Machine teaching to optimize algorithms performances on restricted dataset.

Alexandre CARBONELLI[1], Maurin WOUSSEN [1]

[1]Vibratec, 28 Chemin du Petit Bois, 69130 Écully, France
alexandre.carbonelli@vibratec.fr

Abstract

Machine Learning techniques open important ways of development of physical models in almost every field. Performances reached rely on two major pillars: the (physical) model and the data. When a huge amount of data is available, the intrinsic characteristics of the chosen algorithm become less critical. On the other hand, with limited amount of data, all the human knowledge on the system to be modelled becomes critical to exploit.

One of the multiple applications of Machine Learning (ML) technics concerns the meta-models. Indeed, in this paper, we show how we can bypass a computation scheme by using clever regression models. The approach is performed on a system from which we want to know instantaneously the first natural frequencies without performing each time finite elements (FE) computation. We study the performance reached by studying the number of training cases required to teach the algorithm how to link inputs to outputs within a satisfying accuracy. Different algorithms are tested with very encouraging results as going into higher dimensional problem. The final aim of this study is to provide global guidelines for the most efficient Machine Teaching.

1 Introduction

Reducing the computation time is an objective is many cases. Whether it is for having instantaneously a result or to integrate the meta-model in a bigger loop, the idea remains the same: to link inputs to outputs without going through a whole computation process. The best way to achieve that is for instance to have an analytical formula based on physics consideration which can represent exactly the phenomenon studied. According to the task, it is unfortunately not always possible to use an analytical formula. The typical representative case of such a task would be for instance a classification problem: to determine if a given picture represents a cat or a dog, it seems obvious that we will not find an analytical formula able to do that. To illustrate our idea we can consider Figure 1.

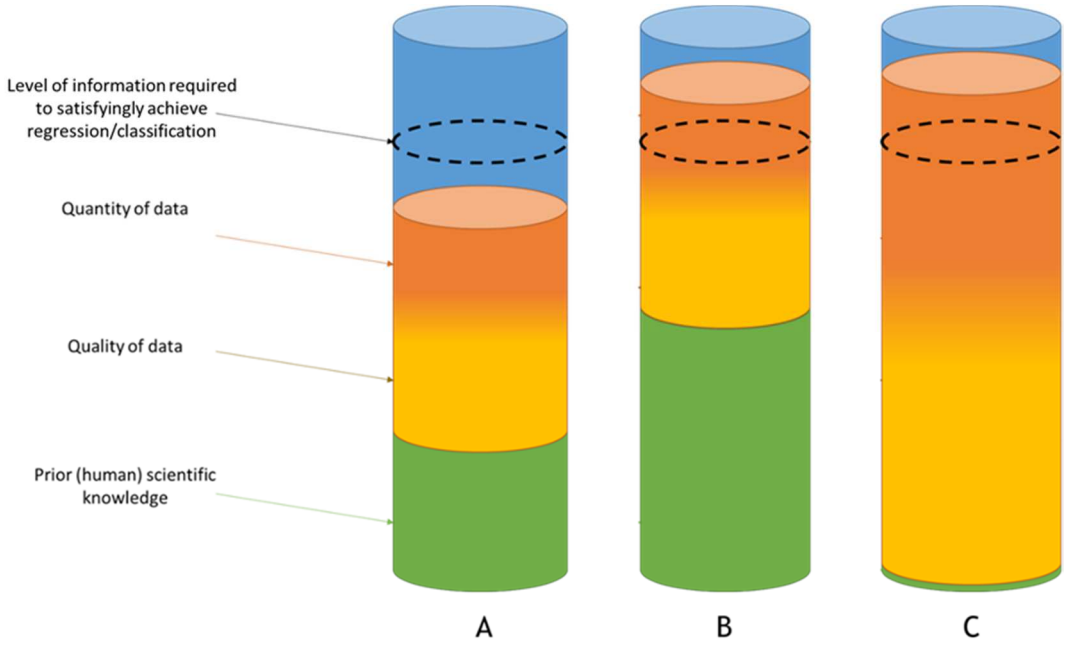

Figure 1: General approach of a regression/classification problem

When we lack information (case A), we can either use scientific knowledge (case B) or generate more data (case C) or of course both in order the fill the gap.
The general parameters suggested to be considered are:

- The level of information required to ***satisfyingly*** achieve the regression/classification implies to set a threshold beyond which we estimate that the task is a success,
- The quantity of data is directly linked to their quality (we consider that in the next part), meaning that many data of poor quality could not be that helpful according to the accuracy aimed for,
- The prior (human) scientific knowledge establishes the base of the whole stack (for instance, an adapted analytical formula fills the gap to the level of information required by its own).

This approach will be later included into a classification task consisting in detecting in a block made of carbon is cracked or not by studying its natural frequencies. We focus in this paper on the determination (through regression) of the natural frequencies of the block without any additional FE computation than those used for the model building. The accuracy of the reference natural frequencies (those obtained for a healthy block) must be high in order to be the more sensitive possible in the incoming classification task.

2 The concept of Machine Teaching

The Machine Teaching concept as we thought it relies on a direct analogy on human learning. Let's consider a student who has a set of examinations.

