11th World Congress on Computational Mechanics (WCCM XI)
5th European Conference on Computational Mechanics (ECCM V)
6th European Conference on Computational Fluid Dynamics (ECFD VI)
E. Oñate, J. Oliver and A. Huerta (Eds)

SURROGATE MODELS FOR SPACECRAFT AERODYNAMIC PROBLEMS

Mikhail Belyaev[1,2,3], Evgeny Burnaev[1,2,3], Ermek Kapushev[1,2], Stephane Alestra[4], Marc Dormieux[5], Antoine Cavailles[5], Davy Chaillot[5] and Eugenio Ferreira[5]

[1]DATADVANCE, llc, Pokrovsky blvd. 3, Moscow, 109028, Russia
e-mail: mikhail.belyaev@datadvance.net, evgeny.burnaev@datadvance.net, ermek.kapushev@datadvance.net
web-page: http://www.datadvance.net

[2]Institute for Information Transmission Problems, Bolshoy Karetny per. 19, Moscow, 127994, Russia
web page: http://www.iitp.ru

[3]PreMoLab, MIPT, Institutsky per. 9, Dolgoprudny, 141700, Russia
web page: http://www.premolab.ru

[4]Airbus Group, St. Martin du Touch, 316 route de Bayonne, 31060 Toulouse Cedex 9, France
e-mail: stephane.alestra@eads.net
web page: http://www.airbus.com

[5]Airbus Defence and Space, Les Mureaux
e-mail: marc.dormieux@astrium.eads.net, antoine.cavailles@astrium.eads.net, davy.chaillot@astrium.eads.net, eugenio.ferreira@astrium.eads.net
web-page: http://airbusdefenceandspace.com/

Key words: Surrogate modelling, data fusion, incomplete factorial DoE, tensor approximation

Abstract. This work concerns a construction of surrogate models for a specific aerodynamic data base. This data base is generally available from wind tunnel testing or from CFD aerodynamic simulations and contains aerodynamic coefficients for different flight conditions and configurations (such as Mach number, angle-of-attack, vehicle configuration angle) encountered over different space vehicles mission. The main peculiarity of aerodynamic data base is a specific design of experiment which is a union of grids of low and high fidelity data with considerably different sizes. Universal algorithms can't approximate accurately such significantly non-uniform data. In this work a fast and accurate algorithm was developed which takes into account different fidelity of the data and special design of experiments.

1 Introduction

Airbus Defence and Space, in the frame of the development of various spacecraft vehicles, is producing aerodynamic models. Different kinds of data are used to build these models:

- Experimental results derived from WTT campaigns, with a high level of confidence.
- Numerical data resulting from CFD simulations, with a lower level of confidence, which depends also on the used numerical method (Euler / RANS, flow regime, etc.)

CFD computations and even more WT campaigns are very costly and it is sometimes not conceivable to perform all flight configurations (surfaces deflections, e.g.) on the whole flight range. An inter-extrapolation process is then necessary to build a complete aerodynamic model. With a lot of input parameters (Mach, Angle of Attack, Angle of Sideslip, surfaces deflections, etc.), this process can be quite time consuming and may lead to inconsistent results with a classical approach. Furthermore available data can constitute anisotropic grids which lead to severely harden the inter-extrapolation process.

The objective is therefore to build a consistent model taking into account all available data, with a fast and rationale method.

Airbus Defence and Space is thus interested by the generation of surrogate models from given spacecraft aerodynamic database. The AErodynamic DataBase (AEDB) provides the aerodynamic coefficients of the vehicle for the different flight conditions and vehicle configurations (notably aerodynamic control surfaces deflections) encountered over the whole mission domain. This type of database is typically included within global vehicle behaviour models. These behaviour models are then used as input to various system studies, such as trajectory and performance analysis, or handling qualities and flight control system analysis.

From a practical standpoint, the AEDB generation process raises the following two fold challenges:

- Multidimensional interpolation/extrapolation: how to cover a prescribed full flight envelope defined in the multidimensional space of flight conditions/vehicle configurations, on the basis of scattered discretized input data?
- Multiple data combination: how to build a consistent and homogeneous aerodynamic database on the basis of multiple input data sets with different levels of fidelity?

