scientist  
Lausanne, Switzerland*

*Nathalie Bartoli  
ONERA, The French Aerospace Lab, Information Processing and Systems Department  
Research Engineer  
Toulouse, France*

### ABSTRACT

This paper presents some multi-fidelity activities in the field of aerodynamic in the ongoing EU-funded research project AGILE. Different computational tools corresponding to different levels of fidelity are used in the project and the idea is to combine all the data in a global surrogate model to reduce the computational time in an optimization process for instance. The paper focuses on the data fusion tool embedded in the AGILE framework with the choice of the surrogate models, the choice of the tool's fidelity and the choice of the sampling points via an iterative process. The paper details the different steps of the proposed approach to set up and operate the data fusion service proposed to the AGILE partners.

**KEYWORDS:** Data fusion, surrogate models, multi-fidelity, AGILE, aerodynamics

### NOMENCLATURE

|          |  |
|----------|--|
| $\alpha$ | Angle of Attack  |
| (R)MSE   | (Root) Mean Square Error   |
| AGILE    | H2020 EU Project: Aircraft 3rd Generation MDO for Innovative Collaboration of Heterogeneous Teams of Experts |
| API      | Application Programming Interface  |
| CFD      | Computational Fluid Dynamics   |
| CSM      | Computational Structural Mechanics   |
| csv      | Comma Separated Values   |
| DACE     | Design and Analysis of Computer Experiments  |
| DC       | Design Campaign  |
| DOE      | Design Of Experiments  |
| EIF      | Expected Improvement Function  |
| Hi-fi    | High fidelity (data)   |
| IP       | Intellectual Property  |
| Lo-fi    | Low fidelity (data)  |
| MDA      | Multi-Disciplinary Analysis  |
| MDO      | Multi-Disciplinary Optimization  |
| POD      | Proper Orthogonal Decomposition  |
| RANS     | Reynolds Average Navier-Stokes   |
| RSM      | Response Surface Model   |
| SMR      | Surface Model Repository   |

## 1. INTRODUCTION

The modern aircraft design philosophy emphasizes on the collaborations and integrations, where the driving force is to build a smooth flow chain to carry out the MDA (Multi-Disciplinary Analysis) process. For the analysis carried out from each single discipline, there are multi-fidelity tools to do the analysis which give out multi-fidelity outputs. One major aspect, the aerodynamic database construction for stability and control analysis, is this case.

Assessment of maneuverability and agility in the conceptual design stage brings great challenges in the design process regarding the stability and control analysis over the entire flight envelope. A large lookup table of forces and moments must be constructed by Computational Fluid Dynamics (CFD) and 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 (LES) 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 much accuracy while saving computational cost, the simplifying assumptions made to solve the standard Navier-Stokes equation should be as reasonable as possible for each single entry. For example, if one is flying at low speed and non-accelerated small angles of attack, the incompressible potential flow models can give reasonable predictions with significant time saving compared with Euler or RANS equations solving.

The second step concerns all the data obtained by the different CFD methods over the entire flight envelope. The data base must be fused, with dense low-cost & low-fidelity data indicating the trend and sparse high-cost & high-fidelity data correcting values [1].

Multi-fidelity modeling method is concerned with finding the means to navigate the flight state parameter space and cover the entire flight envelope in an effective way. This means to devise a procedure to integrate the results from different simulation tools based on different methods and having different complexity levels. Building the multi-fidelity model has two main steps:

1. populate the aero-data over the whole flight envelope by the dense low-fidelity samples;
2. correct the aero-data by the sparse high-fidelity samples.

The data fusion package is developed for the purpose stated above by aiming at building a reliable surrogate model of the aero-database with minimum number (sparse) of the high-fidelity samples including automatically chosen new sample points. The initial starting points are chosen by regular sampling. The package is used as a "service" in the MDA workflow for innovation aircraft design with the EU AGILE Project [2].

This paper focuses on the developed data fusion package which can be used in the AGILE system. Section 2 describes the background about the collaborative framework and the different tools used for some aerodynamic computations. Section 3 details the data fusion strategy to select new samples at each iteration and build the associated surrogate model. Section 4 focuses on the different modules in the data fusion workflow from the surrogate model to the sampling module. Results are presented in Section 5 and Section 6 summarizes and concludes the work.

## 2. BACKGROUND

AGILE is a 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 brings different knowledge and competences regarding aircraft design and optimization. In this context the use of surrogate models is interesting to assemble results coming from various source and level of fidelity.

Aircraft design deals with several different disciplines, each having their own focus. During this process, data need to be exchanged. Managing the interconnections is complex and error-prone. 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) [3, 4], developed by DLR, is widely used in the framework of AGILE.

The RCE integration environment [5] and workflow manager records the sequence and manages the data transport and translation as well as logging the process. RCE, developed by DLR, makes it easy to set up and run a workflow also using modules in which the engineers are not discipline-experts. It is done via Brics [6], which allows to 'call' a module (and so results) on 'the specialist' computer. This person receives a request to proceed 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.

The data fusion targets the large amount of data to be computed for each design cycle in MDO by constructing *surrogate models* for the data production based on variable fidelity analysis tools in the MDO framework together with Kriging, co-Kriging, and adaptive modeling techniques for *fusing* the outputs from each tool. This technique is well established (see for instance the recent review [7]) and constitutes its own challenges in variable fidelity MDA in AGILE system.

In this paper when "variable fidelity" is referred, we always use the nomenclature  $L^*$  to indicate the fidelity Level:

L0: empirical or handbook method;

L1: linearized-equation method, e.g. linear aerodynamics, or vortex lattice method (VLM);

L2: higher fidelity equation solvers with less simplifications (e.g. Euler equations solvers);

L3: highest fidelity equation solvers (e.g. RANS solvers).

To demonstrate proof of concept, AGILE partners have formed teams to run data fusion on RCE framework workflows via Brics, which will provide data fusion solutions for the whole AGILE MDA system. The authors, who are the main partners for the data fusion task, will show the data fusion technology that is applied and used in the AGILE system in this paper.

### 3. SURROGATE MODELS

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 multiplication to much quicker approximate calculations by repeated table lookup, interpolation, and addition.

