Creation of manufacturing cost estimation surrogate models for application in a cross-organizational design workflow

Ton van der Laan[1],
GKN Fokker, Papendrecht, 3351 LB, The Netherlands

Thierry Lefebvre[2], Nathalie Bartoli[3]
DTIS, ONERA, Université de Toulouse, 31000, Toulouse, France
Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, 31000, Toulouse, France

Abstract

Estimating manufacturing cost is a crucial step in evaluating the value of an aircraft design. However, this estimation is often performed only after the design and optimization phases are completed. In this paper, cost estimation is captured in surrogate models, making it available early in the design process. Using surrogate models for cost estimation also addresses intellectual property and complexity challenges. The use case presented in this paper focuses on a seaplane wing. The surrogate models are built using results from a Design of Experiments approach. To obtain these results, a knowledge-based engineering tool is employed to model the seaplane wing, and an open-source cost model is used to estimate the cost. The surrogate model itself is created using a Gaussian process model. Its sensitivity is analyzed, revealing that some design variables can be omitted while maintaining good model quality. Finally, the surrogate model is prepared for integration into a cross-organizational design workflow for the seaplane wing. The developed surrogate model proves useful in addressing intellectual property and complexity issues.

Nomenclature

CSV	Comma Separated Values
DOE	Design of Experiments
IP	Intellectual Property
LHS	Latin Hypercube Sampling
MDM	Multi-Disciplinary Modeller
SoS	System of Systems

I. Introduction

Estimating manufacturing cost is a crucial step in assessing the value of an aircraft design. However, this estimation is often carried out only after the design and optimization phases are completed. Consequently, manufacturing cost is excluded from the optimization process, leading to designs that may not be cost-efficient. The reason for not applying cost estimation models in the optimization is that they are often complex and contain intellectual property data. This paper explores how cost analysis can be integrated earlier in the design process through the use of surrogate models, circumventing the problems of complex and intellectual property issues. The design process envisaged will be cross organizational meaning that it will also have to deal with the challenges encountered by the characteristics of such a process.

In this paper, Section II addresses the context of manufacturing cost analysis and discusses common challenges. The use case for applying cost surrogate models is the optimization of a seaplane wing, which will be described in Section III. To gather the data required for creating the surrogate models, Design of Experiments (DOEs) must be defined; their content and definition strategies will be discussed in Section IV. Based on the DOE results, the surrogate models can be developed, and the process for doing so is explained in Section V. To understand the behavior of the created surrogate models and assess their applicability in optimization workflows, the sensitivities of the surrogate models are

[1] Engineering specialist, Knowledge Tools and Methods, ton.vanderlaan@fokker.com, AIAA Member
[2] Research Engineer, Information Processing and Systems Department, thierry.lefebvre@onera.fr, AIAA Member
[3] Senior researcher, Information Processing and Systems Department, nathalie.bartoli@onera.fr, AIAA MDO TC Member

examined in Section VI. Finally, Section VII describes the cross-organizational workflow in which the surrogate models will be applied.

II. Context of the manufacturing cost analysis

In the Colossus project [1] aircraft platforms are designed to operate in a System of Systems (SoS). The targeted approach aims to consider the SoS level and the Constituent level (i.e. the aircraft platform) with a specific focus on the coupling processes between both. In the design process at aircraft level, the manufacturing cost of major structural components will be included in the optimization process. The Colossus design process is cross-organizational meaning that multiple companies are involved. Enabling manufacturing cost estimation in such a design process encounters several issues:

1. **Data transfer between companies and computer networks.** When transferring data between companies and/or countries it has to go through firewalls and other obstructions that often result in issues that hamper a smooth design process. Furthermore, export compliance rules can also apply.
2. **Protection of Intellectual Property (IP).** Manufacturing cost models often contain IP that companies do not like to disclose to customers or competitors.
3. **Detailed manufacturing and structural information.** Detailed information is often needed for manufacturing cost models. It takes time and complex models to generate this information**.**

In this paper an approach using surrogate models of manufacturing cost is proposed to counter these issues and enable detailed manufacturing cost estimation in a cross-organizational design process. The surrogate models used can be transferred to companies or networks where the design process is executed eliminating the transfer of data between companies in the design process itself. This counters the first issue.
The surrogate models will also ensure that the details of the cost models used will not be visible to other companies or parties in the design process protecting the IP of the manufacturing cost model itself. That said the surrogate models themselves can hold sensitive cost information and therefore become part of a company's IP. In this case the use of surrogate models addresses part of the second issue.
Finally the surrogate models will only have a limited set of input design variables to calculate the manufacturing cost. Therefore no detailed manufacturing and structural information is needed. This addresses the third issue of the list.

In this paper the manufacturing cost model used for the creation of the surrogate models will be a previously developed open source cost model [4]. This cost model produces cost figures at mono-part level. This level is too detailed to be incorporated in a design workflow however, by rolling the result up into a surrogate model application of the cost model is enabled.
The development of the surrogate models will be used as a proof of concept. As a result, not all structural parts of the aircraft design will be covered. In this case the wing design of a seaplane aircraft [2] will be the system of interest covered by the surrogate models. The cost estimation does not incorporate assembly costs, because the aircraft component models discussed in the next paragraph do not have all the data needed for such an estimation.

III. How to create the data required for the manufacturing cost surrogate models

To create the manufacturing cost surrogate model, data has to be created on which to base the surrogate model definition. This basically means determining the inputs and outputs of the surrogate model. To make the cost surrogate models useful in the design and optimization process of the seaplane wing design, the variables used and varied in the design process should be input design variables for the surrogate model. The variables used in the aircraft wing design are:

1. Rear spar position
2. Wing aspect ration
3. Outer section taper ratio
4. Wing weight derived from material application

To generate the cost data for the surrogate models the Multi-Disciplinary Modeller (MDM) [3] of GKN Fokker will be used. This is a Python based tool capable of generating detailed wing models and applying analyses to these models. The MDM models the left-hand side half of the seaplane wing.
To enable the MDM to generate the data required for the surrogate models a wrapper is developed which translates the variables used in the wing design into variables that can be understood by the MDM. To generate the wing other constants that do not vary with the modelled but are used for all experiments. Most important of these constants are:

A. Wing area = 40m2
B. Airfoil shape = NACA0012
C. Airfoil thickness/chord ratio = 0.18
D. Root chord = 2120mm
E. Position of kink = 0.35
F. Front spar position = 0.3
G. Target rib spacing = 500mm

The airfoil shape and thickness/chord ratios are not taken directly from the wing design but are chosen to result in a representative height for spars and ribs, which is most important for the cost estimation process. A schematic top view of a resulting wing half can be seen in Figure 1.