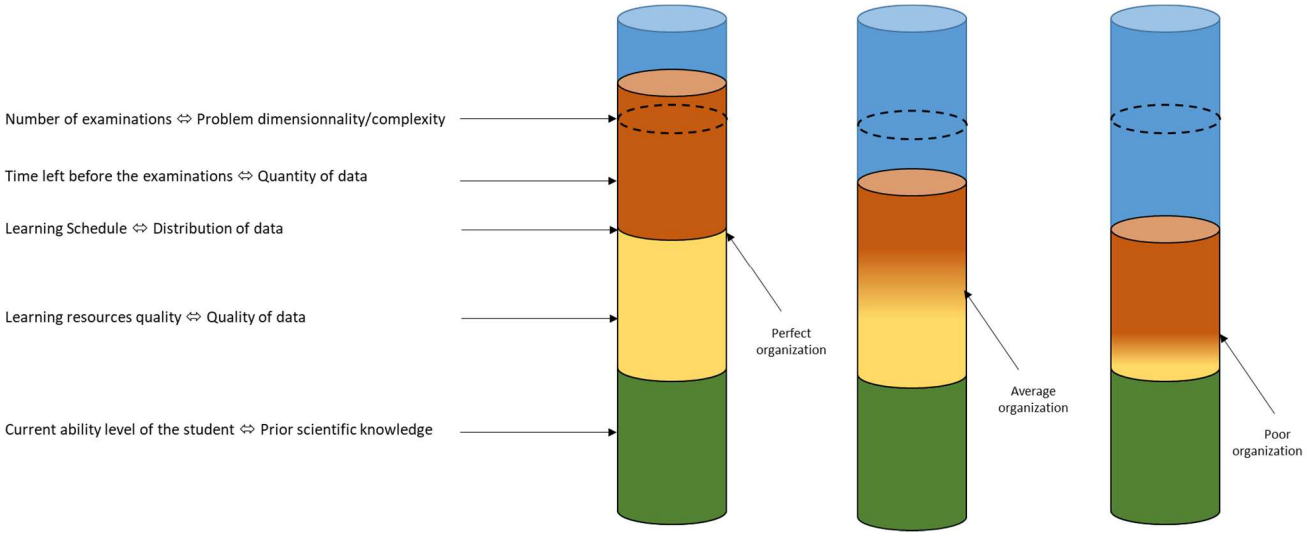

Figure 2: Illustration of the Machine Teaching concept

We suggest on Figure 2 the following symmetries. The number of examinations can be related to the complexity of the task or its dimensionality. The time before the examinations can be related to the quantity of data that one can generate in a given time. The learning resources (writing quality, the teacher implication and so on) can be related to the quality of data (does one need to correct and to complete manually the dataset, taking risk to insert errors…). The current ability of the student can be related on the priori scientific knowledge: what does he knew before any learning? His knowledge basis. Finally, the learning schedule is directly related for us to the concept of Machine Teaching. Indeed, having well distributed data (not necessarily uniform) permits to optimize the supply of each new data, avoid generalization problem, eliminate redundancies and perform better. To continue the analogy, for a student with a limited amount of time, a good schedule learning during which he will be able to the see different cases. The most representative cases will definitely improve his learning compared to spend most of the time on one chapter and ignoring the six left. Therefore a good teaching can be related to one ability to understand the knowledge compulsory to integrate and to provide consistent organization of the data.

3 Presentation of the study case

3.1 The free-free beam

The study of the free vibration of a beam is not that challenging. Nevertheless, classical approach using Euler-Bernoulli's and Timoshenko's beam theories [1-2] provides satisfying as long as some assumptions are satisfied, such as the ratio between the length of the beam and its cross-sectional dimensions (in which case the Euler-Bernouilli's beams present an important error). In our case, we don't want to make any assumptions about the beam behaviour and we use therefore FE model with solid elements to feed the meta-model. The Figure 3 shows two beam samples, which can be rectangular or circular. The Table 1 lists the different parameters considered and their variation ranges and Table 2 sums up the meshing details used.

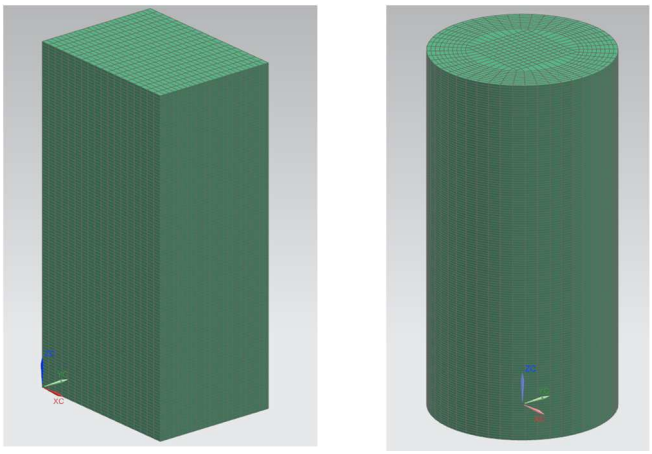

Figure 3: Examples of considered rectangular and circular beams.

Parameters	Minimum	Maximum
Length (m)	2	9
Section[1] (m)	0.3	1
Young's modulus (MPa)	7.92	11.88
Poisson's ratio	0.1	0.4

Table 1: Parameters studied and their variation ranges

	Rectangular beams	Circular beams
Element's type	Cube	Rectangular cuboid
Type of grid	Structured	Structured
Number of elements	50 112	93 600
Number of nodes	55 575	97 999

Table 2: Meshing details

3.2 Automatic generation of samples

The study was realized in two steps, a first naïve approach to the problem, and then a second step where we developed and improved a method. The different cases firstly used to feed the meta-model are generated automatically and are divided in N_1 circular and N_2 rectangular. We separate two different types of set, the training data used for the learning procedure, generated in a grid way, we will explain later the reason of this choice. And the test data, used for the meta-model validation, which are randomly generated, to emphasize the generalization on the entire study domain. We recall the different subsets in Table 3 and Table 4*: Subsets labels of the second step***Erreur ! Source du renvoi introuvable.**.
L, b, h, E, ν respectively stand for the length, the height, the thickness, the Young's Modulus and the Poisson's coefficient of the beam.