Data fusion means the integration of results from different simulation models into one surrogate model. Building a surrogate model is a particular form of data fusion. In AGILE the surrogate models are constructed by different methods provided by partners, which is known as RSM service. The RSM function will be called in the formation of:

$$[\text{force and moment coefficients}] = \text{RSM}(\alpha, \beta, Re, Mach, \dots, D) \quad (1)$$

where the aerodynamic variables are  $\alpha$  the angle of attack,  $\beta$  the angle of sideslip,  $Re$  the Reynolds Number. The data set is denoted by  $D$ . It is impractical for intensive use to ship the large data set  $D$  via Brics Call for every evaluation of  $g$ . Therefore a caching mechanism must be implemented and protocol decided on.

Surrogate modeling has two distinct steps, first the *training* to produce the  $g$ -function, and second, the use of it. The set of inputs used is the training set. In this paper concerning the AGILE applications considered so far,  $y$  is the (computational) data 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  $\hat{y}$  requires at least a flow solution, and possibly a complete aero-servo-elastic simulation. The training should produce a RSM compliant with some accuracy requirement with minimal cost of 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 design of algorithms for the choice of a training set, as described in [8] by AGILE partners providing the different RSMs.

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

$$y = g(x, D) \quad (2)$$

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

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

For the new  $D$ , algorithm  $S$  suggests new samples towards satisfying accuracy requirements:

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

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

The RSM kit contains computer codes for RSM evaluator  $g$ , data fuser algorithm  $M$  and sampling algorithm  $S$ , plus a proviso to evaluate  $f$ . A developed instance, which is ready to run, and improved on further, will contain  $g$ ,  $M$ ,  $S$  and the data set  $D$ . Fig. 1 shows one data fusion workflow built by (partners') interchangeable modules and exchanging data via Brics Call. It should be noted that sending  $D$  via Brics for each call is impractical. 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 sent by 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.

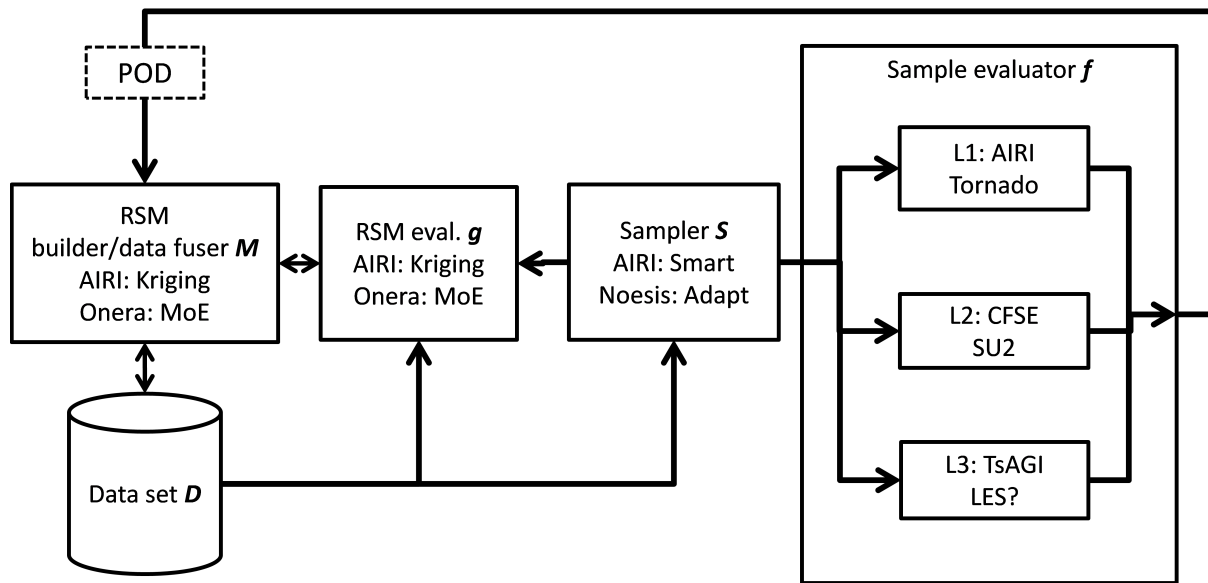


Figure 1: Workflow for data fusion in the MDA chain, interchangeable modules exchanging data via Brics Call.

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 needs distribution and access restrictions. This is discussed in details in the documentation of the RSM Repository [9, 10].

#### 4. THE DATA FUSION

The data fusion package includes the development/delivery of

- (1) surrogate model builder/fuser  $M$  and model evaluator  $g$ ;
- (2) sampling module  $S$ ;
- (3) samples computation module  $f$  (aerodynamic module);
- (4) graphic feedback module.

This package can be used for fusing computed tables of forces and moments, or distributed loads for which POD is used for data compression [11]. Namely, the data fusion service can be used for:

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

2. fusing the **loads** from different tools (sources) and fidelities and storing the surrogate models. The load distributions will be first compressed by POD and then be fused. The POD toolbox contains:

- inputs: a folder with one column-based .csv file for each flight condition ( $\alpha, \beta, Re, Mach$ ), which stores the geometric location on the aircraft (x,y,z) and its corresponding value (force, span-load, or the pressure coefficient  $C_p$ ).
- outputs: (i) the sets of computed POD basis vectors; and (ii) the surrogate model for the POD coefficients of the samples.

#### 4.1. Data fusion in the MDA chain

Fig. 2 spells out the data fusion service in the MDA chain in AGILE system. It has three core modules:

- **Preparation Module:** It prepares the training data storing in the *design study branch*;
- **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 [12] provided by Airinnova AB (AIRI) in *Matlab* based on the DACE toolbox [13], 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 Py *scikit-learn* toolbox. So that it can be stored, transferred, shared and re-produced by partners without Matlab license.
  - (2) Mixture of Experts (MOE) method for surrogate modeling provided by ONERA [14] as an executable file. It will output one (binary) file containing the MOE model parameters corresponding the MOE surrogate trained with incoming training data. The MOE strategy [15, 16] uses a clustering of the training basis into regions where the function to be approximated is expected to be continuous or at least more simple. It strongly relies on the Expectation-Maximization (EM) algorithm for Gaussian mixture models (GMM). With an aim of regression, the inputs are clustered together with their output values by means of parameter estimation of the joint distribution. A local expert is then built (linear, quadratic, cubic, radial basis functions, or different forms of kriging) on each cluster and all the local experts are finally combined using the Gaussian mixture model parameters found by the EM algorithm to get a global model. MOE has been made available to AGILE partners for different applications [17].
- **Aerodynamic Module:** In this module the new samples are computed by aerodynamic tools. In this paper, the L1 aerodynamic data are provided by Tornado, which is a Vortex Lattice Method code [18] from AIRI; L2 aerodynamic data are provided by L2 Tool SU2 [19] which is a computational fluid dynamics simulation software operated at CFSE. Other aerodynamic tools can be integrated into this MDA chain by modifying their API to the data fusion package.