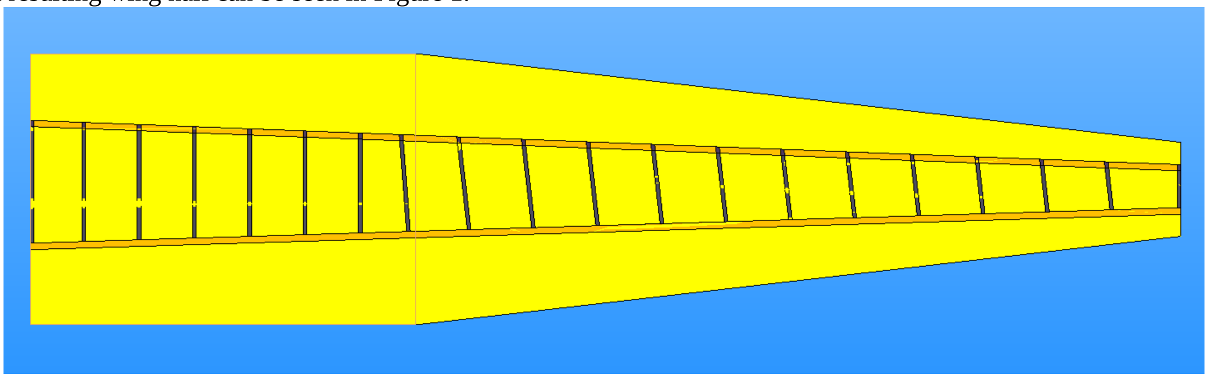


Figure 1: Schematic top view overview of a left side wing box generated by the MDM.

The variables used in the wing design process are handled by the MDM in the following way:

1. Rear spar position, is a direct variable for the MDM.
2. Wing aspect ratio, is used to calculate the wing span using the wing area.
3. Outer section taper ratio is used to determine leading edge sweep angle and the tip chord for the outer section.
4. The wing weight is not a direct input for the MDM. The wing weight is determined by other input variables such as the material definition of all parts in the MDM. As a result, a way will have to determined how to properly create a surrogate model that has weight as an input variable when weight is not an input for the MDM. Furthermore the weight range covered will have to match the output of the other tools in the design framework.

An example of a wing model with different material properties applied can be seen in Figure 2.

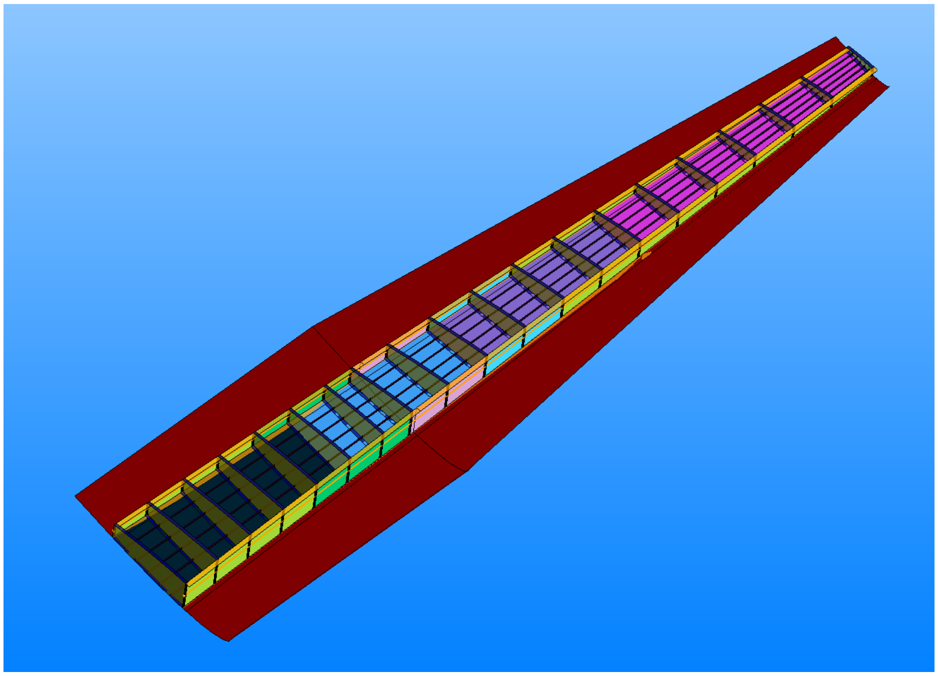


Figure 2: Wing definition in MDM including material applications. Different colors define different material thickness or stack. The upper skin is hidden.

The open source cost model that is used to calculate the part manufacturing cost requires certain data. This data is a combination of geometric data and manufacturing method data. With the MDM both geometric and manufacturing process data are determined. The geometric data is extracted from the geometrical model and includes data about the size, area and mass of each manufactured part. The manufacturing process data consists of manufacturing data per part and generic manufacturing data. The data per part is the manufacturing process used for each part and the generic parts contains data about the manufacturing environment and labor and machine rates.

IV. Determining the DOE content

Designs of Experiments (DOEs) are used to generate the data required for cost surrogate models. For these DOEs the wrapper around the MDM is used. Currently two wrappers are used that enable two different types of DoEs.

In the first wrapper the possible options for the design variables can be specified as a list and a full factorial DoE can be executed. Depending on the number of design variables and on the number of options per design variable this can result in a very large amount of experiments to be executed. This large number of experiments can result in excessive calculation times. To reduce the process time a characteristic of the MDM is used. This is characteristic is lazy evaluation. MDM is written using the commercial Python library Parapy [5]. This library offers the use of CAD elements and also enables lazy evaluation. Lazy evaluation is a programming or modeling technique in which the evaluation of an expression, rule, or computation is postponed until its value is required by another part of the system. This means that a change in experiment variables do not necessarily result to a complete re-evaluation. In case of the full factorial wrapper this is used for evaluating changes in manufacturing method variables. These changes do not change the geometrical model, which is the most computationally expensive to evaluate. By only making a new instance of the MDM for geometrical variable sets and changing the manufacturing variables within that instance valuable computational time can be saved.