[1] Section corresponds to the larger length of the section's beam, i.e. diameter for circular beams or maximum between thickness and height for rectangular beams.

	First Step			
	Rectangular beams		Circular beams	
	Train	Test	Train	Test
Variables involved	b, h, L	b, h, L	b, h, L	b, h, L
Number of samples	512	141	400	75
Generation	Grid	Random	Grid	Random

Table 3: Subsets labels of the first step

	Second Step			
	Rectangular beams		Circular beams	
	Train	Test	Train	Test
Variables involved	b, h, L, E, ν	b, h, L, E, ν	b, h, L, E, ν	b, h, L, E, ν
Number of samples	100	50	100	50
Generation	Grid	Random	Grid	Random

Table 4: Subsets labels of the second step

We begin with 512 samples of rectangular beams for the training process, such as we have eight different values of length, thickness and height to make the grid. The initial distribution is naively uniform.
Its test set counts 141 random samples, with a criteria on the ratio, length over the working dimension of the section, where it can't overstep 3.33 for each configuration. The first training data for the circular beams contains 400 samples, such as we have twenty values of length and twenty values of diameter, to make a grid. We have 75 samples for its test set.

For the second step, we include the material properties in the study and we try to reduce the number of samples needed for the learning. For both of the circular and rectangular cases, we generate a grid such as we have 5 values of length, 5 values of thickness, 2 values of Young's modulus and 2 values of Poisson's ratio. We'll explain later the reasons of that configuration choice. As we add two new features, we also have to create new test sets of 50 samples each, generated randomly.

Finally, a total of 1428 samples have been automatically generated. To give an idea of the length of the process, one computation takes about three minutes. Then, to generate the complete set of input data, about 72 hours have been required.

3.3 Automatic extraction of data

The outputs required are simply the three first traction/compression (Z axis) eigenfrequencies and the three first flexural eigenfrequencies in both directions (X and Y axis). Each computation output file are post-process and for each case the 9 outputs are written related to the parameters defined in part 3.1.

4 Methodology for building the meta-model

Facing a lack of information, we firstly choose the random forest as our starting strategy to predict the eigenfrequencies. Indeed, the ensemble methods can have good results on small subsets, where they start from a high variance and thanks to bagging algorithm, decreases it during the learning process to improve the generalization [4]. However, many parts of the response values are underrepresented, which distorts the training phase, leading to a bad accuracy. We had to change our point of view on the procedure and try to develop a way to guide the algorithm during the training so that we can control how it learns along the process and "help" it. That's why we choose to build a custom learning method, closer to the machine teaching area. This method is based on nonlinear least squares [5] and the use of a specific grid as training data.
We can see in the Table 5: Comparison of relative error between random forest and a custom method a comparison on the worst mode's results, between random forest and our custom method. To do so, we calculate the relative error e such as:

$$e = \sqrt{\frac{(y-\tilde{y})^2}{y^2}} \times 100 \quad (1)$$

with y, the observation value and $\tilde{y}$, the prediction from the modelling function. This error is determined for each point of the test subset.

	Random Forest			Custom Method		
	F1X	F1Y	F1Z	F1X	F1Y	F1Z
Mean (%)	9.57 ; 29.1	13.7 ; 29.3	0.210 ; 5.38 10^{-2}	9.35 10^{-2} ; 0.122	0.100 ; 0.122	0.212 ; 0.265
Max (%)	46.54 ; 154	39.98 ; 156	2.42 ; 0.346	0.264 ; 0.554	0.255 ; 0.554	0.414 ; 0.843

Table 5: Comparison of relative error between random forest and a custom method
Rectangular beams / Circular beams

This table shows the limit of classical machine learning algorithms in this context. We had to modify existing methods to have a better fitting and more complex regression curve.

The main idea behind the method used is to deconstruct the learning procedure, so that we can control the tendency of our response depending on each features separately, during each step. We want to approximate the output with a function such as:

$$f: \quad I_1 \times \ldots \times I_N \rightarrow \mathbb{R} \qquad (2)$$
$$(v_1 \times \ldots \times v_N) \rightarrow f((v_1 \times \ldots \times v_N), \beta)$$

Where N is the number of features, $I_k \subset \mathbb{R}, k = \{1, \ldots, N\}$ the set where the feature is taken from, and β a set of parameters used by the function f. In our situation, we can choose the configuration $(v_1 \times \ldots \times v_N)$ we want to append in our samples set. Then, we choose to build a grid, as a combination of values from chosen subsets. The strategy is given in Appendix.

5 First step

5.1 Strategy

To start the study, several comments have been done to make easier the resolution. We choose to consider each eigenfrequency separately, instead of using multioutput regression algorithms which tends to be less accurate. To approximate the transverse modes along the Y axis, we used the symmetric properties of these modes and take the regression function on the transverse modes along X axis, where we inverse the thickness and height values. We now had to build 6 modelling functions. We also consider that the height does not influence the transverse modes along X axis, so the approximation will involve only two dimensions. For the longitudinal modes, we noticed that the eigenfrequencies are highly correlated with the length, so the regression function will only depend on this feature.
Our custom method were mainly used to build the transverse modes interpolation. So we have two layers, according to the features involved. The thickness layer is approximated with polynomials of second order and the parameters to estimate as related to a rational function such as:

$$f(x) = \frac{\beta_1 x^2 + \beta_2 x + \beta_3}{\beta_4 x^2 + \beta_5 x + 1} \quad (3)$$

Where x is the feature and $\beta_i, i = 1, \ldots, 5$ the parameters. The final function looks like:

$$f(b,L) = \frac{\alpha_1 L^2 + \alpha_2 L + \alpha_3}{\alpha_4 L^2 + \alpha L + 1}\, b^2 + \frac{\beta_1 L^2 + \beta_2 L + \beta_3}{\beta_4 L^2 + \beta_5 L + 1}\, b + \frac{\gamma_1 L^2 + \gamma L + \gamma_3}{\gamma_4 L^2 + \gamma L + 1} \quad (4)$$

With $\alpha_i, \beta_i, \gamma_i, i = 1, \ldots, 5$ the final parameters.
For the longitudinal modes, we did a log-log transformation then a linear regression on the result, such as $f(L) = L^{\beta_1} e^{\beta_2}$.

