

Predictions of the simulated modal properties for ultrasonic welding sonotrodes

Evaluation technical report

Promotion : sep24_cds_int_ultrasonic-welding

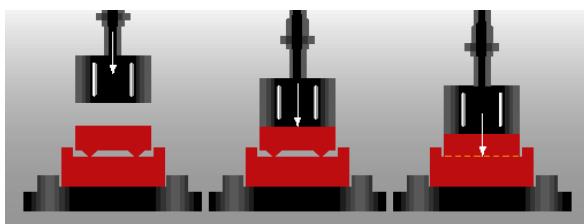
Participants : Brouet François

Evaluation technical report

1 Context

1.1 Ultrasonic welding

The concept of ultrasonic welding involves converting the kinetic energy of vibration into the components to be joined (typically made of plastics or metals). This energy is transformed in heat by the friction between the surfaces. Combined with a mechanical force the material starts to melt and a new homogeneous connection is created.



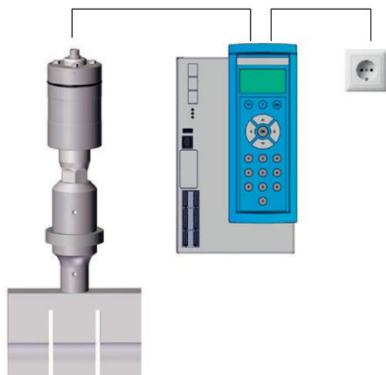
1.2 Sonotrode

The sonotrode is a component of the vibrating stack that oscillates at an ultrasonic frequency. It allows to transfer mechanical vibration energy to the welding part and maintains a constant high level of welding quality.



(a)

- (a) Welding machine
- (b) Ultrasonic stack
- (c) Ultrasonic generator



(b)

(c)

Evaluation technical report

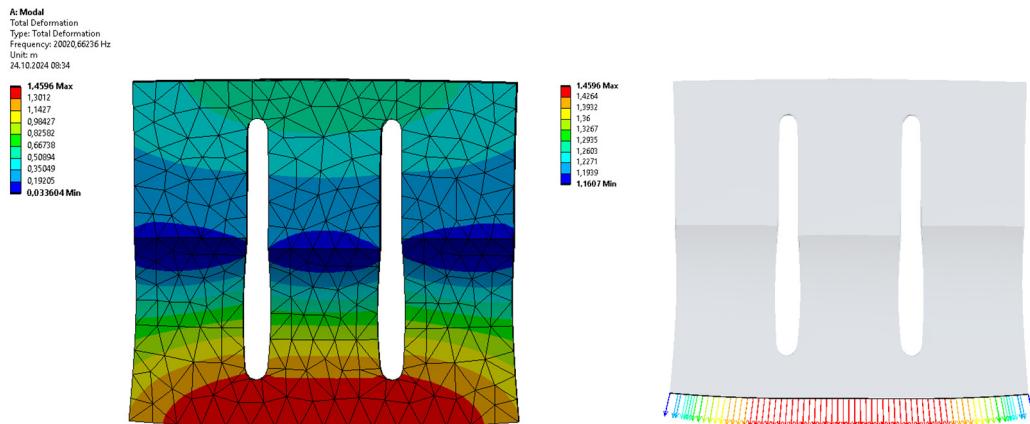


(Example of different sonotrodes)

Like any other solid mechanical part, the sonotrode has multiple vibration modes and the first sonotrode's longitudinal mode is particularly used to drive the vibration. The sonotrode geometry has to be precisely optimized to meet several precise requirements.

Due to the complexity of the sonotrode, the geometry is usually imported to a Finite-Element-Methode (FEM) to determine its modal properties (such as frequency and deformations, and to estimate whether the sonotrode meets certain requirements:

- the longitudinal mode's frequency should be close to the working frequency (in this case 20 kHz)
- The longitudinal mode's displacement should be uniform over the welding surface.
- The longitudinal mode should be isolated from other modes in terms of frequency.



1.3 Tuning of the sonotrode geometry

To meet the specifications, the geometry is tuned iteratively. As the length is used to adjust the longitudinal frequency, other geometric parameters, such as the number of slots, slot position, and thickness, can be modified to enhance deformation and frequency of many modes. These optimizations are performed most of the time manually and depend on the experience of the acoustic engineer.

1.4 Goals

The main goal of this project is to predict the results of FEM simulations without running them. Simulations of complex geometries are computer-intensive, and one iteration can take from a few minutes to an hour. From an economic point of view, the potential savings in calculation time are obvious.

Evaluation technical report

The Machine Learning model not only provides predictions, but also the influence of the parameters. It is of great interest to visualize these effects and compare them to the typical optimization methods used in practice.

From a scientific point of view, the model approximation establishes a multidimensional surface that allows an optimizer to navigate it easily to find the best points. It does not need to run the simulations, so it calculates the best possible geometry to meet the requirements for the sonotrode.

Different projects about the automatic design of sonotrodes are ongoing in the company. The greatest challenges remain:

- To precisely parameterize the geometry (and effectively determine the optimal parameter ranges) in order to accurately represent the actual geometry;
- To simultaneously optimize multiple variables, such as frequencies and uniformity,
- To deal with mode swapping: The oscillation modes are calculated and numbered according to their frequency. During the parameter variation, two modes may exchange positions.

1.5 Target variables

This work focuses on predicting three different values (target variables) using machine learning. For each one, it provides a global analysis that includes visualization, parameter influences, model development, and result interpretation:

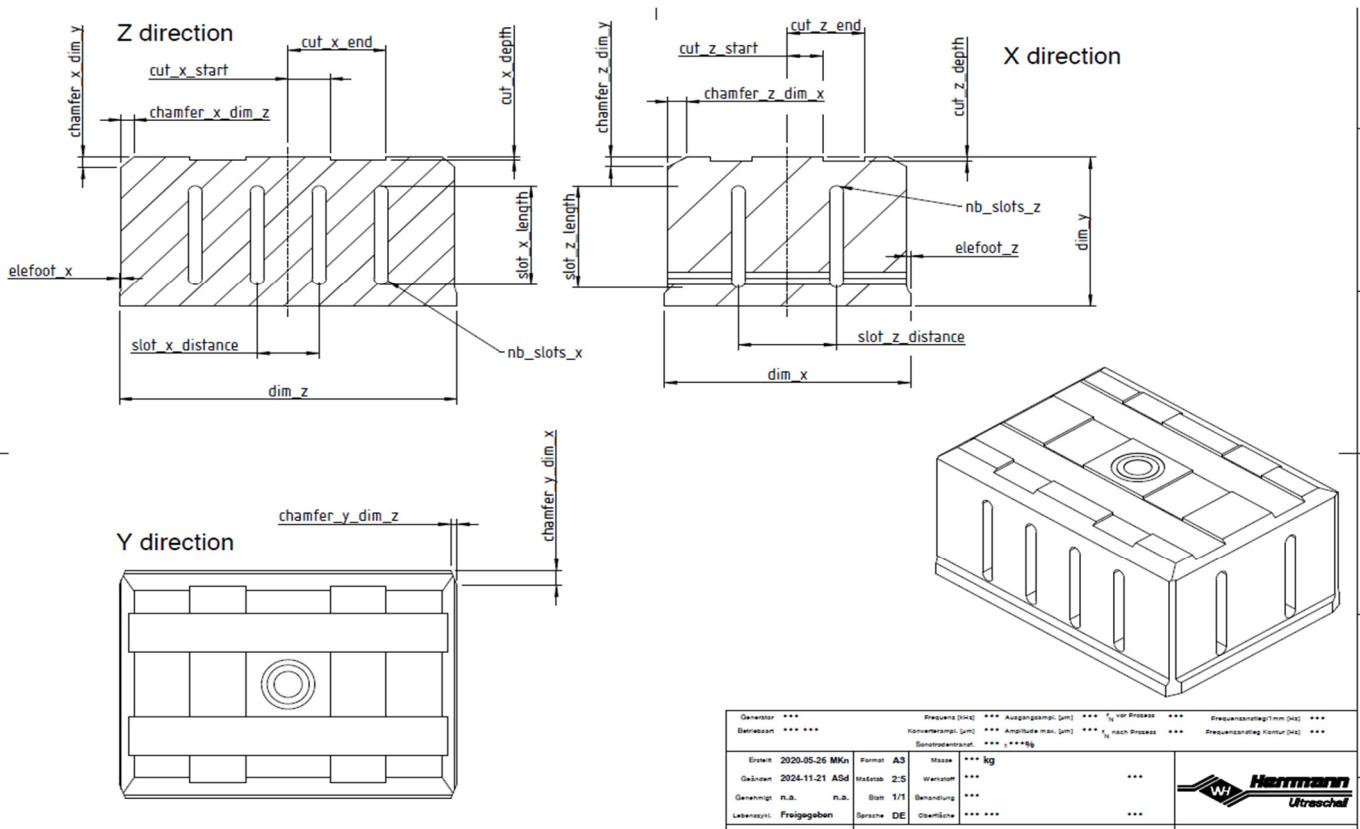
- Prediction of the longitudinal mode's frequency
- Prediction of the frequency of all other modes or alternatively only those modes near the longitudinal mode (greater precision is required, particularly in the vicinity of the first longitudinal mode, for example the 5 modes below and 5 modes above).
- Prediction of the longitudinal mode's displacement along the output surface or alternatively only the uniformity of these displacements (as a scalar number).

2 Data

2.1 Framework

In the study, the software ANSYS is used to simulate the modal properties of the sonotrode. These properties are influenced by the material's properties (assumed constant) and the geometry of the sonotrode, which is described in this project by 23 parameters. A single design point represents a unique geometry as a 23-dimensional vector.

Evaluation technical report



The data was directly generated by simulating many hundreds of geometries. The design points are randomly distributed in a typical parameter space, for a sonotrode with a welding surface of approximately 175 mm (dimension dim_x) and 240 mm (dimension dim_z).

2.1.1 Identification of the longitudinal mode

It would have been possible to identify the first longitudinal mode by analyzing the displacement data, but for a matter of time this step was skipped and performed directly in ANSYS. Following each simulation, ANSYS compared the shape of all modes against a standard vector, calculating the MAC correlation coefficients [1].

2.1.2 Volume of the dataset

An overview of the data is presented in the appendix. In total, the data set spans approximately 24,000 files and occupies around 8 gigabytes.

The design points are listed in the CSV file `design-points-03.csv`. The file contains 8001 lines (design points) and 24 columns (geometrical parameters).

Evaluation technical report

Number of lines in the table : 8001									
# Column	Name of the Column	Variable's type	Description	Is the variable available before prediction	Variable's type	Percentage of missing values	Categorical / Quantitative	Distribution	Comments
		Is the variable a feature or the target ? (Only applicable for supervised learning projects)	What does this variable represent (from a business perspective ?)	Is this variable known before the prediction is made ? (Only applicable for supervised learning projects)	int64, float etc... If "object", develop.	in %		For categorical variables with less than 10 categories, list all categories.	Free text
1	dp_no	feature	Design point number	yes	int64	0,00%	quantitative	between 1000 and 11999	
2	dim_x	feature	Thickness	yes	float	0,00%	quantitative	between 173 and 177	
3	dim_y	feature	Length	yes	float	0,00%	quantitative	between 115 and 125	
4	dim_z	feature	Width	yes	float	0,00%	quantitative	between 238 and 242	
5	nb_slots_x	feature	Number of slots X	yes	int64	0,00%	categorical	3 or 4	
6	slot_x_length	feature	Slot X length	yes	float	0,00%	quantitative	between 74 and 83	
7	slot_x_distance	feature	Slot X distance	yes	float	0,00%	quantitative	between 41 and 55	
8	nb_slots_z	feature	Number of slots Z	yes	int64	0,00%	categorical	2 or 3	
9	slot_z_length	feature	Slot Z length	yes	float	0,00%	quantitative	between 74 and 83	
10	slot_z_distance	feature	Slot Z distance	yes	float	0,00%	quantitative	between 49 and 67	
11	chamfer_x_dim_y	feature	Chamfer XY	yes	float	0,00%	quantitative	between 0,1 and 20	
12	chamfer_x_dim_z	feature	Chamfer XZ	yes	float	0,00%	quantitative	between 0,1 and 20	
13	chamfer_y_dim_x	feature	Chamfer YX	yes	float	0,00%	quantitative	between 0,1 and 20	
14	chamfer_y_dim_z	feature	Chamfer YZ	yes	float	0,00%	quantitative	between 0,1 and 20	
15	chamfer_z_dim_x	feature	Chamfer ZX	yes	float	0,00%	quantitative	between 0,1 and 20	
16	chamfer_z_dim_y	feature	Chamfer ZY	yes	float	0,00%	quantitative	between 0,1 and 20	
17	elefoot_x	feature	Elefant foot X	yes	float	0,00%	quantitative	between 0,1 and 5	
18	elefoot_z	feature	Elefant foot Z	yes	float	0,00%	quantitative	between 0,1 and 5	
19	cut_x_start	feature	Cut off X start	yes	float	0,00%	quantitative	between 20 and 60	
20	cut_x_end	feature	Cut off X end	yes	float	0,00%	quantitative	between 60 and 100	
21	cut_x_depth	feature	Cut off X depth	yes	float	0,00%	quantitative	between 0,1 and 5	
22	cut_z_start	feature	Cut off Z start	yes	float	0,00%	quantitative	between 20 and 45	
23	cut_z_end	feature	Cut off Z end	yes	float	0,00%	quantitative	between 45 and 70	
24	cut_z_depth	feature	Cut off Z depth	yes	float	0,00%	quantitative	between 0,1 and 5	

(Data audit of the design point file)

Overview of the outcome:

The main file mode-ident-full-03.dat lists the main results of the design points: 8001 lines (design points) and 8 columns containing informations on the identified longitudinal mode.