The second wrapper for executing DOEs with the MDM uses a list of experiments defined in an external file. This file is of the Comma Separated Values (CSV) format. The variable values are read from the file and experiments are being executed until the end of the file is reached. Unfortunately it is now known what variables change for each experiments and therefore it is unclear if the geometry changes between experiments. As such lazy evaluation does not have a big effect and for each experiment and new instantiation of the MDM is made.

For the manufacturing surrogate model both wrappers have been used. The first one was used to make a general investigation of the effect of variables and the second one to execute the specific DOEs required for the surrogate model generation.

For the seaplane two different material options will be considered: an aluminium metal option and a composite option. Aluminium and composite designs have different masses and mass ranges also their behaviour can be different. Therefore in a DOE experiment only one material type is considered, either metal or composite material. This also means that hybrid designs where part of the structure is metal and another part is composite will not be covered by the DOEs or surrogate models.

For the first DOE a full factorial search was conducted to explore the design space. In this full factorial the numeric values and material thickness values are covered by three levels, because it is not known if the behaviour of these variables is linear or not. Manufacturing methods are all checked individually. Table 1 recaps all the manufacturing methods per manufactured part considered for the DOE. As can be seen the manufacturing methods are different for metal and for composite components.

Table 1: Manufacturing methods considered in the DoE.

Metal			
Skin	**Spar**	**Rib**	**Stringer**
Stretch_Forming Skin_machining	Speed_Machining RubberForming Machining	RubberForming Machining	RubberForming Machining FoldingCutting SpeedMachining
Composite			
Skin	**Spar**	**Rib**	**Stringer**
Hand-Layup ATP RTM Vacinf	Hand-Layup ATP RTM Vacinf	Hand-Layup ATP RTM Vacinf Double_Diaphragm_Forming	Hand-Layup Double_Diaphragm_Forming

For the material thickness values sets are created. This means that within a set all the part groups have a certain material definition. There is no mixing of material settings between the sets so in an experiment the material options used for the parts will be light, medium or heavy as can be seen in Table 2.
Two different implementations for thickness variations are implemented for metal and composite materials. For composite the material stacks are varied, while for metals the thickness itself is directly varied. Varying the material properties in the parts does not mean they are constant for a part, for skins and spars the materials get thinner near the tip of the wing as can be seen in Figure 2.

Table 2: Material options considered for the different part sets. This will result in different sets of composite stacks for composite parts and different material thicknesses for metal parts.

Material option sets	Material thickness or stack type per manufactured part group			
	Skin	**Spar**	**Rib**	**Stringer**
Light	Light	Light	Light	Very light
Medium	Medium	Medium	Heavy	Medium
Heavy	Heavy	Heavy	Heavy	Heavy

For the three numeric variables that are considered in the DOE three values are chosen: the upper and lower boundaries used in the design process for the seaplane and an intermediate value. This should ensure that any nonlinear influence of these variables on the manufacturing cost is identified. The values used for the numeric variables can be seen in Table 3.

Table 3: Design variable options for numeric variables.

Numeric DoE variables	
Variable name	**Variable options**
Rear_spar_position	0.6, 0.675, 0.75
Aspect_ratio	8, 10.25, 12.5
Outer_taper_ratio	0.35, 0.675, 1

The full factorial DoE results in approximately 10000 experiments for each material option, metal or composite, each experiment takes a few seconds, but this still results in long calculation times. As explained in the previous section, calculation times are reduced by using lazy evaluation when evaluating the different manufacturing options. This principle could also be applied when varying the other variables however in this case this has not been implemented.

When running the DOE for surrogate model generation there is a complicating factor in that the surrogate models are not generated by the same company then the company responsible for running the DOE. This is because the MDM is a proprietary tool owned by GKN Fokker and the surrogate models are generated by ONERA, therefore ONERA defines the content of the DOE required for building the surrogate model. To enable the external definition of DOEs the second DOE wrapper based on CSV has been used. The experiments to be run were provided by ONERA and were run by GKN Fokker. Once the data was generated using the MDM it was provided back in CSV format to ONERA for the construction of the surrogate models.
An example of experiment definitions can be seen in Table 4. In total 96 composite and 104 metal experiment points were defined. Taking into account three different material definitions this results in (96+104)*3 = 600 experiments in total.

Table 4: Example CSV definition of 4 composite experiments.

CSV DOE definition data						
skin_manufacturing method	spar_manufacturing method	rib_manufacturing method	stringer_manufacturing method	rear_spar position	aspect_ratio	taper_ratio outer_wing
Hand_Layup	Hand_Layup	Double_Diaphragm_Forming	Hand_Layup	0.679864	11.04080064	0.560960003
Hand_Layup	Hand_Layup	VacInf	Double_Diaphragm_Forming	0.603139	8.407694648	0.491921692
Hand_Layup	RTM	VacInf	Double_Diaphragm_Forming	0.701823	8.953861949	0.603562815
VacInf	ATP	Hand_Layup	Double_Diaphragm_Forming	0.715058	10.48273905	0.998462156

The data in the CSV definition is not sufficient for running the experiments, it is lacking the material thickness option, so the complete DOE actually consists of running a specified experiment for all the thickness option represented in Table 2, of course taking into account if the experiment is for a metal or composite construction. The resulting DOE results are sent to ONERA to enable surrogate model generation. These results are processed more into a CSV file which is the basis for all subsequent surrogate modeling work. An example of this processed data can be seen in Table 5.

Table 5: Example CSV definition of 4 composite experiments DOE results.