The final goal is to obtain a surrogate model that can automatically interpolate multidimensional data, union of anisotropic grids with non-uniform data and different level of fidelity, and therefore to build a consistent and homogeneous aerodynamic model that can cover all intermediate points within the flight envelope. Last but not least, the surrogate model shall ensure that the exhibited outputs always remain within realistic limits (i.e.

remain meaningful from aerodynamics behaviour standpoint), whatever the input vector content.

2 SURROGATE MODELLING

A surrogate modelling is one of the approaches to solving problems of engineering design actively developing in recent years [1]. In this approach a complex physical phenomenon is described by a simplified (surrogate) model constructed using data mining techniques and a set of examples representing results of a detailed physical modelling and/or real experiments. The problem of approximation of a multidimensional function using a finite set of pairs "point" – "value of the function at this point" is one of the main problems to be solved during construction of the surrogate model.

In this section we will give mathematical statement of approximation and data fusion problems.

2.1 Approximation problem

Let us consider continuous function $g : D \in \mathbb{R}^d \to \mathbb{R}$, where D is a compact set. Let us refer to a set of points Σ and a set of function values at points from the set Σ as a *training set* S

$$S = \{\mathbf{x}_i \in \Sigma, y_i = g(\mathbf{x}_i)\}_{i=1}^N = \{\Sigma, g(\Sigma)\}.$$

Set of input points Σ we will call a *design of experiments* (DoE).

In this paper we consider approximation problem in the following statement.

Problem 1. *Given the training set S, class of functions F and penalty function $P_\lambda : F \to \mathbb{R}$ find such $f^* \in F$ that minimizes the error function $R(f, \Sigma, g(\Sigma))$*

$$f^* = \arg\min_{f \in F} R(f, \Sigma, g(\Sigma)) = \arg\min_{f \in F} \sum_{\mathbf{x} \in \Sigma} (g(\mathbf{x}) - f(\mathbf{x}))^2 + P_\lambda(f) \qquad (1)$$

Introduced penalty function allows to control variability of the approximation model. For example, it can be a norm of the second derivatives of f (see, for instance, smoothing splines [2]) or norm of f in some Hilbert space (kernel ridge regression [3]).

2.2 Data fusion

The data sets considered in this work contains output values obtained from two different sources (for example experimental measurements and CFD simulations). Both sources model the same physical process. However one of them (experimental measurements) is supposed to be more accurate than another (CFD simulations). We will refer to the more accurate one as *high fidelity (HF) model* and denote it by $g_h(\mathbf{x})$. We will refer to the second source as *low fidelity (LF) model* and denote it by $g_l(\mathbf{x})$.

So the training set S is split into 2 parts

- high fidelity sample $S_h = \{\mathbf{x}_i^h, y_i^h = g_h(\mathbf{x}_i^h)\} = \{\Sigma_h, g_h(\Sigma_h)\}$ and

- low fidelity sample $S_l = \{\mathbf{x}_i^l, y_i^l = g_l(\mathbf{x}_i^l)\} = \{\Sigma_l, g_l(\Sigma_l)\}$.

A data fusion task is given high fidelity and low fidelity data sets S_h and S_l construct an approximation $\hat{f}(\mathbf{x})$ of $g_h(\mathbf{x})$.

Usage of LF points allows to build more accurate surrogate models as they contain information about the physical model in regions which don't have HF points. To measure how accurate high or low fidelity value is we introduce confidence levels of the sample w_h (for HF points) and w_l (for LF points), $w_h + w_l = 1, w_h, w_l > 0$. They express our confidence about the accuracy of the source that produced output value at given point and can be thought of as a probability of being the true value of the physical characteristic at given point. For experimental measurements the weights should be greater than the weights for CFD simulations since they are more accurate.

2.3 Design of experiments

One of the peculiarities of the problem considered in this work is a specific DoE. First of all, let us introduce some definitions. Let us refer to sets of points $\sigma_k = \{\mathbf{x}_{i_k}^k \in D_k\}_{i_k=1}^{n_k}$, $D_k \subset \mathbb{R}^{d_k}$, $k = \overline{1, K}$ as *factors*. A set of points Σ_{full} is referred to as a *factorial design of experiments* if it is a Cartesian product of factors

$$\begin{aligned}\Sigma_{full} = &\sigma_1 \times \sigma_2 \times \cdots \times \sigma_k = \\ &\{[\mathbf{x}_{i_1}^1, \ldots, \mathbf{x}_{i_K}^K], \{i_k = 1, \ldots, n_k\}_{k=1}^K\}.\end{aligned} \tag{2}$$

The elements of Σ_{full} are vectors of a dimension $d = \sum_{i=1}^K d_i$ and the sample size is a product of sizes of all factors $N = \prod_{i=1}^K n_i$. If all the factors are one-dimensional Σ_{full} is a multidimensional grid. A subset of $\Sigma \subset \Sigma_{full}$ we will call an *incomplete factorial design of experiments*.