Individual result files: "modes", "nodes", "defs"

The simulation results are stored in three subdirectories and contain the detailed results for one design point (including all modes, node coordinates and deformations).

2.2 Relevance

2.2.1 DataFrames

Due to the large number of files, it is necessary to process and clean the data. To avoid working with raw data in the subsequent steps, all the data is collected and stored in a dictionary of DataFrames.

The dictionary key "parameters" combines the geometrical parameters and the design points specific features (such as the identification of the longitudinal mode).

all_data['parameters']	
0	dp_no dim_x dim_y dim_z nb_slots_x slot_x_length slot_x_distance nb_slots_z slot_z_length slot_z_distance ... cut_z_start cut_z_end cut_z_depth mode_no freq rel_dev mean_in_disp mean_out_disp mode_gain mode_mac
1	1000 175.00 120.00 240.00 4 77.000 48.000 2 77.000 58.000 ... 25.000 55.000 0.10000 NaN NaN NaN NaN NaN NaN
2	1001 175.06 121.95 240.31 4 75.993 43.235 2 78.219 59.508 ... 25.293 66.060 1.99450 77.0 19217.965 0.92217 -0.371 0.280 0.75609 0.921784
3	1002 173.45 122.77 238.13 4 81.168 45.859 2 78.924 57.208 ... 26.997 56.481 4.80300 79.0 19386.659 0.95697 -0.386 0.293 0.75940 0.923429
4	1003 175.85 116.71 238.10 4 77.532 45.260 2 80.099 49.215 ... 22.264 51.204 0.81847 79.0 20087.459 0.51795 -0.337 0.321 0.95450 0.971387
5	1004 173.51 116.37 238.30 4 82.428 51.007 2 74.757 50.840 ... 37.379 58.304 1.02820 80.0 20209.177 0.79699 -0.351 0.319 0.90825 0.948260
6
8074	11995 175.18 119.50 239.53 2 78.314 80.925 2 77.038 54.535 ... 30.256 66.937 4.66870 55.0 20351.555 1.20382 -0.309 0.295 0.95621 0.913093
8075	11996 176.27 117.29 240.59 2 77.933 71.773 2 79.789 63.279 ... 33.988 64.708 4.83750 58.0 20703.179 1.93926 0.355 -0.259 0.73108 0.764613
8076	11997 174.01 117.81 239.26 2 80.402 81.905 2 79.398 60.146 ... 39.064 50.663 1.04200 56.0 20621.116 1.13667 -0.322 0.304 0.94193 0.919224
8077	11998 174.61 119.50 239.41 2 74.848 74.107 2 74.870 57.402 ... 35.845 56.585 4.99810 55.0 20058.828 4.79138 -0.112 0.218 1.94744 0.328624
8078	11999 176.58 122.07 238.37 2 78.904 68.789 2 77.839 61.872 ... 33.453 60.300 4.89750 55.0 19915.683 0.86152 0.273 -0.286 1.04981 0.936144

8079 rows × 31 columns

Evaluation technical report

The key “modes” includes all the modes with position between 25 and 100:

all_data[fam]['modes']							
dp_no	mode_no	freq	rel_dev	mean_in_disp	mean_out_disp	mode_gain	mode_mac
0	1000	25	11705.549	120.61732	-0.121	0.002	0.01985
1	1000	26	11987.927	412.69239	-0.009	0.002	0.26013
2	1000	27	12187.894	474.99649	0.000	0.001	2.26471
3	1000	28	12318.561	11.80434	-0.065	-0.046	0.71102
4	1000	29	12598.709	847.60025	0.001	0.000	0.92230
...
607463	11999	96	27491.449	967.39541	0.001	-0.001	1.31907
607464	11999	97	27520.148	245.47758	-0.001	-0.005	4.99439
607465	11999	98	27581.676	422.81995	-0.002	0.004	1.66456
607466	11999	99	27587.681	100.51100	0.008	0.017	2.04339
607467	11999	100	28165.665	1277.52327	-0.000	0.001	4.75540

607468 rows × 8 columns

The key “nodes” accesses the list of nodes on the output surface (whose coordinates change with the design points):

all_data[fam]['nodes']				
dp_no	node_no	x_coord	y_coord	z_coord
0	1000	1	-120.015000	0.0
1	1000	11	-120.015000	0.0
2	1000	12	-120.015000	0.0
3	1000	13	-120.015000	0.0
4	1000	14	-120.015000	0.0
...
3960055	11999	10463	61.672218	0.0
3960056	11999	10464	79.323372	0.0
3960057	11999	10487	45.118361	0.0
3960058	11999	10506	75.736356	0.0
3960059	11999	10519	-29.625969	0.0

3960060 rows × 5 columns

The largest DataFrame, which contains all the node displacements for each mode between 25 and 100, can be found in the “defs” key:

all_data[fam]['defs']																				
dp_no	node_no	mode25	mode26	mode27	mode28	mode29	mode30	mode31	mode32	...	mode91	mode92	mode93	mode94	mode95	mode96	mode97	mode98	mode99	mode100
0	1000	1	0.134142	0.216114	0.193092	-0.297784	0.046671	0.101679	-0.038922	0.039048	-0.290045	-0.30238	-0.418750	-0.396922	0.167523	-0.29450	0.305611	0.242859	-0.131780	-0.102837
1	1000	11	0.133235	-0.224205	0.177632	-0.296075	-0.053283	-0.110080	-0.023488	0.037248	-0.259536	-0.325200	-0.413388	-0.399802	0.185901	0.24049	0.349572	0.235923	0.139851	-0.066484
2	1000	12	0.067980	-0.244073	0.067255	-0.170133	0.017068	-0.028849	-0.032300	-0.002932	0.002639	-0.088270	-0.416453	-0.475318	0.080366	0.157874	-0.031110	0.535567	-0.200790	-0.409373
3	1000	13	-0.013566	-0.263616	-0.039998	-0.027363	0.124321	0.080527	-0.013113	-0.022632	-0.169028	0.010970	0.005401	-0.042836	-0.001065	0.016994	-0.298478	0.215014	-0.454417	-0.438804
4	1000	14	-0.091604	-0.232150	-0.146628	0.080373	0.196288	0.157715	-0.005128	-0.053346	0.000723	0.090866	0.728555	0.792384	-0.042708	-0.064504	-0.377651	-0.633504	-0.411714	-0.202440
...	
3960055	11999	10463	-0.048392	0.036274	0.020404	-0.088032	0.035213	-0.071034	0.067050	-0.045100	-0.146413	-0.018394	0.147541	0.098663	-0.081684	0.124870	-0.087999	0.102367	-0.105086	0.046667
3960056	11999	10464	-0.032425	0.054994	0.007382	-0.005498	0.003142	-0.191339	-0.002823	-0.108213	-0.120918	-0.181788	0.103812	-0.004514	0.023668	0.056033	-0.066901	0.127630	-0.009349	-0.064460
3960057	11999	10487	-0.101641	-0.017423	0.007340	-0.169164	0.021515	0.072380	0.100115	-0.023856	-0.022090	0.094887	0.080802	0.128938	-0.073307	-0.007754	-0.059983	0.085003	0.046614	-0.041904
3960058	11999	10506	-0.038694	-0.091874	-0.020720	-0.017423	-0.021544	-0.079995	0.002790	-0.038876	-0.297812	-0.063521	-0.124353	-0.047298	-0.063314	0.202607	-0.001829	-0.277119	0.012272	0.220565
3960059	11999	10519	-0.129445	-0.015920	-0.101730	-0.090920	-0.029353	-0.227684	-0.004785	0.028464	-0.199744	0.196219	-0.285314	0.312668	0.087459	-0.015949	-0.062190	-0.209838	0.573700	0.157085

3960060 rows × 20 columns

2.3 Pre-processing and Feature engineering

2.3.1 Data cleaning

A data cleaning was performed at the very beginning to remove the missing values and duplicated records. Some simulations had failed during the phase of data generation and some values were invalid (among other results, the identification of the longitudinal mode).

2.3.2 Standardization

Many algorithms require that each feature be rescaled (except for tree-based models), especially if outliers are present. The Standardization function was chosen because the geometric parameters were selected in predefined ranges using Latin Hypercube Sampling.

Evaluation technical report

The `StandardScaler()` transformer utility was used to standardize the data. It will be integrated into the model pipelines. The statistics of each feature are computed on the training set, and then the features are centred and scaled (zero mean and unit variance).

2.3.3 Additional features

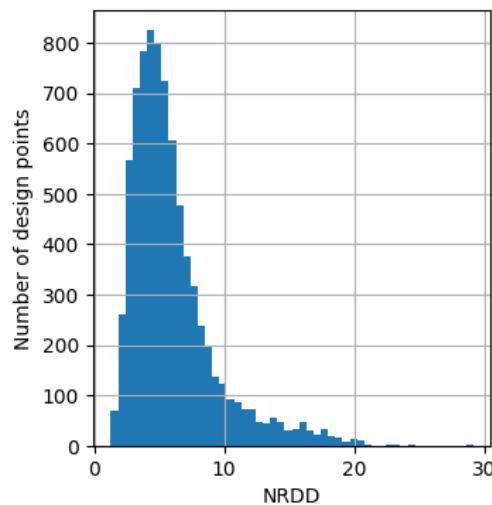
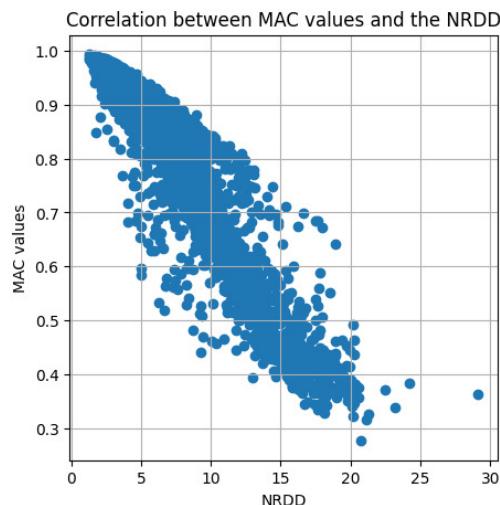
2.3.3.1 Slot class

The number of slots was selected from [2, 3, 4] in the X dimension and [1, 2] in the Z dimension. They are forming design classes (“slot classes” designated as “`nb_slots_x - nb_slots_z`”):

	<code>nb_slots_x</code>	<code>nb_slots_z</code>
1	4	2
2001	3	2
4004	4	1
6079	2	2

2.3.3.2 Uniformity of the longitudinal mode’s displacements

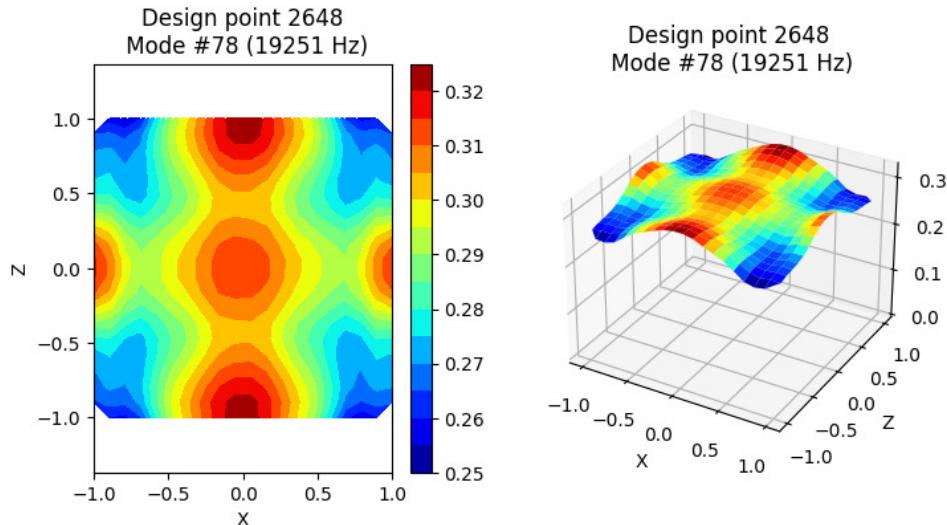
Besides the longitudinal mode’s MAC value, a new measure called the “Norm of the Relative Displacement Deviation” (NRDD) is proposed to assess the uniformity of its displacement. This metric is obtained by taking the average of the output surface’s displacement and the relative difference of each node’s deviation from this mean value. The deviations are squared, summed up for all nodes. The root value gives the deviation criterion: a low value shows a great displacement uniformity at the output surface (a perfect uniform mode shape would have zero deviation). That criterion will be one additional target for the models, since we want to predict the uniformity without calculating explicitly the longitudinal mode’s displacements.