5.2 Results

5.2.1 Longitudinal modes

The relative error mentioned (1) is used to explore the function's accuracy and validate the model. We deduct some statistic information of its tendency, such as the mean above the values, the standard deviation around it and maximum and minimum value of the error data. This procedure is kept for all the results in the study.
The mentioned data are presented in **Erreur ! Source du renvoi introuvable.**.

	Rectangular beams			**Circular beams**		
	F1Z	F2Z	F3Z	F1Z	F2Z	F3Z
Count	141	141	141	75	75	75
Mean	0.212	0.202	0.198	0.265	0.263	0.296
Standard Deviation	$9.31\ 10^{-2}$	$8.99\ 10^{-2}$	0.138	0.215	0.241	0.295
Min	$7.63\ 10^{-2}$	$4.86\ 10^{-2}$	$9.36\ 10^{-3}$	$1.12\ 10^{-2}$	$2.01\ 10^{-3}$	$1.53\ 10^{-3}$
Max	0.414	0.432	0.608	0.843	0.973	1.42

Table 6: Error (%) for longitudinal modes with circular beams during the study's first step

We can already be satisfied by the error which is less than 1%, except for the third longitudinal modes with the circular beams, where we notice a peak at 1.42%. All the means are around 0.2% of error which is promising for the efficiency of the method used.

We also represent the error depending on the ratio, to enhance the influence of ratio as major error vector. The results are represented in the Figure 4.

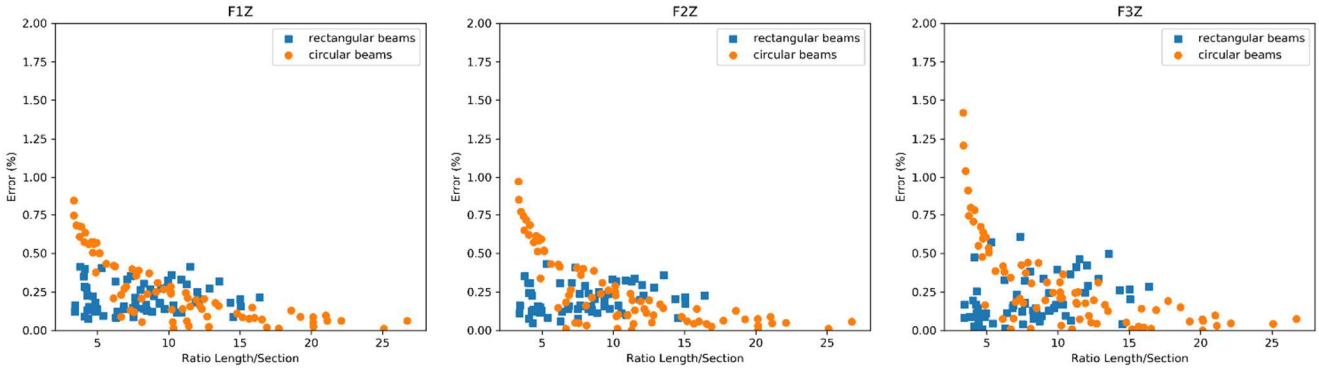

Figure 4: Errors in function the ratio for longitudinal modes during the study's first step

The error has a dependency with the ratio which increases when this ratio decreases. The difference of behavior between circular and rectangular beams is due to the difference of the grid's construction used.

5.2.2 Transverse modes

Table 7: Error (%) for transverse modes with rectangular beams during the study's first step and Table 8*: Error (%) for transverse modes with circular beams during the study's first step* contain the statistic of the error on the transverse modes. We still have an error less than 1% except a peak for the third modes which results of its higher complexity. The means tends to evolve around 0.15% which is still promising.

	F1X	F2X	F3X	F1Y	F2Y	F3Y
Count	141	141	141	141	141	141
Mean	$9.35\ 10^{-2}$	0.128	0.126	0.100	0.141	0.139
Std. Dev.	$6.76\ 10^{-2}$	$8.58\ 10^{-2}$	$8.98\ 10^{-2}$	$6.55\ 10^{-2}$	$8.53\ 10^{-2}$	$8.31\ 10^{-2}$
Min	$1.04\ \ 10^{-3}$	$2.44\ 10^{-3}$	$1.45\ 10^{-3}$	$5.45\ 10^{-3}$	$6.17\ 10^{-3}$	$5.08\ 10^{-4}$
Max	0.264	0.323	0.305	0.255	0.324	0.344

Table 7: Error (%) for transverse modes with rectangular beams during the study's first step

	F1X	F2X	F3X	F1Y	F2Y	F3Y
Count	75	75	75	75	75	75
Mean	0.122	0.160	0.185	0.122	0.153	0.187
Std. Dev.	0.130	0.153	0.187	0.130	0.153	0.187
Min	$9.52\ 10^{-4}$	$1.21\ 10^{-3}$	$2.92\ 10^{-3}$	$9.52\ 10^{-4}$	$1.21\ 10^{-3}$	$2.92\ 10^{-3}$
Max	0.554	0.877	1.23	0.554	0.877	1.23

Table 8: Error (%) for transverse modes with circular beams during the study's first step

We have globally the same accuracy as for the longitudinal modes, the distribution is wider which can be explained by the more complex behavior of these quantities. The use of 2 layers here can also add some deviation for some values.