The DoE considered in this paper is a union of several anisotropic grids (here anisotropy means significantly different sizes of the grids), see Figure 1. It is an incomplete factorial DoE.

Such designs are rather complicated for approximation methods which don't use any knowledge about the data structure. Universal approximation algorithms (like Gaussian Process regression [5]) explicitly or implicitly assume that the DoE is rather uniform (there are no big holes in the DoE). Union of anisotropic grids is sufficiently a non-uniform DoE and this fact is the main reason to use special approximation technique. If we have big regions without training data then we have to control model smoothness in order to avoid oscillations and overshoots. Gaussian Process regression has some kind of smoothness control (not direct!) but it works only in the neighborhood of the training data (see Figure 2). To solve this problem we use a specific approximation algorithm described in the following section.

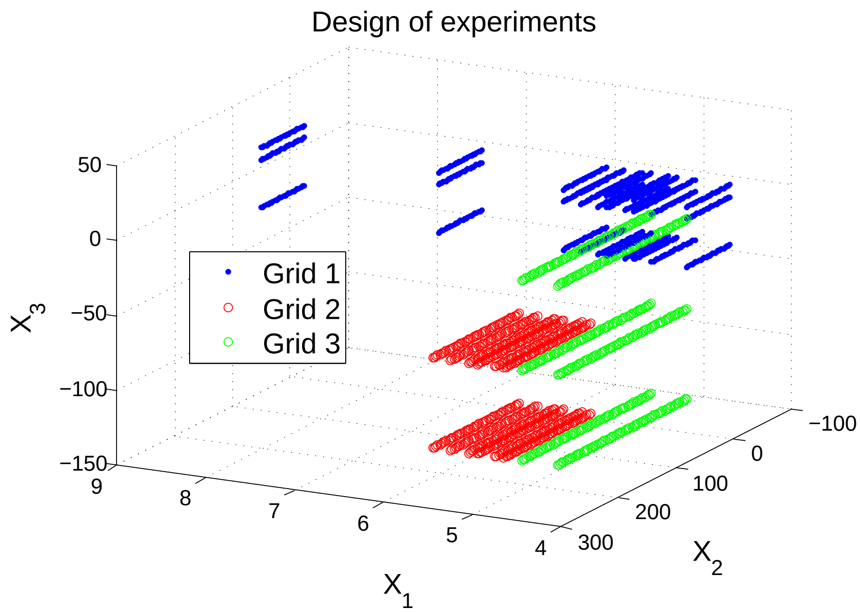


Figure 1: Design of experiments

3 TENSOR PRODUCT OF APPROXIMATIONS

The approximation $\hat{f}(\mathbf{x})$ will be modeled by the linear expansion in a dictionary of parametric functions. The algorithm consists of the following steps:

1. Choose class of functions F.
2. Choose penalty function P_λ.
3. Solve optimization problem (1).

The dictionary of functions is chosen in the following way. Let $\Delta_k = \{\psi^k_{j_k}\}_{j_k=1}^{p_k}$ be a dictionary of functions defined on D_k (for all $k = \overline{1,K}$). The dictionary of functions defined on $D_1 \times D_2 \times \cdots \times D_K$ is formed as tensor product of functions from Δ_k

$$\Delta_{tensor} = \{\psi^1_{j_1} \otimes \psi^2_{j_2} \otimes \cdots \otimes \psi^K_{j_K}, \{j_k = \overline{1,p_k}\}_{k=1}^K\}.$$

The class of functions F will be a linear expansion in a dictionary Δ_{tensor}. It means that the model $f(x) \in F$ can be written as $f(x) = \sum_{j_1,\ldots,j_K} \alpha_{j_1,j_2,\ldots,j_K} * \psi_{j_1}(\mathbf{x}^1) * \psi_{j_2}(\mathbf{x}^2) * \cdots * \psi_{j_K}(\mathbf{x}^K)$.