The left graph highlights the correlation between the introduced metric NRDD and the MAC value. The well-identified longitudinal modes with almost ideal output displacements achieve high values of MAC (around 1). The design points with lower MAC values may have more candidates for the longitudinal mode, which can result in poor uniformity of the output surface.

The right graph represents the distribution of the NRDD. Most values are clustered around 5.0. The worst uniformity is found for NRDD = 29.1, while the most uniform displacement field is reached for NRDD = 1.29 (as indicated below the DP 2648 plotted surfaces).

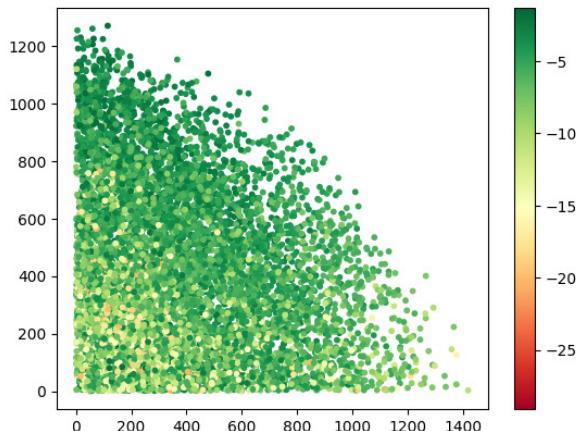
Evaluation technical report



2.3.3.3 Frequencies above / below

We will see later that the gaps between the frequencies matter. The graph, showing the frequency gaps above and below the longitudinal mode, reveals that none of the design points has a longitudinal mode isolated on both sides with more than 800 Hz. The information is crucial for the sonotrode design engineer when he is optimizing the frequency isolation of the longitudinal mode.

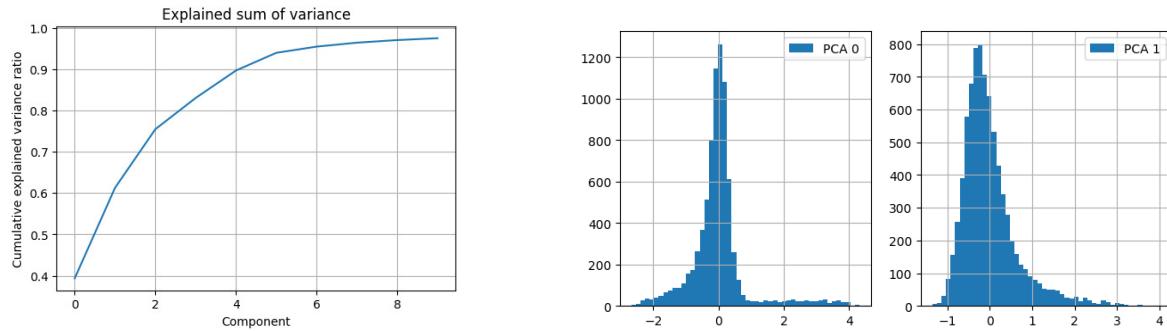
The color information carried by the NRDD shows the trend of poor uniformities when the frequency gaps are smaller (we will get on that later in the model interpretation chapters).



2.3.4 Dimension reduction

A Principal Component Analysis (PCA) was conducted in the field of the longitudinal mode's displacements. The aim of the procedure is to reduce the number of variables (here, $20 \times 20 = 400$ displacement values). To do so, the data was mapped into a ten-dimensional vector space. In general, the PCA applies to a set of explanatory variables. However, in this particular case, the displacements are part of the target values.

Evaluation technical report

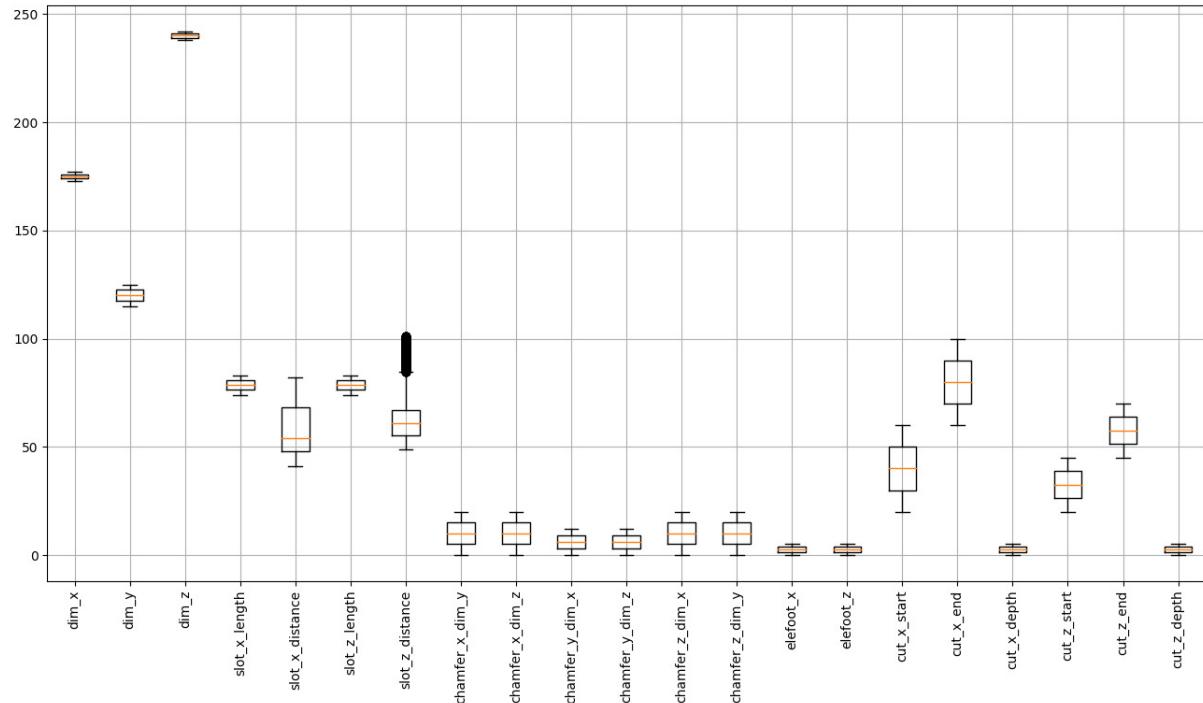


The cumulated sum of the explained variance on the left diagram shows that a reduction to 10 dimensions keeps more than 95% of the information. We will attempt to predict the first two components, whose distributions are illustrated in the charts on the right.