#### 4.2. The "smart sampling"

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 (R)MSE. The sampling identifies a number of sampling locations. The response at the sampling points can be evaluated in parallel. Due to the inherent nature of the sampling algorithms  $S$  (shown below), the  $S$  is strongly coupled with the RSM evaluator  $g$ , namely, Kriging or co-Kriging surrogate models. All the tools use data read/written in CPACS format, which enables easy switching tools. A *variable* fidelity model for data fusion is constructed following steps:

1. *Initialization.* Initialize the samples, populate the whole domain by the lowest fidelity samples.
2. *Sampling.* Propose new sample points (at high fidelity level) based on *smart sampling*, and evaluate them.
3. *Surrogate model construction/update.* Make a surrogate model from the above samples by Kriging/co-Kriging interpolation.

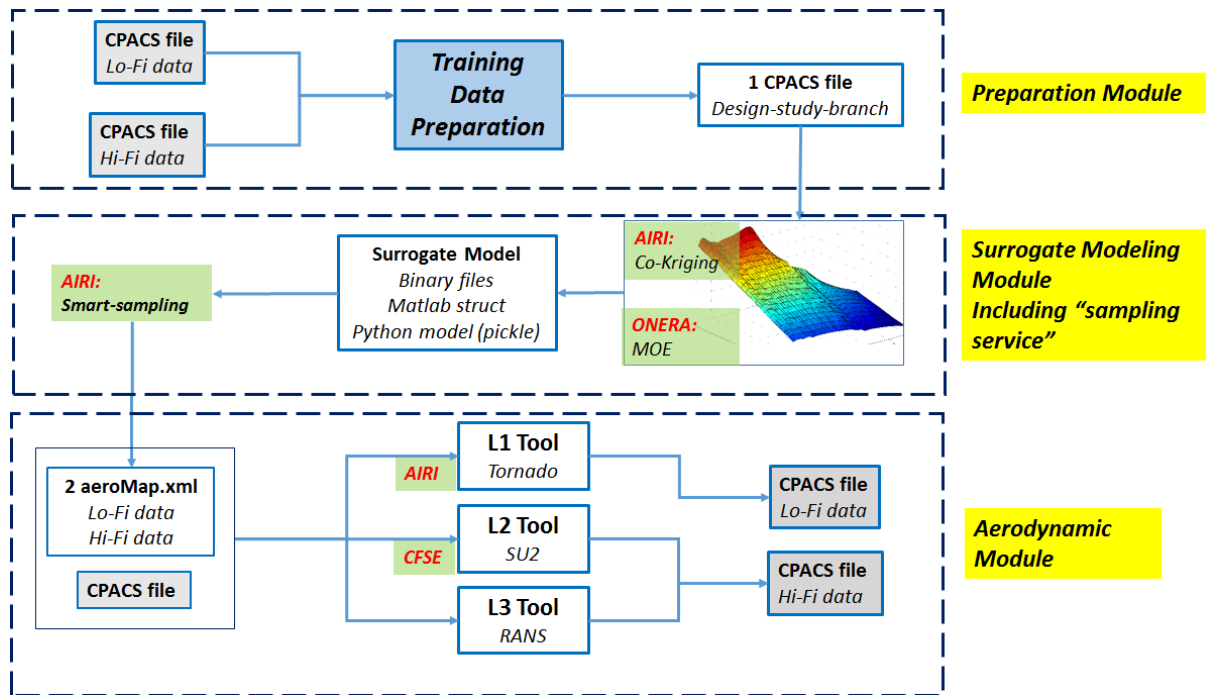


Figure 2: Data fusion service in the MDA chain with 3 stages.

4. *Go to next fidelity level.* Use the surrogate model obtained from previous step as lower fidelity samples. Iterate from 2-4 steps.
5. *Stop* if the convergence criteria are met at the highest fidelity level.

Fig. 3 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*. If the stopping criteria are met, the iteration ends. Otherwise new samples are suggested according to the listed methods in Fig. 3. 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.

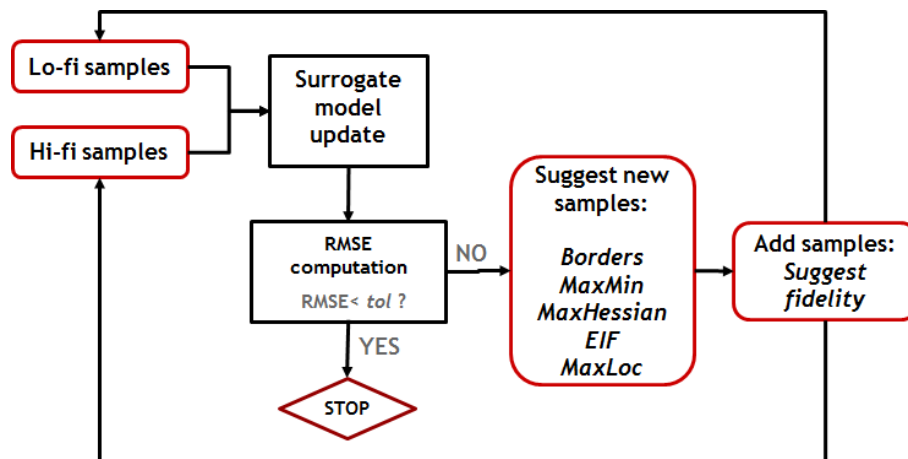


Figure 3: The "smart sampling" technology workflow.

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

- *Borders*: only applies when loading the initial samples (iteration=0). 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).

- *MaxMin*: finds the local optimum 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 if tool fidelity is not specified.
- *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 (see Section 6). This method only becomes active if *Find MaxMin* fails.
- *EIF*: finds the EIF location [12] 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. The new samples shall be computed by L1 tool (lo-fi) by default.
- *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 previous iterations:  $RMSE(k+1) > f \cdot RMSE(k)$  where  $0 < f < 1$  is the improvement factor, thus the sampling method shall be "switched" to *MaxLoc*.