For the penalty function we will use variability of a function along some factor (or group of factors). For example, to penalize the variability along the first factor the following penalty function is used

$$P_\lambda(f) = \left\| \left(\frac{\partial^2 f}{(\partial \mathbf{x}^1)^2} \right) \right\|^2 = \sum_{i_1,\ldots,i_K} \left(\sum_{j_1,\ldots,j_K} \alpha_{j_1,j_2,\ldots,j_K} * \frac{\partial \psi^1_{j_1}}{\partial \mathbf{x}^1}(\mathbf{x}^1_{i_1}) * \psi^2_{j_2}(\mathbf{x}^2_{i_2}) * \ldots, \psi^K_{j_K}(\mathbf{x}^K_{i_K}) \right)^2.$$

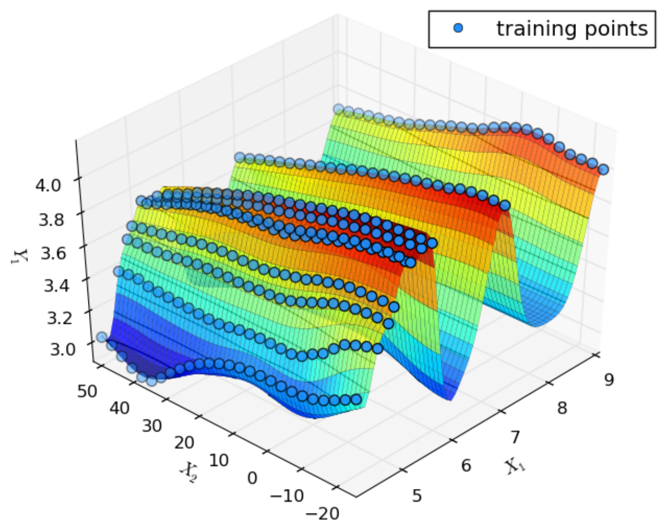


Figure 2: Example of degenerate Gaussian Process regression model.

Also note that the chosen penalty function is quadratic over decomposition coefficients $\mathrm{A} = \{\alpha_{j_1,j_2,\ldots j_K}, \{j_k = \overline{1, p_k}\}_{k=1}^K\}\}$. This penalty function can be generalized to control variability over several factors, for the details see [4].

Here we describe a simplified version of the last step since it is important for the data fusion problem. Due to the choice of penalty function the original optimization problem (1) can be reduced to minimization of the following function:

$$\widetilde{R}(\mathrm{A}) = (\widetilde{\mathrm{Y}} - \widetilde{\Psi}\mathrm{A})^{\mathrm{T}} W (\widetilde{\mathrm{Y}} - \widetilde{\Psi}\mathrm{A}) + \mathrm{A}^{\mathrm{T}}\Omega\mathrm{A}.$$

where

- $\widetilde{\mathrm{Y}}$ is an extended vector of training values and $\widetilde{\Psi}$ is an extended matrix of regressors (see details in [4]);
- Ω is a square penalty matrix (see details in [4]);
- W is a diagonal weighting matrix (sizes are $N * N$):

$$W_{i,i} = \begin{cases} 1, & \text{if } \mathbf{x}_i \in \Sigma \\ 0, & \text{if } \mathbf{x}_i \in \Sigma_{full} \backslash \Sigma. \end{cases}$$

Optimization problem $\min \widetilde{R}(\mathrm{A})$ can be solved in a very efficient way using special structure of the training set and tensor calculus [4].

4 DATA FUSION BASED ON TENSOR PRODUCT OF APPROXIMATIONS

In this section we will describe 3 different approaches to data fusion problem. All of them are based on tensor product of approximations on incomplete factorial DoE (iTA). We suppose that both HF points (Σ_h) and LF points (Σ_l) are incomplete factorial DoE.

4.1 Merged solution

Main idea of this approach is simple: we should merge CFD (low fidelity) and experimental (high fidelity) data into one sample using the following rule.