2.4 Visualization

2.4.1 Distribution of the explanatory variables

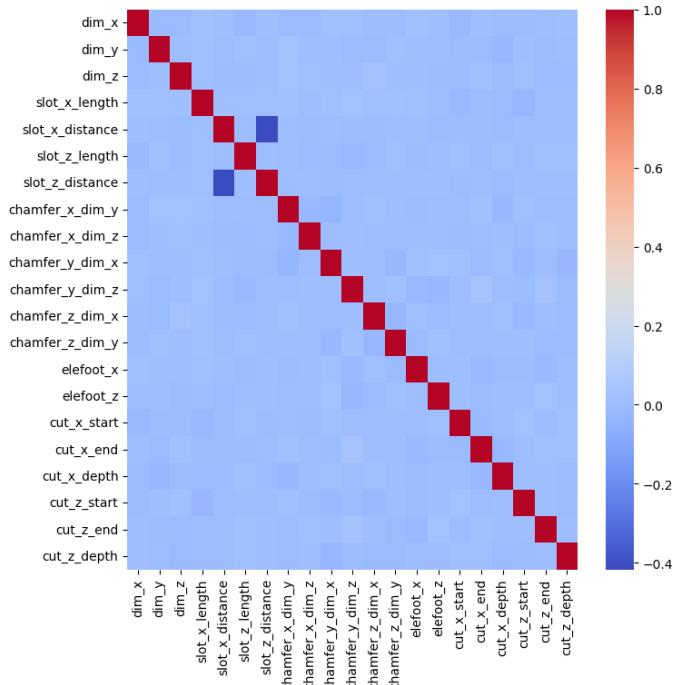
The explanatory variables are the 23 parameters that define the geometry of the sonotrode. The box plots show that most of these variables are normally distributed and centred around zero, which makes sense given that the design points were generated by Latin Hypercube Sampling. Only the variables “slot_x_distance” and “slot_z_distance” seem to have a different distribution. We will see later that these values depend on the number of slots.



The welding surface dimensions (175×240 mm) may vary by up to ± 2 mm. In the design phase, the design engineer can adjust the sonotrode slightly in width or thickness to obtain better results, while ensuring full coverage of the weld and avoiding contact with other components.

Evaluation technical report

2.4.2 Correlation between the variables

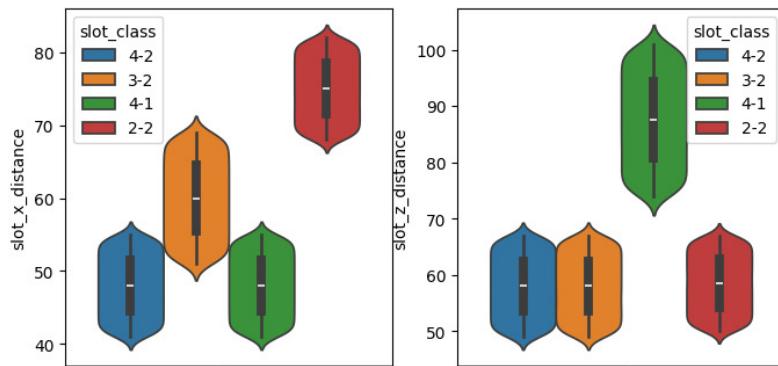


(Correlation heatmap between the explanatory variables)

All the variables are completely independent due to the Latin-Hypercube parameters' sampling. Only the slot distances are correlated to each other: the reason is that their nominal values are calculated directly from the number of slots (before applying the variation of $\pm 15\%$):

$$\text{slot_x_distance} = \frac{\text{dim}_z}{\text{nb_slots_x}+1}$$

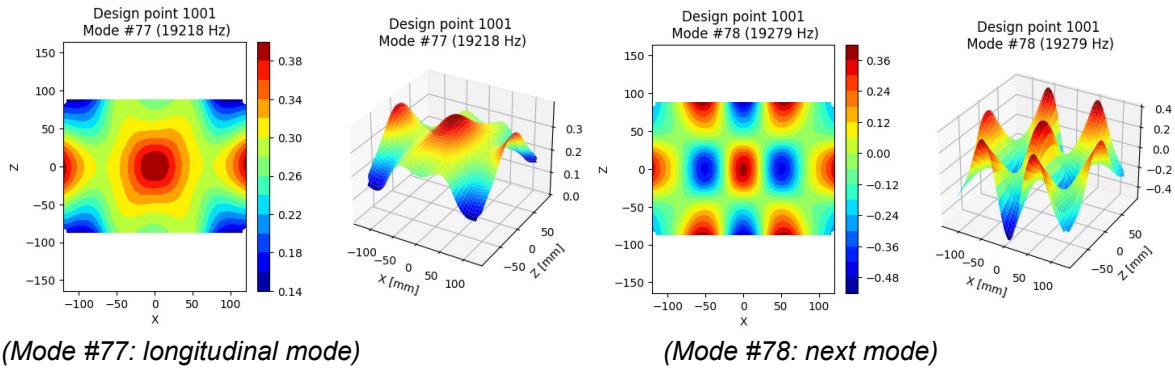
$$\text{slot_z_distance} = \frac{\text{dim}_x}{\text{nb_slots_z}+1}$$



2.4.3 Visualization of the mode shapes

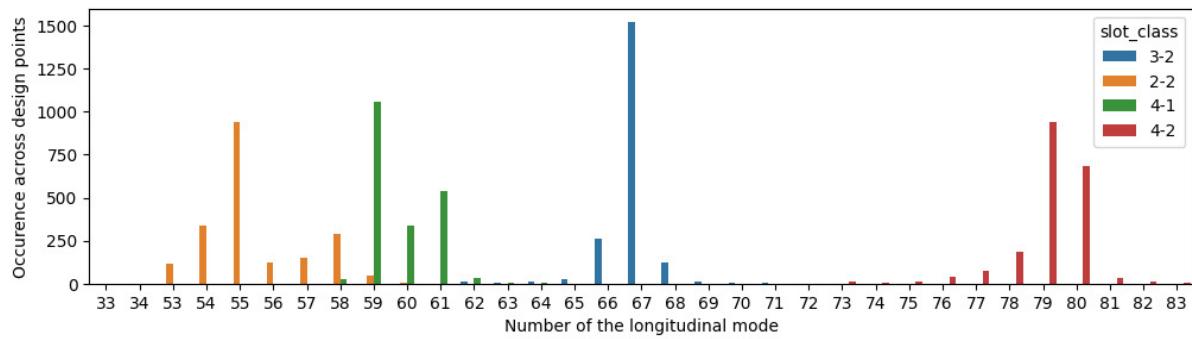
The displacements of the nodes located on the output surface can be visualized by looking at their X and Z coordinates. The following two graphs show the displacements of mode #77 (identified as the longitudinal mode) and of the next mode #78:

Evaluation technical report



2.4.4 Distribution of the identified longitudinal modes

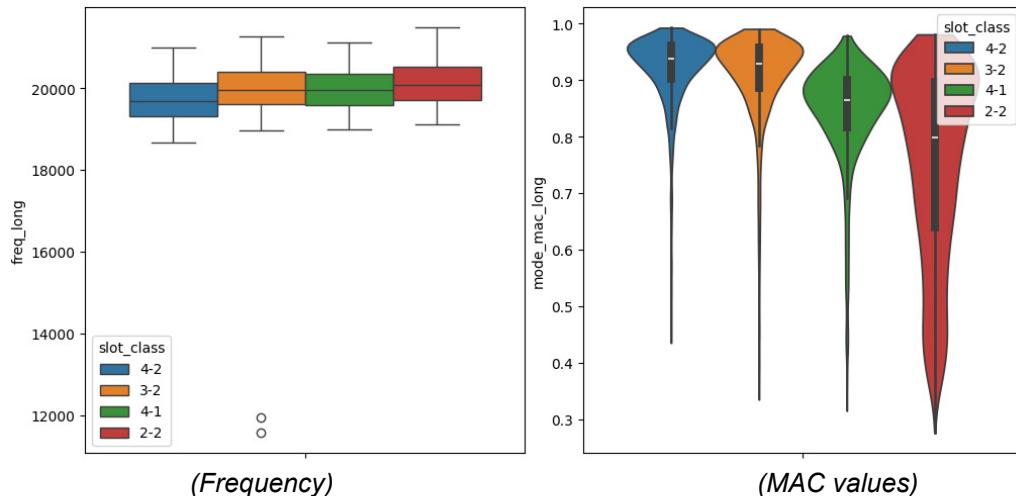
As previously mentioned, the longitudinal mode was directly identified at the end of the FEM simulation by comparing the MAC values between all the calculated modes.



The number of the longitudinal mode is located between 50 and 85 and it strongly depends on the slot class. Adding of slots to the sonotrode increases the complexity of the geometry, which leads to an increase in the number of vibration modes for a given frequency range (called “modal density” see [2]).

2.4.5 Distribution of the target variables

Frequency and MAC value of the longitudinal modes

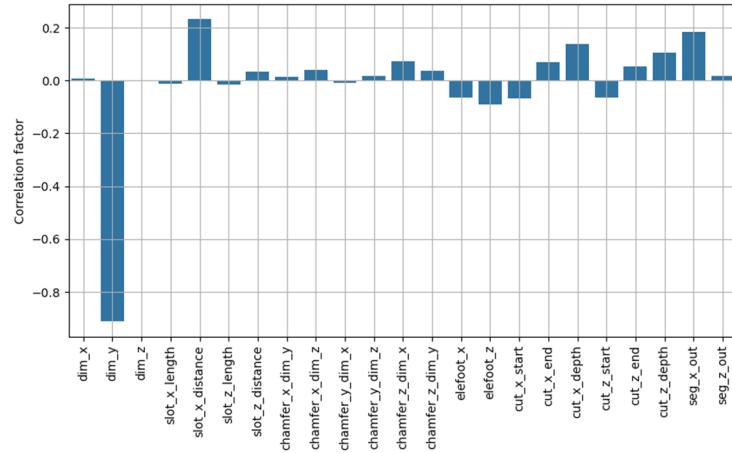


Evaluation technical report

The boxplot of the frequencies highlights two outliers around 12 kHz that will be excluded from the dataset. The MAC values of the longitudinal mode are higher for the slot classes “4-2” and “3-2”. For sonotrodes with dimensions of about 175×240 mm, it is beneficial to choose 2 slots across the thickness and 3 (or 4) slots across the width to ensure even displacement at the output surface.

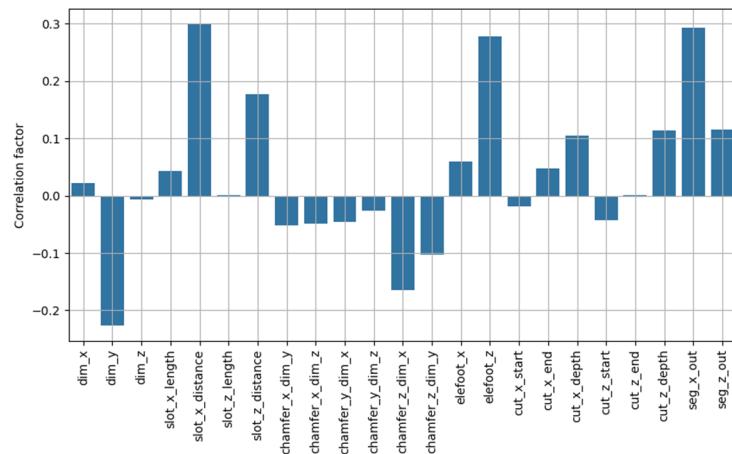
2.5 Relationships between explanatory variables and target variables

2.5.1 Frequency of the longitudinal mode



As expected the longest-influencing factor is undoubtedly the sonotrode length (`dim_y`), which is typically adjusted to match the driving frequency of 20 kHz. The other influential variables are `nb_slots_x`, `slot_x_distance`, and `slot_z_distance`.

2.5.2 Normalized Relative Displacement Deviation



The influences are more evenly dispersed across the parameters than for the longitudinal frequency. Many parameters, such as `slot_x_distance`, `elefoot_z`, `seg_x_out`, and `dim_y`, can be used to adjust displacement uniformity. These influences are crucial and very helpful for the acoustic design engineer. They show the different leverage points that achieve more uniform output displacements.

Evaluation technical report

3 Modelling

3.1 Classification of the problem

Supervised learning uses labelled training data to learn a mapping between input and output variables. This allows for the creation of precise output when given new input values. Our case of supervised learning is a regression problem. It involves predicting numerical values that fall within a continuous range, such as sales figures or housing prices. It is a common technique used to establish a relationship between an input variable (X) and an output variable (y).