As stated above, the new samples suggested by all of the methods (except *Borders*) 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 lo-fi until error estimate satisfies lo-fi criterion. Then hi-fi points are filled in until the error estimate between response surface and hi-fi is small enough (i.e., fulfills the hi-fi criterion). Note that this may "waste" lo-fi calculations in regions where lo-fi is bad so hi-fi is necessary.

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

- 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,  $L_k - > L_{k+1}$ , where  $k$  is the current fidelity level.
- Another indication is that the maximum RMSE is *not* significantly improved compared with the previous iteration. This means either the methods or the fidelity shall be "switched".

## 5. RESULTS

The data fusion is implemented in the AGILE workflow as a "customer" since it will use the service provided by RCE and Brics. On the other hand, it will provide the data fusion service as a "provider" within the workflow. The results from both aspects will be shown in this section, with the concept of "fusing" aerodynamic data.

### 5.1. Data fusion service implemented on AGILE MDA system

The Data fusion service, which can only work in the best way that it correctly links to other tools/modules. To achieve that, it makes use of and is implemented on AGILE MDA system. One important aspect leads to a successful integration/implementation is to prepare proper API for all the tools involved. As shown in Fig. 4: it has (1) CPACS file which contains the (previous iteration) dataset  $D_k$ , (2) Data Fusion algorithms  $M$  &  $S$ , and (3) L1-L3 tools to enrich the dataset. Those modules/tools are connected by RCE, and callable via Brics. There are some details one needs to consider carefully. For example, for fusing aerodynamic coefficients, the inputs are CPACS files including the aircraft configuration information, the Data Fusion (fuser  $M$ ) only takes aero-coefficients (dataset) computed from L1-L3 tools, and the sampling  $S$  only suggests new flight domain(s) to be computed. There are questions need to be asked (and solved):

1. How to convert the configuration data (CPACS) into a valid mesh file?
2. How to transfer/pass the mesh file to the corresponding CFD solver?
3. How to transfer/pass the newly suggested domains (from  $S$ ) to the aero-solvers?
4. How to identify the validity of domains of the surrogate models?

Question 1 and question 3 are solved either by modifying tools' API to adapt to CPACS format (common i/o), or by agreeing on a (newly defined) common i/o format for all the tools. For question 2, currently (in this paper) the solution is that we choose one single partner (CFSE) to do both mesh creation and CFD computations offline (no mesh transfer via Brics happens). Future work will cover question 3 via different partners and considerations for the identification of the domains validity of the surrogate models.

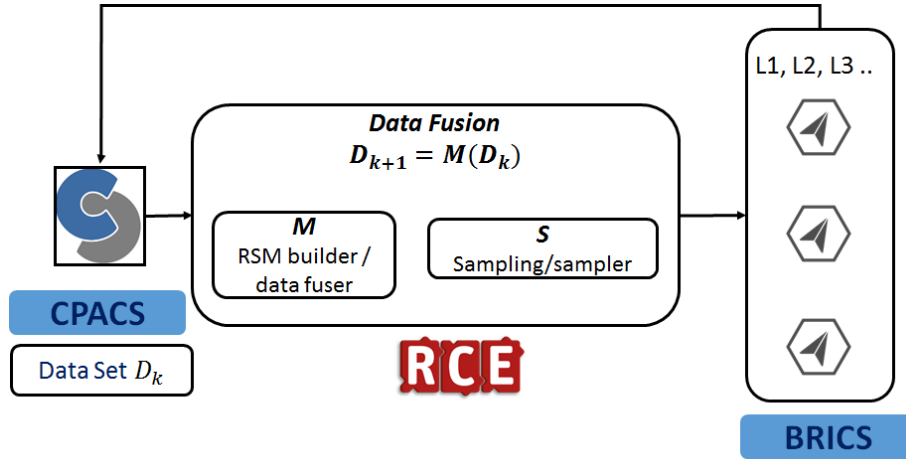


Figure 4: Illustration of the Data Fusion implemented in the AGILE MDA system.

## 5.2. Test case: aerodynamic coefficients fusion

Figure 5a shows the aircraft configuration used in this test case: the aircraft is the reference aircraft used in AGILE Design Campaign 1 (DC-1), it was analyzed and simulated by AGILE MDA system, but no experimental data are available. The reference aircraft, or DC-1 aircraft, is defined in CPACS format. The initial samples (aerodynamic coefficients, which stores in the node */analyses/aeroPerformanceMap*) are computed with regular sampling by L1 tool Tornado and L2 tool SU2 (Fig. 5b).

The initial sample domain between Mach [0.2 0.9] and AoA [-5 15] deg has 20 lo-fi samples and 6 hi-fi samples and are stored in 2 CPACS files respectively (one for each fidelity). The surrogate model is built by co-Kriging from AIRI (Data Set  $D_k$ ). The stopping criterion is set that the maximum deviation ( $err$ ) is  $\max(err) < 10\%$ , which,

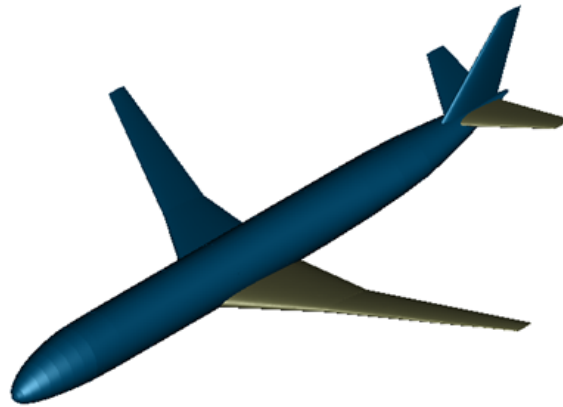
$$err = \frac{RMSE}{\max(y_{hi-fi})} \quad (5)$$