CSV DOE results data											
ID	material	stacking	skin_manufacturing method	spar_manufacturing method	rib_manufacturing method	stringer_manufacturing method	rear_spar position	aspect_ratio	taper_ratio outer_wing	total_mass	total_cost
0	Composite	Heavy	Hand_Layup	Hand_Layup	Double_Diaphragm_Forming	Hand_Layup	0.679864	11.04080064	0.560960003	142.4721	29017.76
1	Composite	Heavy	Hand_Layup	Hand_Layup	VacInf	Double_Diaphragm_Forming	0.603139	8.407694648	0.491921692	111.7418	25878.84
2	Composite	Heavy	Hand_Layup	RTM	VacInf	Double_Diaphragm_Forming	0.701823	8.953861949	0.603562815	140.3977	29875.41
3	Composite	Heavy	VacInf	ATP	Hand_Layup	Double_Diaphragm_Forming	0.715058	10.48273905	0.998462156	177.5303	55133.9

The surrogate models are created to analyze cost in a cross organizational workflow, this is described in Section VII. To ensure the resulting surrogate models are applicable in this workflow they have to cover the complete design space or possible outcomes of other tools in the workflow. An analysis of the results of these tools showed that the 600 experiment DOE did not cover the complete mass range from the other tools in workflow. As can be seen from the workflow tool and original DOE rows in Table 6. This means that if a surrogate model was only based on these 600 points it would need to be extrapolated in order to be functional in the workflow. In order to avoid this extrapolation a new lighter material variant was added to the experiments. This results in an additional 200 experiments. With these experiments added the DOE dataset now encompasses the complete mass range of the workflow tools ensuring that the resulting surrogate models only uses interpolation and no extrapolation, as can be seen in Table 6.

Table 6: Overview of minimum and maximum total wing masses for workflow tool and DOE results.

Minimum and maximum total mass of wing		
	Minimum Mass	**Maximum mass**
Workflow tool metal	203.58 kg	401.76 kg (when removing 782.13 kg outlier)
DOE results metal original	236.76 kg	554.59 kg
DOE results metal including very light	183.63 kg	554.59 kg
Workflow tool composite	144.74 kg	276.37 kg
DOE results composite original	162.55 kg	356.62 kg
DOE results composite including very light	141.38 kg	356.62 kg

For the creation of the surrogate models the full factorial DOE dataset is not necessary and would probably even be too large. Therefore, based on the full factorial dataset a new DOE is defined that will form the bases for the surrogate models that are generated. In this case a Latin Hypercube Sampling (LHS) has been used using the formalism defined in the SMT toolbox [7] and particularly in the *smt-design-space-ext* package [8]. A total of 14 variables are listed in Table 7, four of which are active only when metal is selected as the material, and four others when composite is selected. A DOE of 600 + 200 = 800 points is built in order to cover the design space with 416 metal points and 384 composite points.

Table 7: Design variables for the hierarchical DOE. The mass figures account for half a wing.

Variables	Nature	Number of levels	comment
1. Material choice	Categorical	2 levels	Metal/Composite
2. Rear spar position	Float	See Table 3	
3. Aspect ratio	Float		
4. Outer taper ratio	Float		
5. Mass for Metal, 6. Composite	Float	[91,278], [70,179]	
7. Skin manufacturing for Metal, 8. Composite	Categorical	2 levels, 4 levels	See Table 1 for the manufacturing options Metal en metallic methods are considered different variables
9. Spar manufacturing for Metal, 10. Composite	Categorical	3 levels, 4 levels	
11. Rib manufacturing for Metal, 12. Composite	Categorical	2 levels, 5 levels	
13. Stringer manufacturing for Metal, 14. Composite	Categorical	4 levels, 2 levels	

Some details can be found in the online documentation for the hierarchical LHS[4].

V. Creating the cost surrogate models using the DOE results

As the objective is to create a unique surrogate model for metal & composite material, a Gaussian process model with a specific kernel has been chosen. The Gaussian process model [6] (also denoted as Kriging) provides a probabilistic model with both a prediction value and an estimation of the associated uncertainty. This uncertainty information is very useful to identify areas where the prediction is not accurate enough in order to add points in an active learning process. The manufacturing process depends on both the material and the part (skin/spar/rib/stringer) so a hierarchical structure has to be introduced.

Using the 800 DOE points, a split of 10% between training and validation points is done: 720 points are used for the training and 80 points for validation. Different Gaussian process hyperparameters have been tested in order to reduce the error on the validation set, in particular for both the hierarchical kernel and the categorical one. Depending on the

[4] https://smt.readthedocs.io/en/stable/_src_docs/applications/Mixed_Hier_usage.html#id6

choice, the number of parameters to evaluate via the maximum likelihood estimation could increase as could the CPU time. Table 8 lists the choice made to get the best result with the DOE data available**.**

Table 8: Choice of the hyperparameters for the Gaussian process model[5].

Hyperparameter	Choice
Hierarchical kernel	Algebraic distance
Categorical kernel	Gower distance
Continuous kernel	Power exponential correlation (with power=1.9)
Regression term	constant

With the settings a surrogate model can be created using the DOE results the resulting model has a relative error of 0.66% on the validation set. In the DOE results the metallic and composite results are combined. The model prediction with the associated confidence interval can be seen in Figure 3. As expected, the training points (blue dots) lie on the diagonal (an accurate prediction, since Gaussian processes are interpolation models), and the test points (orange crosses) are not far from it.

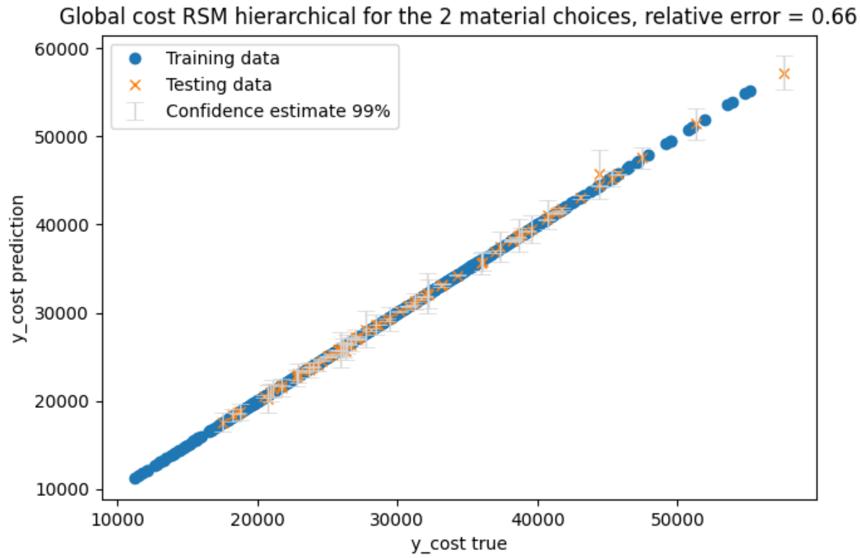


Figure 3: Cost prediction versus true cost value for the training and validation points. A confidence interval of 99% is plotted for each test point