We have a good accuracy, even if we have some peaks for small ratio which confirming the changing behavior of this area as shown Figure 5.

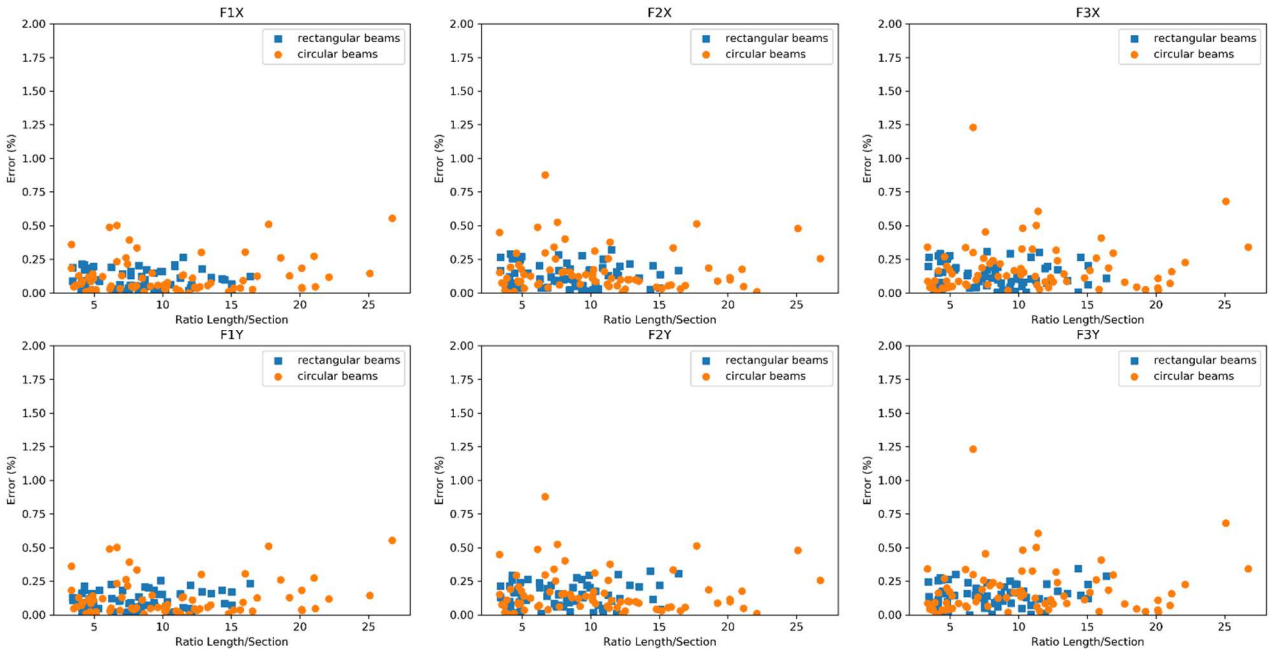

Figure 5: Error in function of the ratio for transverse modes during the study's first step

5.3 Critics and way of improvement

Even if the accuracy could be higher, we have satisfying results for this first meta-model, which gives us 1% error on our generalization data. This results is satisfying for the model. As we are in a problem with maximum two dimensions, having only one feature to fit the longitudinal works well in this specific case.

Moreover, we want to include in the second step the material properties of the beam, so we will have to be more flexible on the complexity of our regression function, even for the longitudinal modes. A more specific way of determining the training samples has to be built, so that we minimize its size. The next step will be to set a guideline to choose an initial subset and optimize the number of samples needed to have a great accuracy.

6 Second step

6.1 Strategy

As we have two new features involved, another grid needs to be made. We want to minimize the number of samples, so the first idea is to keep as much variance as possible in the domain of study, while maintaining the properties of a grid. Nevertheless, during the study, we set a constrain on the ratio, where for a beam too close of a cube, we are not sure to find a third longitudinal mode, due to its geometry. To respect this constrain, instead of taken the length between 2 and 9 meters, we set its boundaries between 3.33 and 9 meters.

The optimization is mainly done by the number of values we choose for each feature. Thanks to theoretical properties, we know that the eigenfrequencies are linearly dependent with Young's modulus and Poisson ratio [6]. So for the layers associated to material features, we needed two values each to make a linear regression. Then for the other features, we want to approximate data with curves which have a certain complexity, so we decided to choose five values as our beginning training set. The five values will allow us to choose function including five parameters and this gives us a first flexibility on function complexity.

We took the same comments as the previous step, where only six modes are approximated and the height is not included as a relevant feature for both the transverse and longitudinal modes. Moreover, for the rectangular beams, we generate randomly the height always strictly lower than the thickness, so that we avoid redundancy in the training data.

We also improve the custom method, now for each layer and for each parameters, the function to interpolate can be chosen. This increases a lot the complexity of the final function and allows us to fit more precisely our data and have a visualization of the fitting at each step of the process. We chose among the following functions to adjust the data:

- Linear : $f(x) = \beta_1 x + \beta_2$
- Polynomial of degree 2 : $f(x) = \beta_1 x^2 + \beta_2 x + \beta_3$
- Polynomial of degree 3 : $f(x) = \beta_1 x^3 + \beta_2 x^2 + \beta_3 x + \beta_4$
- Rational function of first order : $f(x) = \frac{\beta_1 x + \beta_2}{\beta_3 x + 1}$
- Rational function of second order : $f(x) = \frac{\beta_1 x^2 + \beta_2 x + \beta_3}{\beta_4 x^2 + \beta_5 x + 1}$
- Exponential : $f(x) = \beta_1 e^{\beta_2 x}$

With $\beta_i, i = 1, \dots, 5$ the parameters to adjust.