The RMSE is computed by cross validation, that

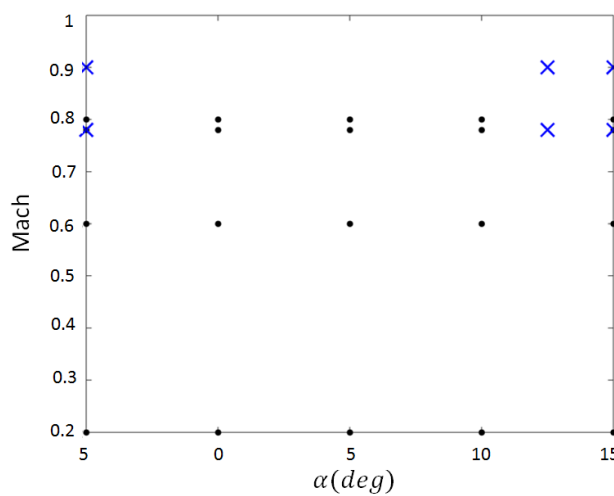
$$\mathbf{RMSE}: \hat{s}(x^*) = (\rho^2 \hat{\sigma}_c^2 + \hat{\sigma}_d^2 - \mathbf{c}^T \mathbf{C}^{-1} \mathbf{c} + \frac{1 - \mathbf{1}^T \mathbf{C}^{-1} \mathbf{c}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}})^{1/2} \quad (6)$$

where  $\mathbf{C}$  is the covariance matrix for data combining both low and high-fidelity observations, and  $\mathbf{c}$  is the  $m_c + i$ th column of  $\mathbf{C}$ , if the prediction is made at the  $i$ th high-fidelity points, and the Lo-Fi data have  $m_c$  design sites. Details can be referred in Chapter 8 in [12].

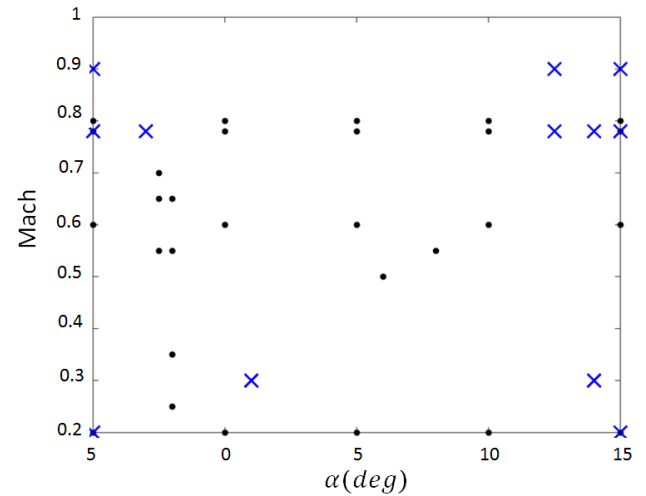
Fig 6 shows an iterative process of data fusion for aerodynamic coefficient  $C_D$  (Fig. 7 is the corresponding error). After 7 iterations ( $k = 7$ ), the stopping criteria is met and the final surrogate model  $M$  is built from the enriched data set  $D_{k=7}$ . The final sample domain is shown in Fig. 5c, with 29 lo-fi samples and 12 hi-fi samples. The sampling methods for each iteration are chosen according to rules stated in Section 4.2. Note that between iteration 2 and 3, the method to choose new sample locations was switched from *Max-Min* to *MaxHessian*, which means the *Max-Min* method does not find a local optimum of the response surface, so that the method falls to the next one in the sequential way, *Find MaxHessian*. At iteration 6, the RMSE was not significantly improved compared with previous iteration (see Fig. 7), so (for the next iteration) the method was switched to Method *MaxLoc*.



(a) 3-D view of DC-1 aircraft.

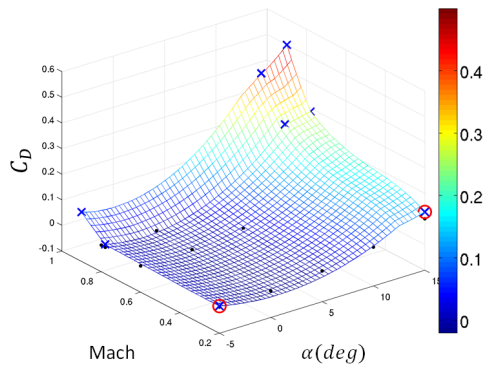


(b) Domain of the initial samples.

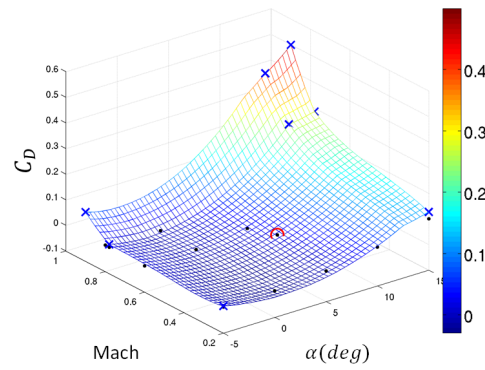


(c) Domain of the final samples.

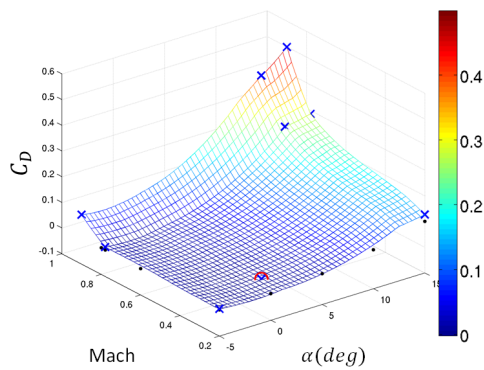
Figure 5: DC-1 aircraft configuration and sample domains (black dots are relative to lo-fi samples and blue cross to hi-fi samples).



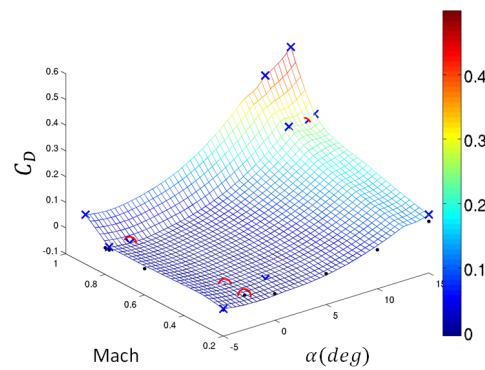
(a) Iteration 1. Method: Check the border.



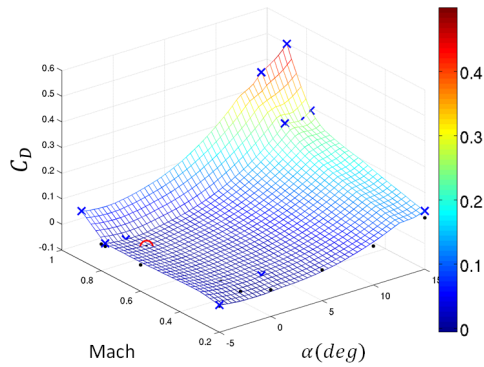
(b) Iteration 2. Method: Find Max-Min.