In order to try and improve the surrogate model the composite and metal DOE points were also used as separate sets to create a specialized metal and a specialized composite surrogate models. In this case the metal model has a relative error of 0.04% and the composite a relative error of 1.0%. This shows that the composite DOE results are more difficult to capture in a surrogate models. The small improvement in surrogate model quality and the increase complexity of having to manage two different surrogate models does not warrant the split for use in optimization workflows. However it does show that adjusting the inputs and definition of the surrogate models can show some interesting behavior. This will be investigated further in the next section.

[5] https://smt.readthedocs.io/en/stable/_src_docs/applications/Mixed_Hier_surr.html

VI. Cost surrogate model sensitivities

The cost surrogate model used in the wing design workflow uses all variables available in the workflow plus a definition of the manufacturing methods to be used for all manufactured parts. All the input variables and their valid ranges can be seen in Table 9. The surrogate model might not be sensitive to all the input variables in the model. In this section the sensitivities of the input variables will be investigated.

Table 9: Complete surrogate model input variable overview.

Surrogate model variables	
Variable	**Possible values**
Material type	Metal/composite
Rear spar position	Between 0.6 and 0.75
Wing aspect ratio	Between 8 and 12.5
Outer section taper ratio	Between 0.35 and 1
Wing mass	Metal between 183.63 and 554.59 Composite between 141.38kg – 356.62kg
Skin manufacturing method	Methods from Table 1
Spar manufacturing method	Methods from Table 1
Rib manufacturing method	Methods from Table 1
Stringer manufacturing method	Methods from Table 1

To test the sensitivities some variables can be excluded from the surrogate model generation. This will probably result in models that are worst then the original however the interest lies in how much worst and also if this degradation of quality is acceptable.

A dataset of 20% of the original data points will be used to test the quality of the model, meaning they are not part of the training set. The surrogate model is a Gaussian process model so all training points will have an exact match in the surrogate model. In the tables below (

Table 10, Table 11, Table 12) different active variable sets, meaning variables are included in the building of the surrogate model, are executed and commented. In the tables the training points are represented by dots and by definition lie on the diagonal of the diagram. The test points are crosses and indicate the quality of the surrogate model.

Table 10 Base line surrogate model and surrogate models omitting geometrical variables

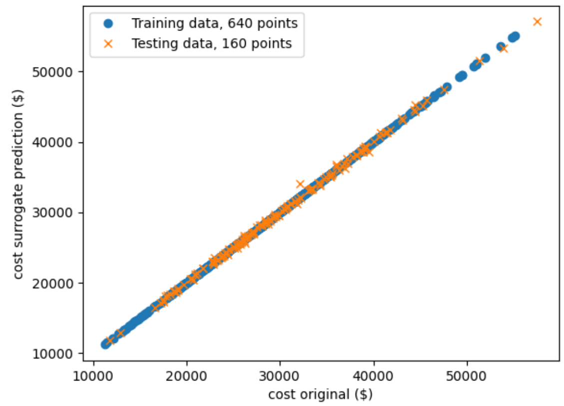
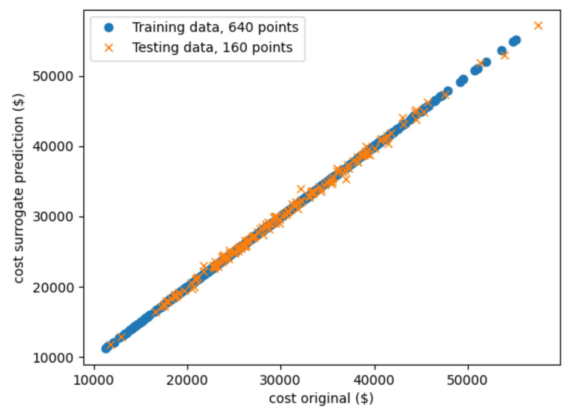
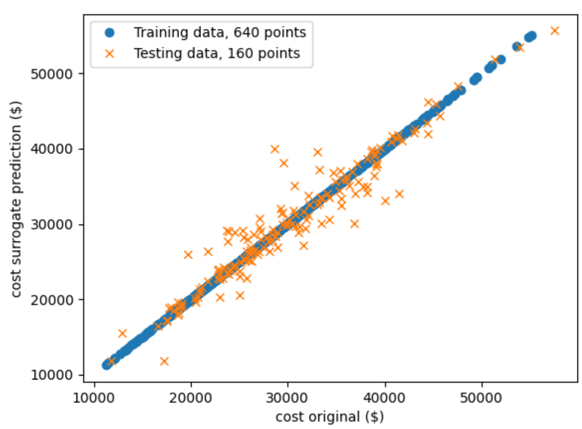
Set name	Description	Model graph
1. Full original model	The training and testing points lie on the diagonal which means that the predictions of the test points closely match the original data	
2. Omitting taper and wing aspect ratio variables	The training points and test points lie very much on the diagonal. There is not much difference with respect to the full original model. Hereby we can conclude that the taper and wing aspect ratio variables do not have a big influence on the surrogate model.	
3. Omitting all geometric variables	The test data lies of the diagonal especially in the middle part of the cost range. This indicates that the model is not very accurate there. So this model is probably not useful.	