6.2 Results

6.2.1 Longitudinal modes

In the Table 9*: Error (%) for longitudinal modes with rectangular beams during the study's second step*, we present the results of longitudinal modes for the second step of the study.

	Rectangular beams			Circular beams		
	F1Z	F2Z	F3Z	F1Z	F2Z	F3Z
Count	50	50	50	50	50	50
Mean	0.378	0.412	0.582	0.215	0.278	0.351
Standard Deviation	0.138	0.223	0.442	0.221	0.263	0.291
Min	$1.62\ 10^{-2}$	$6.31\ 10^{-3}$	$4.20\ 10^{-4}$	$5.03\ 10^{-4}$	$6.00\ 10^{-5}$	$8.71\ 10^{-4}$
Max	0.570	0.954	2.30	0.565	0.763	0.886

Table 9: Error (%) for longitudinal modes with rectangular beams during the study's second step

The error is slightly higher than the first step and increases with the order of the mode, but the means are under 0.6% of error, which is a promising accuracy for this initial grid.

The Figure 6: Error in function of the ratio for longitudinal modes during the study's second presents

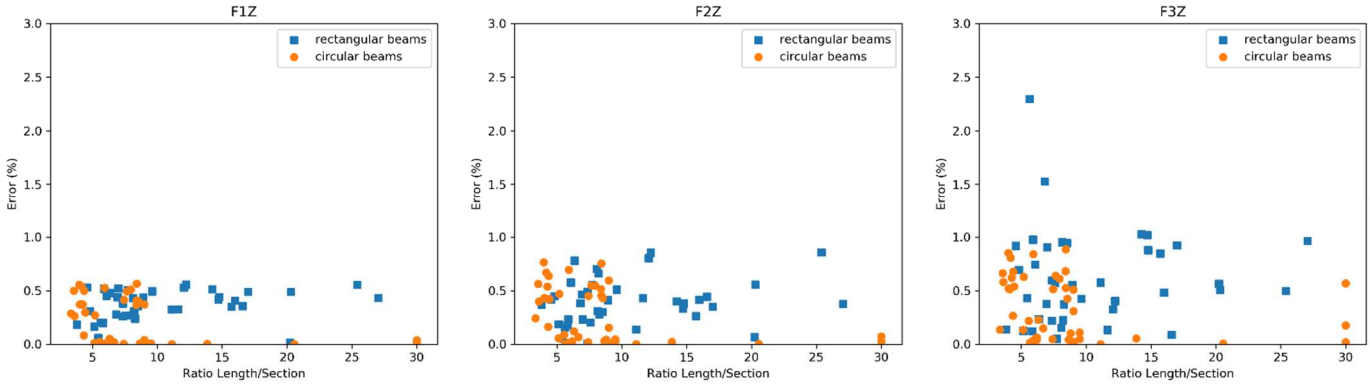

Figure 6: Error in function of the ratio for longitudinal modes during the study's second

the error depending on the ratio of the beams.
The rise of the error depending on the mode's order is more prominent here, with still some perturbations for small ratio.

6.2.2 Transverse modes

In the Table 10*: Error (%) for transverse modes with rectangular beams during the study's second step* and

Table 11: Error (%) for transverse modes with circular beams during the study's second step, we present the results of transverse modes for the second step of the study.

	F1X	F2X	F3X	F1Y	F2Y	F3Y
Count	50	50	50	50	50	50
Mean	0.430	0.487	0.512	0.399	0.678	0.829
Standard Deviation	0.203	0.250	0.379	0.166	0.494	0.623
Min	$1.93\ 10^{-2}$	$6.66\ 10^{-2}$	$2.76\ 10^{-2}$	$3.79\ 10^{-2}$	$1.45\ 10^{-2}$	$6.23\ 10^{-1}$
Max	1.05	1.24	1.92	0.837	2.48	3.03

Table 10: Error (%) for transverse modes with rectangular beams during the study's second step

	F1X	F2X	F3X	F1Y	F2Y	F3Y
Count	50	50	50	50	50	50
Mean	0.186	0.309	1.36	0.189	0.309	1.36
Standard Deviation	0.173	0.472	1.00	0.172	0.470	1.00
Min	$4.02\ 10^{-3}$	$1.26\ 10^{-4}$	$2.73\ 10^{-2}$	$4.02\ 10^{-3}$	$1.26\ 10^{-4}$	$3.73\ 10^{-2}$
Max	0.475	3.25	3.64	0.475	3.24	3.63

Table 11: Error (%) for transverse modes with circular beams during the study's second step

The same remark as the previous section can be made, the error is higher than the first step but still acceptable, considering the number of training data.

In the Figure 7*:* Error in function of the ratio *for transverse modes during the study's second*, the graphs show the limits of the method and the grid, where we have generalization issues, especially for small ratios and third modes.