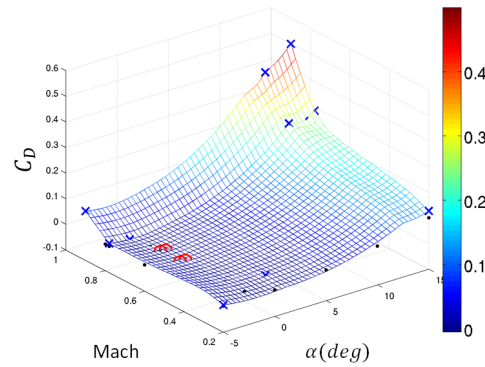
(c) Iteration 3. Method: Find Max Hessian



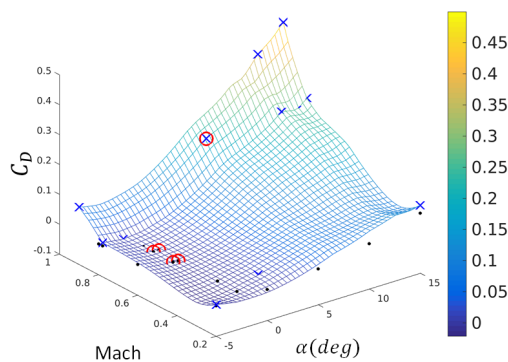
(d) Iteration 4. Method: Find Max-Min.



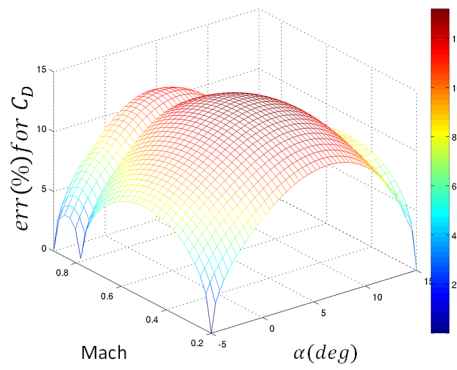
(e) Iteration 5. Method: Find Max-Min.



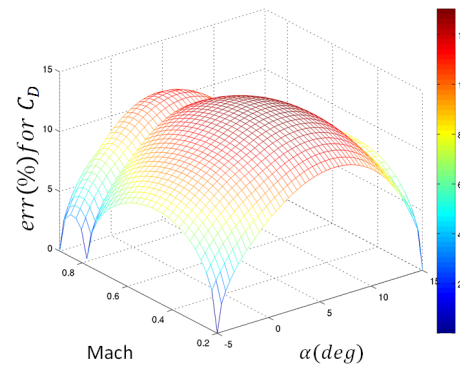
(f) Iteration 6. Method: Find Max-Min. RMSE is *not* significantly improved.



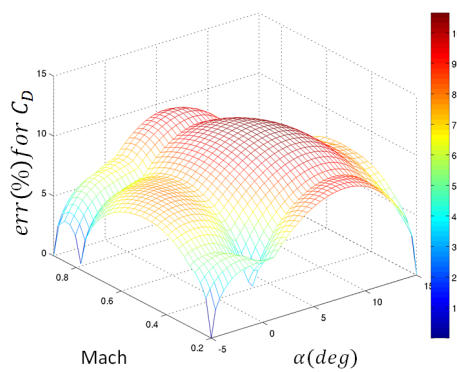
(g) Iteration 7. Method switched: MaxLoc. Stopping criteria met.



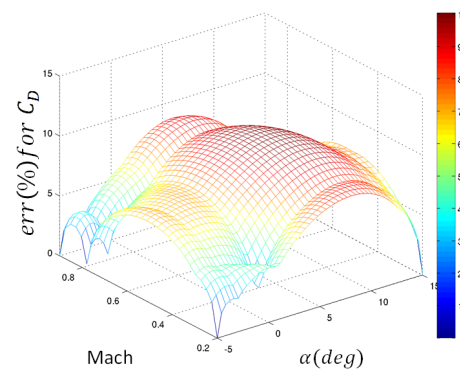
(a) Iteration 1.



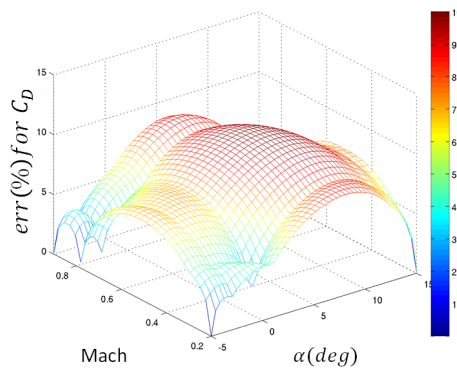
(b) Iteration 2.



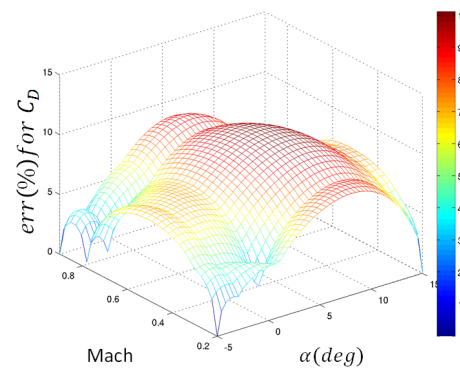
(c) Iteration 3.



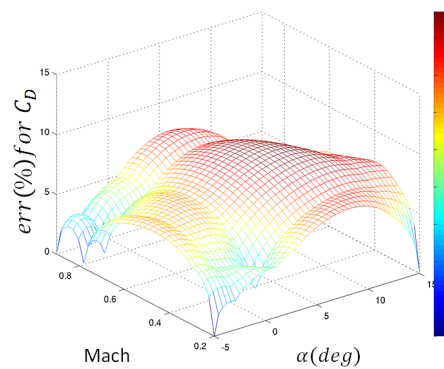
(d) Iteration 4.



(e) Iteration 5.



(f) Iteration 6.



(g) Iteration 7.

Figure 7: The errors for DC-1 aircraft data fusion iteration on  $C_D$  coefficient.