Table 11: Surrogate models omitting manufacturing methods variables

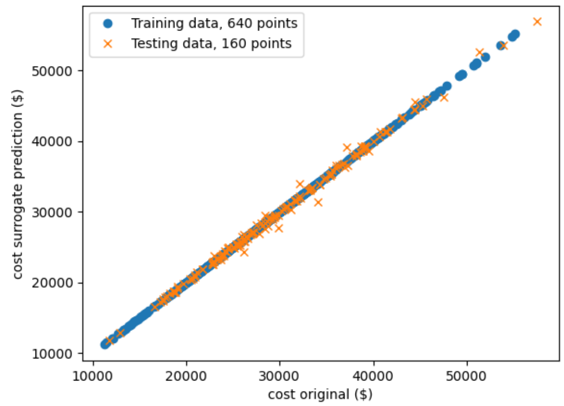
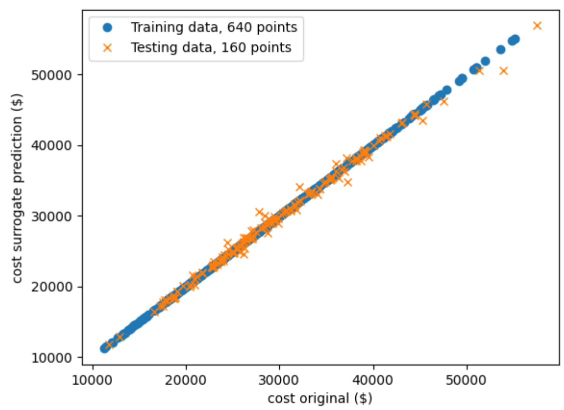
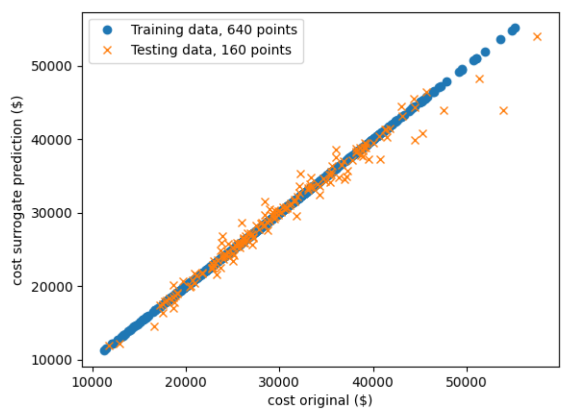
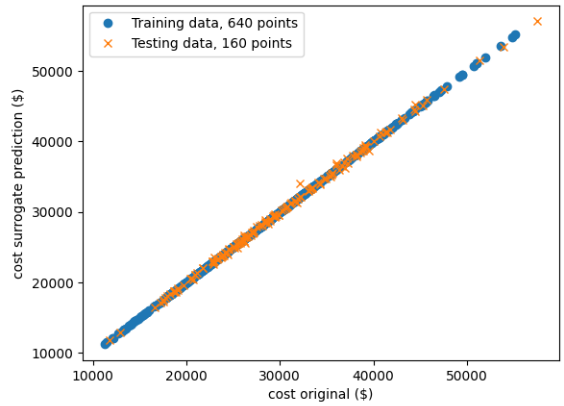
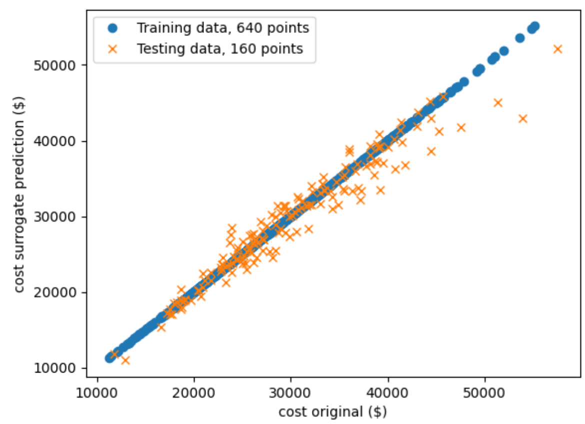
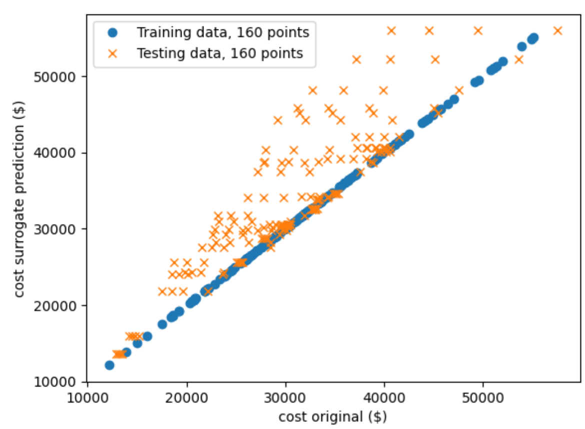
Set name	Description	Model graph
4. Rib and stringer manufacturing methods omitted	The testing points lie closely to the diagonal which means that the predictions of the test points still provide a good surrogate model. Indicating that the rib and stringer manufacturing methods do not have a big influence on the quality of the surrogate model. This could be due to the low mass of the ribs and stringers compared to the wing box total mass.	
5. Spar, Rib and stringer manufacturing methods omitted	By adding the spar manufacturing methods to the omitted variables we still see all test points close to the diagonal, so the spar manufacturing method also does not seem to have a big influence on the surrogate model quality. Again its mass might be too low to have a big influence.	
6. Omitting all manufacturing method variables	Even when omitting all manufacturing method variables the test points still mostly lie close to the diagonal. For the higher cost estimate the test point are underestimating significantly.	