1. Merge sets of training points Σ_h and Σ_l into one set $\Sigma = \Sigma_h \cup \Sigma_l$;
2. Create corresponding set of the training values Y:
   - $y_i = y_i^l$ if $\mathbf{x}_i \in \Sigma_l$ and $\mathbf{x}_i \notin \Sigma_h$ (only low fidelity value is known), set $W_{i,i} = w_l$;
   - $y_i = y_i^h$ if $\mathbf{x}_i \notin \Sigma_l$ and $\mathbf{x}_i \in \Sigma_h$ (only low fidelity value is known), set $W_{i,i} = w_h$;
   - $y_i = w_l y_i^l + w_h y_i^h$ (where $w_h = 0.8$ and $w_l = 0.2$ are given confidence levels) if $\mathbf{x}_i \in \Sigma_l$ and $\mathbf{x}_i \in \Sigma_h$ (both high and low fidelity values are known), set $W_{i,i} = w_l + w_h$;
3. Construct iTA model using created sample and weighting matrix W.

This approach intends to fit LF model in regions where only LF values are given, HF model in regions with only HF values and weighted sum of LF and HF values in other regions.

4.2 Fused solution

This approach based on another idea. Let us approximate bias of the low fidelity function $g_l(x)$ with respect to the high fidelity $g_h(x)$. Then we can use this approximation to estimate values of the high fidelity function $g_h(x)$ in points where only low fidelity value is known.

1. Find set of points Σ_{both} for which $\mathbf{x}_i \in \Sigma_l$ and $\mathbf{x}_i \in \Sigma_h$ (both high and low fidelity values are known).
2. Calculate $Y_{diff} = \{y_h(\mathbf{x}) - y_l(\mathbf{x})\}$ for all points $\mathbf{x}$ from the set Σ_{both};
3. Construct model f_{diff} using $\{\Sigma_{both}, Y_{diff}\}$ as training sample;
4. Find set of points $\widetilde{\Sigma}_l = \{\mathbf{x}, \mathbf{x} \in \Sigma_l, \mathbf{x} \notin \Sigma_h\}$ (only low fidelity value is known);
5. Estimate high fidelity values in points from the set $\widetilde{\Sigma}_l$ using constructed model f_{diff}: $Y_{est} = f_{diff}(\widetilde{\Sigma}_l) + Y_l(\widetilde{\Sigma}_l)$.

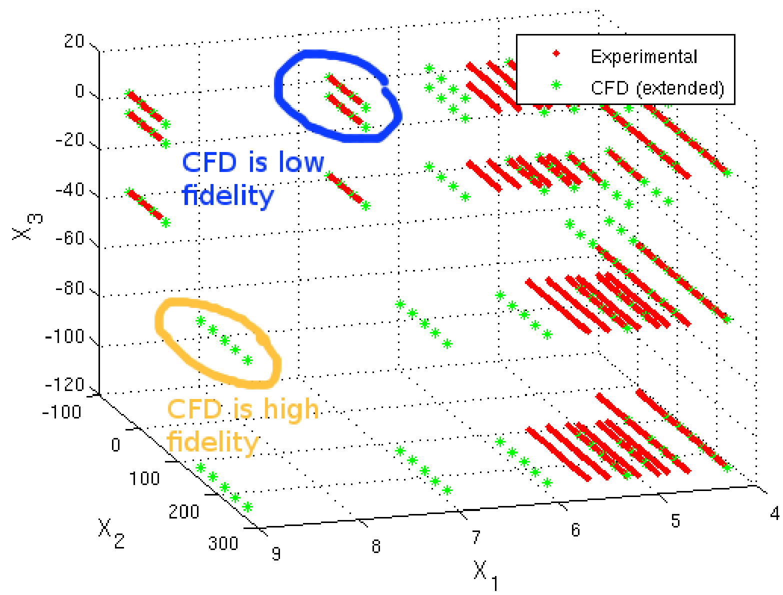


Figure 3: Local fidelity of CFD data.

6. Add $\{\widetilde{\Sigma}_l, Y_{est}\}$ to high fidelity sample as new data.

7. Construct Fused model using extended training sample $\{\widetilde{\Sigma}_l \cup \Sigma_h, Y_{est} \cup Y_h\}$.

4.3 Local fidelity solution

In previously described approaches we used the global fidelity, i.e. the confidence level doesn't depend on the location of the point and its neighborhood. However, if region contains lots of high fidelity experimental points then in this region CFD points should be treated as low fidelity (and can be even discarded), whereas in region without experimental points CFD points become high fidelity data. Therefore we come to the idea of *local fidelity*.