## 6. CONCLUSIONS AND FUTURE DEVELOPMENTS

This paper represents how the data fusion technique applies to AGILE MDA system. This is particularly useful when the computational dataset can not be provided locally, and it is the usual case when the multiple fidelity analysis happens, different tools are run at different sites, the dataset can get enriched simultaneously while the system is running and finally a fused dataset can be obtained. This fused dataset can be further used for example, as aero-inputs when simulating the flight performance.

The future work will cover the identification of the domain validity of the surrogate models. It is related to the corresponding problems, for example, to fuse the aerodynamic coefficients and derivatives for stability and control analysis, the domain space will cover the whole flight envelope, especially at the boundaries, i.e., the extreme flight conditions. In this case, the DC-1 aircraft as a conventional transport airliner, the validity of the domain, will not cover, for instance, the high angles of attack at high Mach numbers. Otherwise the sampling rules will fail and the new suggested samples will always locate around those locations since they are highly non-linear. Mapping the existing flight envelope to the domain space can solve this problem. If the flight envelope is not explicitly defined, a reasonable guess by e.g. aerodynamicist or flight mechanist will also work. To implement it based on the existing pattern of the cube-type domain, one may shrink/stretch each dimension of the design space so the number of the intervals remains unchanged.

The future developments will also involve the comparisons of the surrogate models from different dimensions:

- comparison of the surrogate models based on samples provided by (1) different CFD tools within the same fidelity with the same mesh (e.g., L2 samples can come from CFSE (SU2) and TsAGI); or (2) the same CFD tool with different meshes (provided by different meshing tools).
- comparison of the surrogate models improved by different sampling techniques/services.

To conclude, the data fusion service is now proposed to the AGILE partners as a remote service and will be used for the novel aircraft configurations in the frame of Design Campaign 3 of AGILE.

## ACKNOWLEDGMENTS

The research presented in this paper has been performed in the framework of the AGILE project (Aircraft 3rd Generation MDO for Innovative Collaboration of Heterogeneous Teams of Experts) and has received funding from the European Union Horizon 2020 Programme (H2020-MG-2014-2015) under grant agreement n° 636202. The Swiss participation in the AGILE project was supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0162. The authors are grateful to the partners of the AGILE consortium for their contribution and feedback.

## REFERENCES

1. Ghoreyshi, M., Badcock, K., and Woodgate, M., "Accelerating the numerical generation of aerodynamic models for flight simulation," *Journal of Aircraft*, Vol. 46, No. 3, 2009, pp. 972.
2. "AGILE EU Project portal," <http://www.agile-project.eu>.
3. Böhnke, D., Nagel, B., Zhang, M., and Rizzi, A., "Towards a collaborative and integrated set of open tools for aircraft design," *51st AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition, grapevine (Dallas/Ft. Worth Region), Texas, USA, 2013*, pp. 7–10.
4. "CPACS – A Common Language for Aircraft Design," <https://software.dlr.de/p/cpacs/home/>, accessed 28 August 2017.
5. Seider, D., Fischer, P., Litz, M., Schreiber, A., and Gerndt, A., "Open Source Software Framework for Applications in Aeronautics and Space," *IEEE Aerospace Conference*, March 2012.
6. Baalbergen, E., Kos, J., Louriou, C., Campguilhem, C., and Barron, J., "Streamlining cross-organisation product design in aeronautics," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 2016, pp. 0954410017716480.
7. Fernández-Godino, M. G., Park, C., Kim, N.-H., and Haftka, R. T., "Review of multi-fidelity models," *arXiv preprint arXiv:1609.07196*, 2016.
8. Lefebvre, T., Bartoli, N., Dubreuil, S., Panzeri, M., Lombardi, R., Della Vecchia, P., Nicolosi, F., Ciampa, P., Anisimov, K., and Savelyev, A., "Methodological enhancements in MDO process investigated in the AGILE European project," *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Denver, Colorado, USA, AIAA 2017–4140*, 2017.

9. "Deliverable D6.7 AGILE: Architecture for dynamic configuration of MDO workflow/AGILE framework components, In progress," August 2017.
10. "Surrogate Model Repository (SMR) Status and Demonstration," presentation at AGILE M25 Meeting, Moscow, Russian, July 2017.
11. Zhang, M., Tomac, M., Wang, C., and Rizzi, A., "Variable Fidelity Methods and Surrogate Modeling of Critical Loads on X-31 Aircraft," *51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, Grapevine, Texas, USA*, Vol. 1081, 2013, pp. 7–10.
12. Forrester, A., Keane, A., et al., *Engineering design via surrogate modelling: a practical guide*, John Wiley & Sons, 2008.
13. Lophaven, N., Nielsen, H., and Sondergaard, J., "DACE: A Matlab Kriging Toolbox, Version 2.0, August 1, 2002," *Technical University of Denmark*, 2000.
14. Bartoli, N., Kurek, I., Lafage, R., Lefebvre, T., Priem, R., Bouhrel, M.-A., Morlier, J., Stilz, V., and Regis, R., "Improvement of efficient global optimization with mixture of experts: methodology developments and preliminary results in aircraft wing design," *17th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Washington D.C., USA, June 2016.
15. 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 and Multidisciplinary Optimization*, Vol. 43, No. 2, 2011, pp. 243–259, 10.1007/s00158-010-0554-2.
16. Liem, R. P., Mader, C. A., and Martins, J. R. R. A., "Surrogate Models and Mixtures of Experts in Aerodynamic Performance Prediction for Mission Analysis," *Aerospace Science and Technology*, Vol. 43, 2015, pp. 126–151.
17. Lefebvre, T., Bartoli, N., Dubreuil, S., Panzeri, M., Lombardi, R., Lammen, W., de Wit, B., Zhang, M., van Gent, I., and Ciampa, P. D., "Overview of MDO enhancement in AGILE project: a clustered and surrogate-based MDA use case," *6th CEAS Aerospace Europe Conference*, paper no.956, Bucharest, Romania, October 2017.
18. T., M., "Using Internet Interactions in Developing Vortex Lattice Software for Conceptual Design," 2003, Software available from <http://redhammer.se/tornado/>, July 2009.
19. Thomas D. Economon, Francisco Palacios, S. R. C. T. W. L. J. J. A., "SU2: An Open-Source Suite for Multiphysics Simulation and Design," *AIAA Journal*, March, Vol. 54, No. 3 : pp. 828-846, 2015.
20. Tomac, M. and Eller, D., "From Geometry to CFD Grids: An Automated Approach for Conceptual Design," *Progress in Aerospace Sciences*, Vol. 47, No. 11, 2011, pp. 589–596, doi: 10.1016/j.paerosci.2011.08.005.
21. Si, H., "TetGen: a Quality Tetrahedral Mesh Generator and 3D Delaunay Triangulator," Tech. rep., User's Manual, WIAS Technical Report No. 13, 2013.