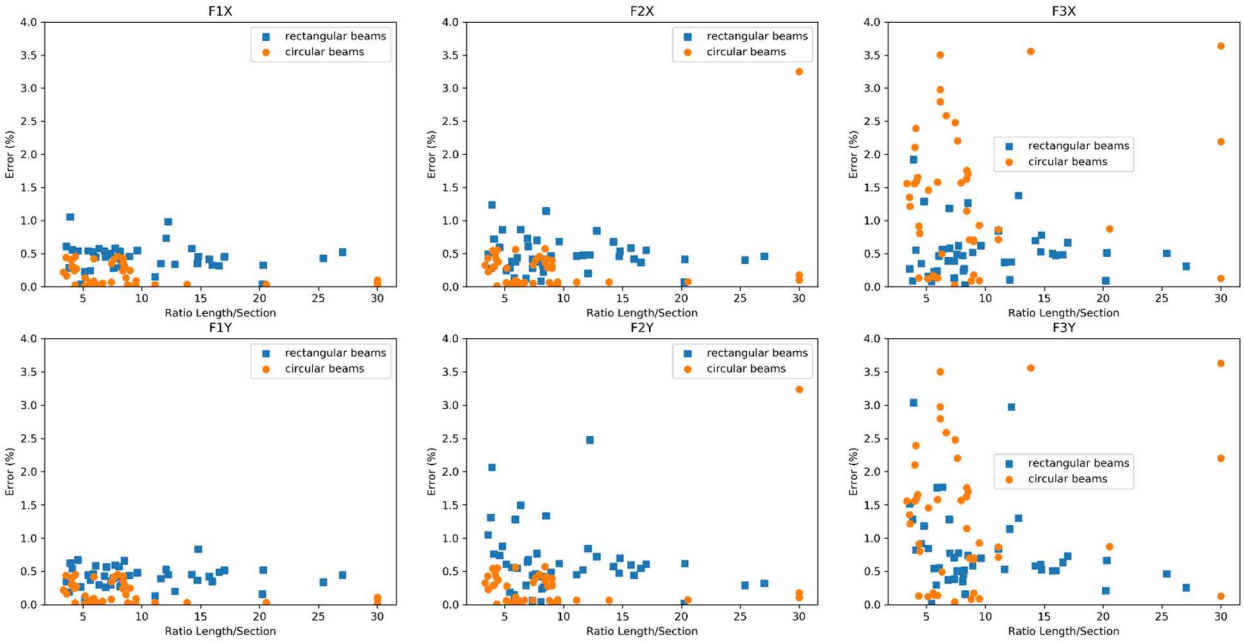

Figure 7: Error in function of the ratio for transverse modes during the study's second

6.3 Critics and ways of improvement

The error increases but is still acceptable. Moreover, we have to consider the context of the learning which is more complex than the previous step. The error is twice higher than the first step, under 3% error of accuracy. It is worth remembering that we added two features and decreases the training data from 512 to 100 samples.

A first improvement would be to stabilize generalization's error. To do so, we will propose a guideline to select leading samples to add in the training data which will increase the accuracy. The criteria to build the initial grid can also be improved. Indeed, we didn't train the model with length under 3.33

meters and it impacts the generalization as we see in the results. A new methodology could spare the lack information in this ratio's area.

6.4 Guideline for sample selection

We propose a first guideline to increase the precision by adding some relevant samples. The initialization starts with the construction of a first grid, with a minimum number of samples. Here it will depends on the nature of the features and the domain of study. If some follows a specific tendency, then we know the minimum number of requisite values to approach them.
In order to get a satisfying learning, we need to have a uniform representation of the response on its boundaries. Then, with the modelling function learnt with the initial grid, the purpose is to find the missing configuration with that function, to restore a uniformly represented response.
We can either use directly the reciprocal function of our final regression function, or generate a set of values with that function and determine the closest ones to our wanted configuration. Once we have all the inputs needed, we have to adjust those values to a grid which will be added to the initial grid and used in the next iteration.

7 Discussion and perspectives

The first aim of this study was to build a model to obtained directly the 9 first eigenfrequencies of interest of a circular or rectangular beam in order to lately integrate this model in a larger loop which purpose is to classify carbon blocks whether there are susceptible to contained cracks or not, and this the more precisely possible. This has been achieved with success.
The second aim was to investigate the concept of Machine Teaching in the sense that we were searching for guidelines to minimize the number of input data required conserving a given accuracy goal. To illustrate the idea, we can think about Figure 1, and more particularly at the link between the quality and the quantity of data. Indeed, by structuring cleaver the distribution of the input data taking into account the produced output data, we strongly believe that we can reach higher levels of regression or classification ability while minimizing the number of initial computations required to feed the model. That is of first importance when the computation time last hours or days.
This study based on a simple model permitted to not be hampered by too long computational times in order to find the optimal strategy. The further work will consist in applying those guidelines on a more complex problem and to implement it for optimization purposes.

References

[1] Méndez-Sánchez, R. A., Morales, A., & Flores, J. (2005). Experimental check on the accuracy of Timoshenko's beam theory. *Journal of Sound Vibration*, *279*, 508-512.
[2] Mokhtari, A., Sarvestan, V., Ghayour, M., & Mirdamadi, H. R. Free vibration of undamped Euler-Bernoulli and Timoshenko beams by using spectral finite element method for different boundary conditions.
[3]Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*(Oct), 2825-2830.
[4] Qi, Y. (2012). Random forest for bioinformatics. In *Ensemble machine learning* (pp. 307-323). Springer, Boston, MA.
[5] Kenneth Levenberg (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares.
[6] Michel Lalanne, Patrick Berthier, Johan Der Hagopian. (1984). Mechanical vibrations for engineers. *J. Wiley.*

A Appendix

The grid used is built with the following pattern:

$$
\begin{aligned}
&choose\ m_1\ points\ in\ I_1, such\ as\ , \left\{v_1^{(1)}, \dots, v_{m_1}^{(1)}\right\} \\
&\qquad\qquad\qquad\vdots \\
&choose\ m_N\ points\ in\ I_N, such\ as\ , \left\{v_1^{(N)}, \dots, v_{m_N}^{(N)}\right\} \\
&for\ i_1 = 1, \dots, m_1 : \\
&\quad \ddots \\
&\qquad for\ i_N = 1, \dots, m_N: \\
&\qquad\quad Add\ \ (v_{i_1}^{(1)}, \dots, v_{i_{N_1}}^{(N)})
\end{aligned}
$$

This kind of combination leads us to $\prod_{i=1}^{N} m_i$ samples. In this way of data generation, we can explore a domain study with a chosen discretization, which can be useful in our case. A second benefit of the grid, is the possibility to observe the response depending on only one feature, where all the others are fixed. We will use this ability for the visualization.