The difference between local and global approaches can be roughly explained using simplified data (see Figure 4). The global fidelity model try to catch the difference between HF and LF and use it to estimate HF values in [0.8, 1]. The local fidelity model doesn't trust HF data in [0.8, 1] (because there is no points nearby) and prefers to use LF data instead.

The concept of local fidelity can be implemented as follows. For each point $\mathbf{x} \in \Sigma_{both}$ we estimate the output value $\tilde{y}$ as

$$\tilde{y}(\mathbf{x}) = \frac{\tilde{w}_h(\mathbf{x})}{\tilde{w}_h(\mathbf{x}) + \tilde{w}_l(\mathbf{x})} g_h(\mathbf{x}) + \frac{\tilde{w}_l(\mathbf{x})}{\tilde{w}_h(\mathbf{x}) + \tilde{w}_l(\mathbf{x})} g_l(\mathbf{x}), \quad (3)$$

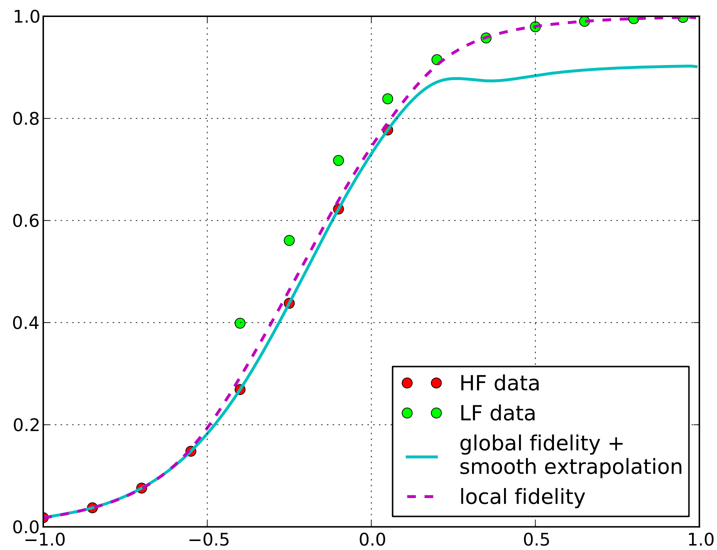


Figure 4: Local fidelity vs. Global fidelity.

where weights $\tilde{w}_h, \tilde{w}_l$ are calculated according to the idea of local fidelity:

$$\tilde{w}_k(\mathbf{x}) = \sum_{\mathbf{x}' \in \Sigma_k} \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{\sigma^2}\right), \quad k = h \text{ or } k = l.$$

If there are a lot of HF points in the neighborhood of point $\mathbf{x}$ then the value of $\tilde{w}_h(\mathbf{x})$ will be large, if there are few points in the neighborhood then $\tilde{w}_h(\mathbf{x})$ will be small. The same holds for $w_l(\mathbf{x})$. Therefore, in regions with large amount of HF points and low amount of LF points the value $\tilde{y}(\mathbf{x})$ will be close to $g_h(\mathbf{x})$ and vice versa.

Local fidelity data fusion algorithm:

1. Build $\hat{f}_h$ approximation of g_h using HF points.
2. Build $\hat{f}_l$ approximation of g_l using LF points.
3. For $\forall \mathbf{x} \in \Sigma_h \cup \Sigma_l$ calculate HF and LF weights $\tilde{w}_h$ and $\tilde{w}_l$.
4. Using equation 3 calculate Y_{est} for the set $\Sigma_h \cup \Sigma_l$.
5. Using $\{\Sigma_h \cup \Sigma_l, Y_{est}\}$ as a training set construct iTA model.