## APPENDIX A: HESSIAN OF KRIGING PREDICTOR

Recall Formula 2.16 in [13]<sup>a</sup>:

$$\hat{y}(x) = f(x)^T \beta^* + r(x)^T \gamma^* \quad (7)$$

where  $f(x)$  is the regression basis function at  $x$ ;  $r(x)$  = vector of values of correlation functions between  $x$  and data sites  $S^i, i = 1, 2, \dots, m$ ;  $\beta^*$  and  $\gamma^*$  are the coefficients computed by DACE.

The Hessian  $H_{n \times n}$  per data dimension  $p$ :

$$\frac{\partial^2}{\partial x_k \partial x_\ell} \hat{y}(x) = \beta^{*T} \cdot \frac{\partial^2}{\partial x_k \partial x_\ell} f(x) + \gamma^{*T} \cdot \frac{\partial^2}{\partial x_k \partial x_\ell} r(x) \quad (8)$$

where  $f, r \in \mathbb{R}^{m \times 1}$ ,  $\beta^*, \gamma^* \in \mathbb{R}^{m \times p}$ .

So the computation of Hessian  $H$  is converted to the computation of second derivatives of  $f(x)$  and  $r(x)$ .

### Second derivatives of $f(x)$ computation

If *regpoly0* or *regpoly1* (regression model polynomial order is 0 or 1),  $\frac{\partial^2}{\partial x_k \partial x_\ell} f(x) = 0$ .

If *regpoly2* (regression model polynomial order is 2),  $\frac{\partial^2}{\partial x_k \partial x_\ell} f(x) = \text{const.}$

<sup>a</sup>Nomenclatures and symbols are consistent with DACE toolbox manual

### Derivatives of $r(x)$ computation

First derivatives are already available in DACE if *predictor* is given one  $x$  only to predict (Jacobian of the correlation model computed in **function corr.m**), pp.17-18 in [13]. Data sites  $S^i$  with coordinates  $S_j^i, i = 1, 2, \dots, m; j = 1, 2, \dots, n$

$$\mathbf{r}^T = [r_1(x), r_2(x), \dots, r_m(x)] \quad (9)$$

where,

$$r_i(x) = \prod_{j=1}^n R(\theta_j, x_j - S_j^i) \quad (10)$$

The first derivative:

$$\frac{\partial r_i}{\partial x_k} = R'(\theta_k, x_k - S_k^i) \prod_{j \neq k}^n R(\theta_j, x_j - S_j^i) \quad (11)$$

The second derivatives:

$$\frac{\partial^2 r_i}{\partial x_\ell \partial x_k} = \begin{cases} R''(\theta_k, x_k - S_k^i) \prod_{j \neq k}^n R(\theta_j, x_j - S_j^i), \ell = k \\ R'(\theta_\ell, x_\ell - S_\ell^i) R'(\theta_k, x_k - S_k^i) \prod_{j \neq k, \ell}^n R(\theta_j, x_j - S_j^i), \ell \neq k \end{cases} \quad (12)$$

So the **corr** function must be coded to deliver  $R, R'$  and  $R''$  and some new parameters to allow **predictor** to deliver all the data for many  $x$  at a time. A variant **predictorHess** is made based on DACE function **predictor** to calculate the Hessian of the *predictor*. The maximum Hessian method for sampling is strongly coupled with the surrogate modeling.

## APPENDIX B: L1-L2 TOOLS DESCRIPTION

### L1 Tool Tornado:

The L1 aerodynamic tool used for this workflow is the VLM code Tornado [18]. Tornado is coded in Matlab and within AGILE, it has been fully adapted to CPACS, integrated onto RCE and is callable via Brics.

**L2 Tool SU2:** The L2 aerodynamic tool used for this workflow is the open-source code SU2 develop by the university of Stanford, it is able to solve unstructured meshes. To use SU2 in the datafusion process, the CPACS file received is converted, then opened with the mesh generator software SUMO [20]. A surface mesh is create on the aircraft skin and is expended with the help of TetGen [21] to create an unstructured Euler mesh. Additional calculation parameters, such as the convergence criteria, can be modified before launching the SU2 solver. Once results are satisfying enough (generally convergence of the residuals and forces are checked), coefficients are extracted automatically and placed in the CPACS file, which is sent back to be integrated in the new surrogate model. All theses operation take place in a RCE workflow shown on Fig. 8, it can be modified and adapted to request.

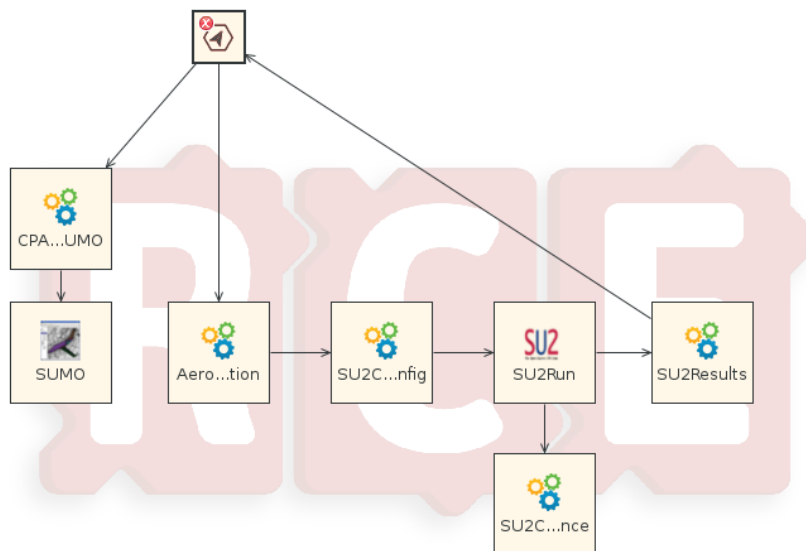


Figure 8: L2 aerodynamic workflow operated at CFSE