Table 12: Surrogate models omitting material type and mass variables

<table>
<tr><th>Set name</th><th>Description</th><th>Model graph</th></tr>
<tr><td>7. Material type variable omitted</td><td>The training and testing points lie on the diagonal. The model resembles the original model. It seems that the material type variable has no influence on the quality of the surrogate model. However this could be because the different material types are also contained in the part manufacturing methods.</td><td>


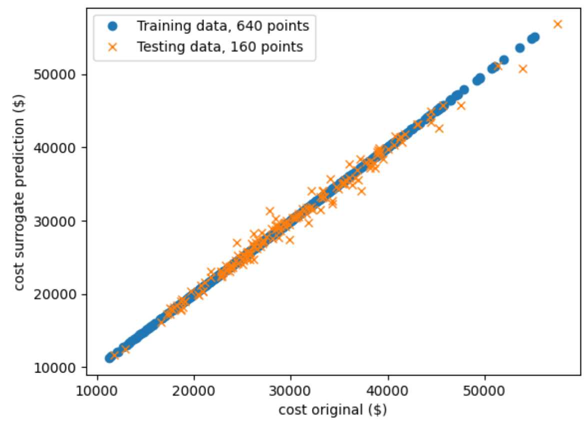
</td></tr>
<tr><td>8. Material type and manufacturing method variables omitted</td><td>The testing points lie of the diagonal and are more of the diagonal then in 6, indicating that the material type variable and manufacturing methods variables interact.</td><td>


</td></tr>
<tr><td>9. Omitting the mass variable</td><td>The testing points lie of the diagonal. Also the number of training points is reduced because of double entries in the training set which are removed. Interestingly the surrogate model seems to always overestimate the cost of the original.</td><td>


</td></tr>
</table>

Based on the results of the different surrogate models is seems reasonable to expected that a surrogate model where only the material type, skin manufacturing method and rear spar position are used to make the surrogate model would give acceptable results. The results can be seen in Table 13.

Table 13 Surrogate model only based on material type, skin manufacturing method and rear spar position

Set name	Description	Model graph
10. Only Material type, skin manufacturing method and rear spar position variables active	The testing points lie relatively close to the diagonal and the absolute different does not vary a lot across the cost range. This seems to be an acceptable surrogate model.	

The surrogate model seems to be acceptable however to be sure more investigations will be required, as this is only an initial indication. This does also coincide with engineering logic. Material type determines material cost and has a big influence on the manufacturing methods that can be chosen for the different manufactured parts. The skins are the biggest and heaviest parts of the wing box and therefore its manufacturing process will have the most significant effect of all manufacturing methods on the manufacturing cost of the total wing box. Finally the mass of the wing box directly related to how much material is used. The amount of material affects the manufacturing cost though the material cost directly and also indirectly to the time it takes to manufacturing a part.

The analysis shown in this section is only an experiment with switching variables on and off for surrogate model generation. More in debt analysis will be required to truly understand the sensitivities in the data. However what is shows that is always valuable to analyse the data when making surrogate models and to try and relate surrogate model behaviour to engineering physics.

The Python software code and data used for the creation of the surrogate models shown in the diagrams in this section can be found here: https://doi.org/10.5281/zenodo.17799280.

VII. Application of the cost surrogate model in a cross organizational workflow

The cost surrogate models will be part of a larger cross organizational workflow that optimizes a seaplane wing design. The role of the cost surrogate models in this workflow is to estimate the manufacturing cost of the wing box mono-parts. Even though, the wing design itself is not influenced by the cost surrogate models, the aim is to estimate the manufacturing cost of the optimized wing. The workflow itself consists of multiple tools and/or surrogate models. The main element of the workflow that has interaction with the cost surrogate models is the element performing the structural optimization of the wing box. In the workflow itself this is either the Proteus tool [9] or a surrogate model representing the results from Proteus. An overview of how the manufacturing cost surrogate models relates to the structural sizing process in the workflow can be seen in Figure 4.

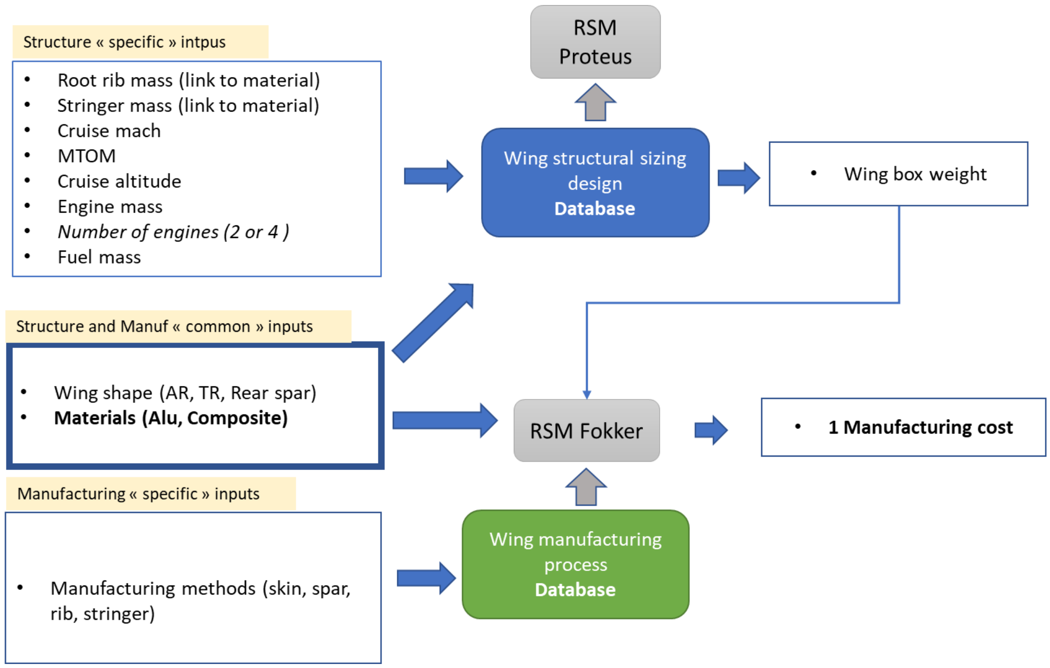


Figure 4: Position of the manufacturing cost surrogate models in the workflow.

There are several advantages of using the cost surrogate model in the workflow these are:

- **Computational speed**, reduce the computational effort to run the workflow.
- **Intellectual property (IP) protection**, the surrogate model encompasses the cost models used to create the surrogate model and therefore protects the IP within those models.
- **Simplify data flows in the workflow.** The surrogate models can be stored in binary files and run on any computer that has Python and SMT installed. Therefore the tools do not have to run in a designated place and is not affected by company infrastructure and firewall limitations.

To ensure the cost surrogate model works properly in the workflow it has to cover the complete workspace as covered by the other tools in the workflow. This has been achieved by taking into account all variables varied in the workflow as shown in Table 3 and ensure the complete mass range coming from the structural optimization is covered.