We start by unrolling the grid's construction operations, by fixing each feature, then we have a 2D graph with the response associated to the set $\left\{v_1^{(N)}, \dots, v_{m_N}^{(N)}\right\}$. Then, we interpolate the graph by a chosen function, with the nonlinear least squares algorithm, which gives us the best parameters to fit the interested values. After repeating this curve fitting for each configuration on the last loop, we interpolate the parameters of the chosen function with the set $\left\{v_1^{(N-1)}, \dots, v_{m_{N-1}}^{(N-1)}\right\}$. The process is done when all the features were involved in the learning process.

We can illustrate that decomposition with an example. Let's consider a problem with three real input features, *A*, *B* and *C,* and *F* as a real output feature. In order to make a grid, we choose arbitrary two points for each input, and by applying the previous building method, we get the following combinations (Table).

A	**B**	**C**	**F**
$\boldsymbol{A_1}$	B_1	C_1	F_1
$\boldsymbol{A_1}$	B_1	C_2	F_2
$\boldsymbol{A_1}$	B_2	C_1	F_3
$\boldsymbol{A_1}$	B_2	C_2	F_4
$\boldsymbol{A_2}$	B_1	C_1	F_5
$\boldsymbol{A_2}$	B_1	C_2	F_6
$\boldsymbol{A_2}$	B_2	C_1	F_7
$\boldsymbol{A_2}$	B_2	C_2	F_8

Table 12: One grid example

Then, we can apply the decomposition, which can be illustrated with a graph, like the one in Figure 8. The set of nodes corresponding to one feature is a step of the algorithm, which can also be described as a feature layer. Only the last layer will proceed the response, corresponding to the first step of the method. Then all the layers will deal with the parameters of the function fitted in the layer before.

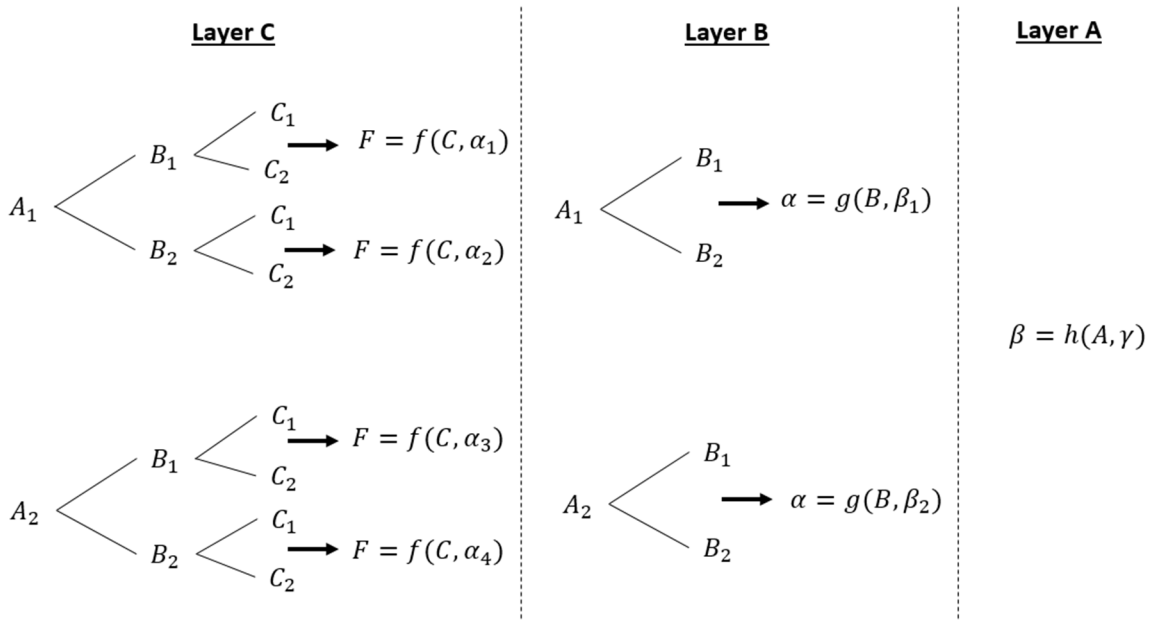

Figure 8: Graph of each layer of the example

The purpose is to manually select the best fitted function among a lot of possibilities, in terms of intrinsic properties such as convexity, asymptotic behavior or the level of complexity. The construction of the final modelling function is done by a composition of all the functions involved during each feature's layer. In our example, the final regression function would be:

$$\tilde{F}\ (A, B, C) = f(C, g\big(B, h(A, \gamma)\big)) \tag{5}$$

The interest of this method, is firstly the high complexity of the final function, often with several input features, where we adjust several functions on a 2D graph. It also decomposed the training phase in several steps, where a human can control the level of accuracy. The possibility of choosing the modelling function at each layer for each parameter, depending of the tendency followed by the values is a determinant way to obtain maximum precision with a highly restricted data set.