4.4 Comparison of proposed solutions

Now let us compare three proposed solutions on real data. Training sample has the following characteristics:

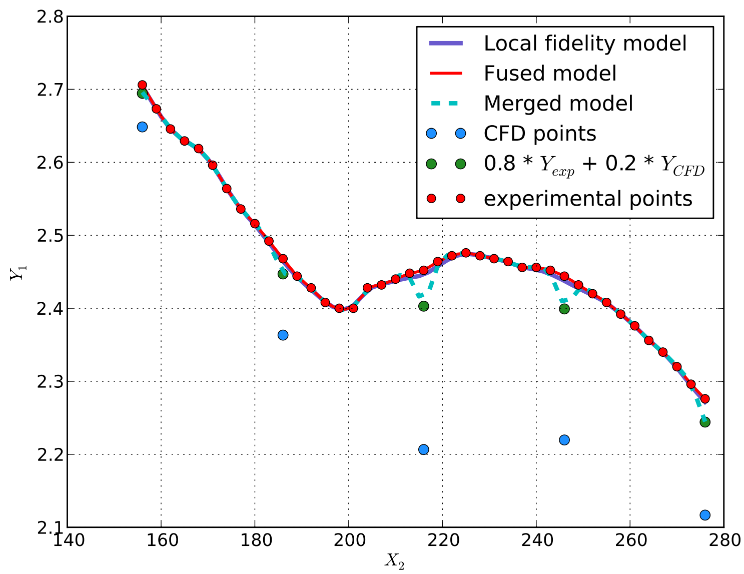


Figure 5: One-dimensional slice of constructed model.

- Input dimension: 3
- Output dimension: 3
- High fidelity sample size: 1846
- Low fidelity sample size: 180

Merged solution and Fused solution have some disadvantages. Figure 5 illustrates one-dimensional slice of approximations. As it can be seen Merged solution has small singularities at points where both high and low fidelity values are given.

The Fused solution is based on the difference between high fidelity model and low fidelity model. The difference model is built using points for which both CFD and experimental values are known. Such points dont cover the whole domain, so we have to extrapolate in other regions. Such procedure is rather inaccurate and introduces large uncertainties. Particularly, approximation can take values of the same magnitude as the HF and LF values. Such behavior is not reasonable from physical point of view. Figure 6 illustrates this problem.

Figure 7 depicts two-dimensional slices of obtained surrogate models. In this figure blue points denote HF points which were removed from the training sample in order to see the behavior of approximation in regions which have HF points but don't have LF points. Local fidelity surrogate model provides smooth approximation without overshoots and oscillations. One can see that merged and fused solutions can change their behavior significantly in such regions which leads to overshoots.

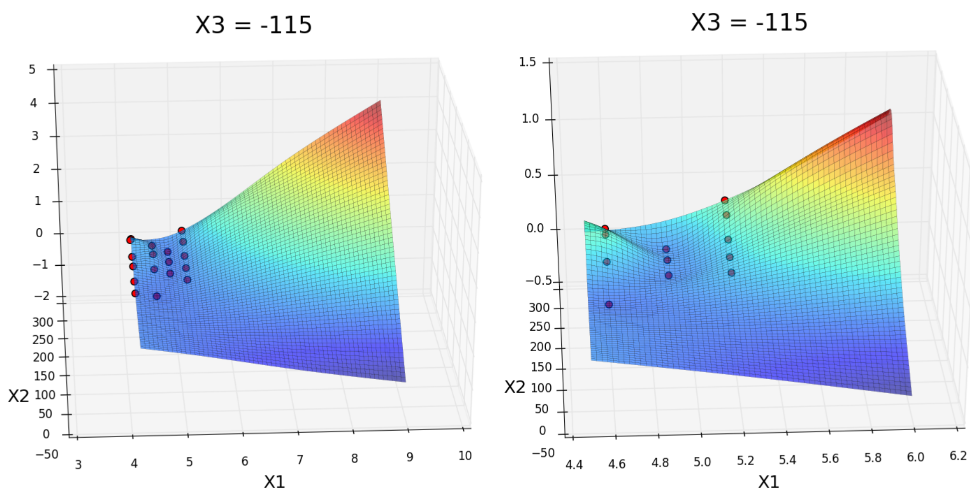
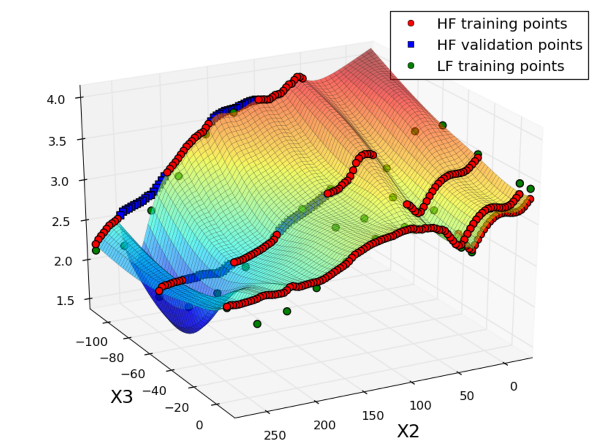


Figure 6: Problem with fused solution. Usual view (left) and enlarged view (right) of approximation of difference between HF and LF models. Approximation takes large values in region without training points.