The cost surrogate model in the workflow is limited to covering the mono part cost of the wing torsion box elements. Therefore it cannot calculate the complete manufacturing cost of the wing itself. The parts covered are skin panels, spars, ribs and stringers and can be seen in Figure 5 and Figure 6.

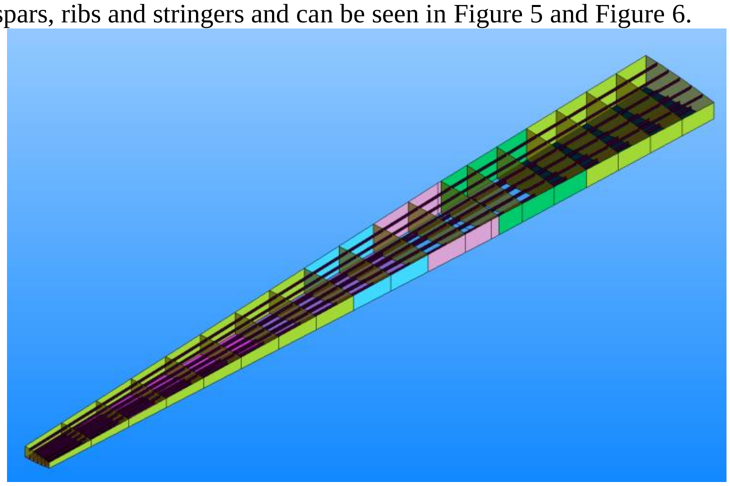


Figure 5: Wing box covered by the cost surrogate model

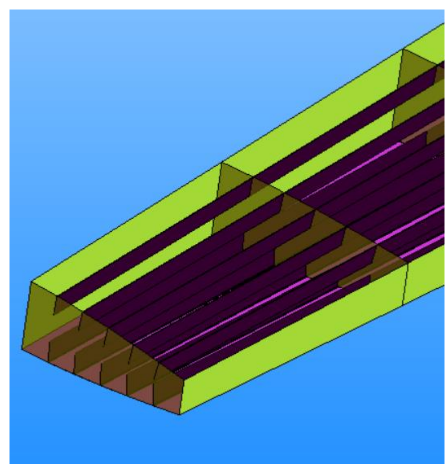


Figure 6: Wing box detail

Unfortunately the complete wing optimization workflow with the cost estimation surrogate model is not up and running yet so results are not yet available. However the applicability of the surrogate models is assured because the in- and outputs of the surrogate models match the ones used by other tools in the workflow.

VIII. Conclusions and future work

In this paper, it was demonstrated that a cost estimation model for wing box parts can be captured using a surrogate model. The steps required to generate such models were explained. To produce the data needed for the surrogate model, detailed geometric models and an open-source cost model were used. The dataset was selected to ensure that reliable surrogate models could be generated, which were then used in a design workflow to determine the manufacturing cost of a seaplane wing box. Finally, the sensitivities of the surrogate models were analyzed, and it was shown that applying engineering judgement when analyzing the surrogate model dataset can be beneficial, as it may simplify the dataset required for building the surrogate model.

Future work will include the finalization of the cross-organizational workflow in which the surrogate model will operate. Furthermore, the applicability of the surrogate model will be further tested by entering structurally optimized design concepts into the Multi-Disciplinary Modeler (MDM) and comparing the calculated cost in the MDM with results from the surrogate model. Finally, more advanced tools will be used to analyze the correlation within the surrogate model dataset to better understand the value and applicability of the model created.

Acknowledgments

The research presented in this paper has been performed in the framework of the COLOSSUS project (Collaborative System of Systems Exploration of Aviation Products, Services and Business Models) and has received funding from the European Union Horizon Europe program under grant agreement No. 101097120.

Data availability and replication of results

The python software code and data used for the creation of surrogate models discussed section VI can be found here: https://doi.org/10.5281/zenodo.17799280.

IX. References

[1] Prakasha, P.S., Naeem, N., Amadori, K., Donelli, G., Akbari J., Nicolosi, F., Knöös Franzén, L., Ruocco, M., Lefebvre, T., and Nagel, B., "COLOSSUS EU Project – Collaborative SoS Exploration of Aviation Products, Services and Business Models: Overview and Approach", ICAS2024-1138.

[2] Tuccilo, M., Ruocco M., "Aircraft Design Capabilities for a System-of-Systems Approach (eVTOL and Seaplane Design)", 14th EASN International Conference on "Innovation in Aviation & Space towards sustainability today & tomorrow", Thessaloniki, Greece, 8–11 October 2024.

[3] Berg, T. van den, Laan, A.H. van der, "A multidisciplinary modeling system for structural design applied to aircraft moveables", in AIAA AVIATION 2021 FORUM, VIRTUAL EVENT, 2021.

[4] Laan, A.H. van der, Johnman, A., Puffelen, T. van, Nolet, S., Manen B. van, Daugulis, E., Berg, T. van den, "An open source part cost estimation tool for MDO purposes" in AIAA AVIATION 2021 FORUM, VIRTUAL EVENT, 2021.

[5] https://www.parapy.nl/, cited 27-10-25

[6] Williams, C. K., Rasmussen, C. E. (2006). Gaussian processes for machine learning (Vol. 2, No. 3, p. 4). Cambridge, MA: MIT press.

[7] Saves, P., Lafage, R., Bartoli, N., Diouane, Y., Bussemaker, J., Lefebvre, T., Hwang, J. Morlier, J., Martins, J.R.R.A (2024). SMT 2.0: A Surrogate Modeling Toolbox with a focus on Hierarchical and Mixed Variables Gaussian Processes. Advances in Engineering Sofware.

[8] Saves, P., Halle-Hannan, E., Bussemaker, J., Diouane, Y., Bartoli, N. (2025). "Hierarchical Modeling and Architecture Optimization: Review and Unified Framework", arXiv:2506.22621.

[9] Werter, N. P. M, Breuker, R. de, "A novel dynamic aeroelastic framework for aeroelastic tailoring and structural optimisation," Composite Structures, pp. 369-386, 2016.