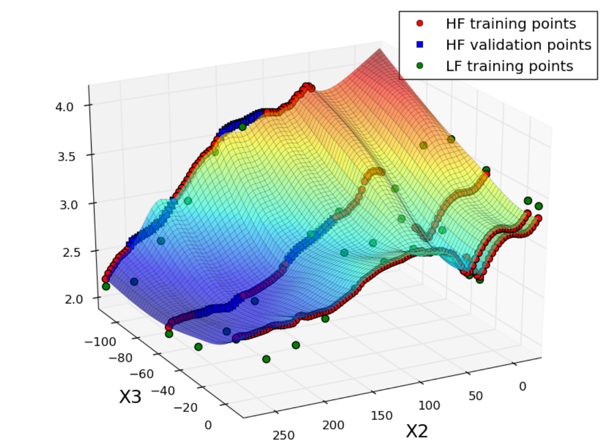
Thus, local fidelity solution doesn't have both mentioned disadvantages of merged and fused solutions and provides rather smooth and reasonable approximation. However, in some cases local fidelity solution can be less "physical" than fused solution. Local fidelity model interpolates the HF data in regions where only HF data is available while there is difference between LF and HF models and the both values $g_l(\mathbf{x})$ and $g_h(\mathbf{x})$ should be used and taken with corresponding confidence levels. So the fused solution can be more suited for such situations from physical point of view.

5 CONCLUSIONS

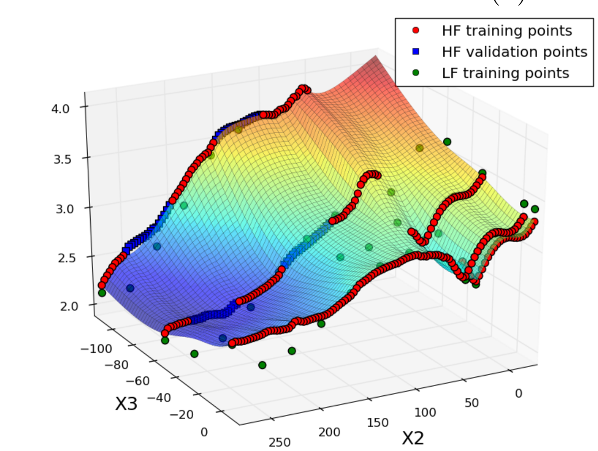
In this work we considered the data fusion problem where the design of experiments is a union of several anisotropic grids of HF and LF points. Several approaches (merged solution, fused solution, local fidelity solution) based on iTA approximation technique has been developed to solve this problem. The iTA technique takes into account structure of the data set and thus builds accurate approximations in a very efficient (in sense of computational complexity) way. The approach based on idea of local fidelity can provide good approximations but its behavior is not always physical. The fused approach is the most promising solution for now as it is able to follow physics of the process. The merged approach is mainly used for benchmarking purposes. The work on the both promising approaches (local fidelity solution and fused solution) is not finished. Further development should be put on removing discussed disadvantages.



(a) Merged solution



(b) Fused solution



(c) Local fidelity solution

Figure 7: 2D-slices of approximations

REFERENCES

[1] Forrester A., Sobester A., Keane A. *Engineering Design via Surrogate Modelling. A Practical Guide*. Wiley. 2008.

[2] de Boor C. *A Practical Guide to Splines, 2nd edition*. Springer-Verlag. Berlin 2001.

[3] Schölkopf B., Herbrich R., Smola A. *A generalized representer theorem Computational Learning Theory*. (2001) V. 2111. P. 416-426

[4] Belyaev M. *Approximation problem for factorized data*. Artificial intelligence and decision theory. (2013) V. 2013. P. 24-39

[5] Rasmussen, Carl E. and Williams, Christopher. *Gaussian Processes for Machine Learning*. MIT Press, 2006.