The most commonly used metrics to evaluate the quality of predictions are: Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) or the R^2 score function (coefficient of determination).

The R^2 score was chosen as the primary metric for evaluating the model because it is the default scoring metric for both LightGBM and HistGradientBoosting regressors in Scikit-learn. Moreover, R^2 is a universal and absolute metric that does not require any reference point to validate the model's accuracy. If the R^2 score is greater than 0.75, the model is considered good [5]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

The numerator is the sum of squared residuals, and the denominator is a data-dependent constant value. R^2 is a dimensionless measure, with a best possible score of 1.0. A zero score indicates that the model does not outperform a simple average.

In the case of multiple target variables (like predicting several frequencies), the wrapper *MultiOutputRegressor* will be used to build a single regressor. The score method's “*multioutput*” parameter, which is set to default (“*uniform_average*”), weights all the target variables equally when calculating the R^2 score.

3.1.1 Target variables

As previously mentioned, the objective is divided into three different problems, but the explanatory variables remain identical (sonotrode geometry). Only the target variables will differ from each other:

- Variable `freq_long`: Longitudinal mode frequency
- Variables `freq_50`, `freq_51`, ..., `freq_85`: All frequencies

For the prediction of displacements, many target variables were tested to compare the best estimators:

- The variable `disp_long_pca_0` corresponds to the first component of the reduced set of displacements
- Variable `mode_mac_long`: MAC value of the longitudinal mode
- Variable `long_disp_nrdd`: NRDD criterion.

The transformation of the target variable before training a regression and using it for prediction can be of benefit for non-linear problems. It increases the R^2 score. Many transformations have been tested on the `long_disp_nrdd` variable to compare their effect on the accuracy.

Evaluation technical report

3.2 Model choice & Optimization

3.2.1 Model selection with Scikit-Learn

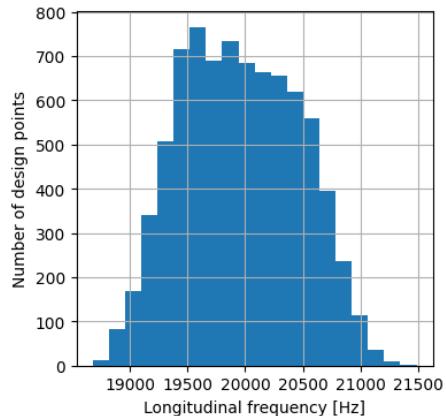
This section details the method for choosing algorithms. The approach remained consistent across all three objectives:

- First, scale the explanatory variables.
- Choose the target variable to predict. Display its distribution.
- Split the dataset into 80% training and 20% testing subsets.
- Use the LazyPredict and FLAML packages to get a first impression of the most appropriate estimators.
- Perform a cross-validation (CV) test on the best candidates to confirm the results of LazyPredict and FLAML

The CV evaluates the estimators again by taking a random part of the training data as a validation set. It then computes a performance measure. The process is repeated 5 times with different splits, and the average and standard deviation of the scores are used as performance criteria.

3.2.1.1 Frequency of the longitudinal mode

The distribution of the target variable `freq_long` is almost symmetrical and nearly centred around 20 kHz.



Model	Mean of score R ²	Standard dev. of R ²
LightGBMRegressor	0.985	6.14 10 ⁻⁴
HistGradientBoostingRegressor	0.985	3.32 10 ⁻⁴

Both models provide good predictions with standard parameters, according to the mean score although the HistGradientBoosting regressor has a more reproducible score. Gradient boosting algorithms are particularly suitable for handling complex problems and large datasets. They can identify patterns and relationships. We have decided to retain both models for the next optimization phase because they can predict the frequency with comparable precision.

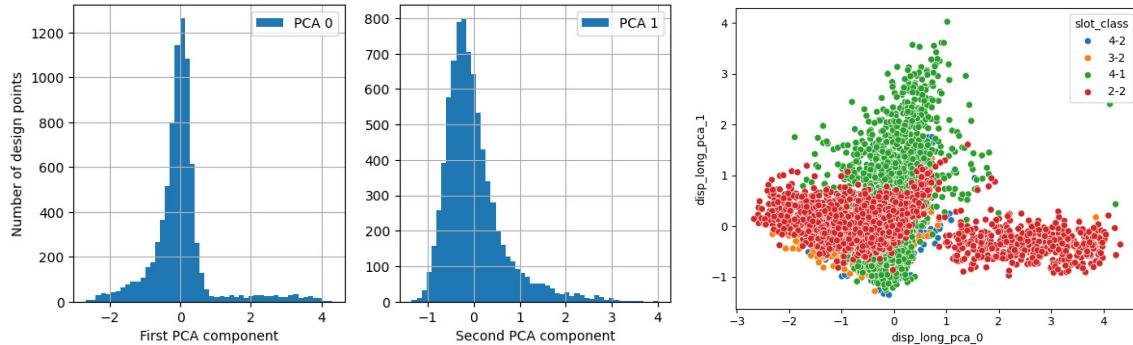
3.2.1.2 Displacements of the longitudinal mode

We will present the most useful results from the model benchmark, which was run with different target variables and required intensive computations.

Evaluation technical report

3.2.1.2.1 First PCA component as target variable (`disp_long_pca_0`)

The first two PCA reduced displacements do not seem to have any discernible pattern:

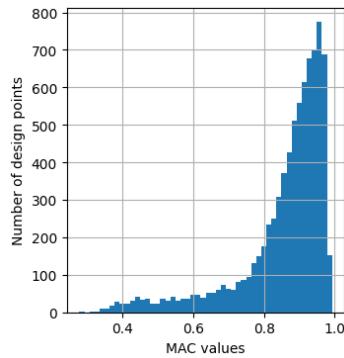


Model	Mean of score R ²	Standard dev. of R ²
ExtraTreesRegressor	0.273	0.018
LightGBMRegressor	0.218	0.036

So far, efforts to directly predict the subset of displacements have not yielded satisfactory results. As an alternative, it was decided to predict either the uniformity criterion NRDD or its transformed versions.

3.2.1.2.2 MAC value as target variable (`mode_mac_long`)

The mode (most frequent value) of the histogram of the MAC values is located around 0.95.



Model	Mean of score R ²	Standard dev. of R ²
ExtraTreesRegressor	0.580	0.020
LGBMRegressor	0.693	0.024
XGBRegressor	0.540	0.028

The R² score is higher than for the prediction of the first PCA component.

Evaluation technical report

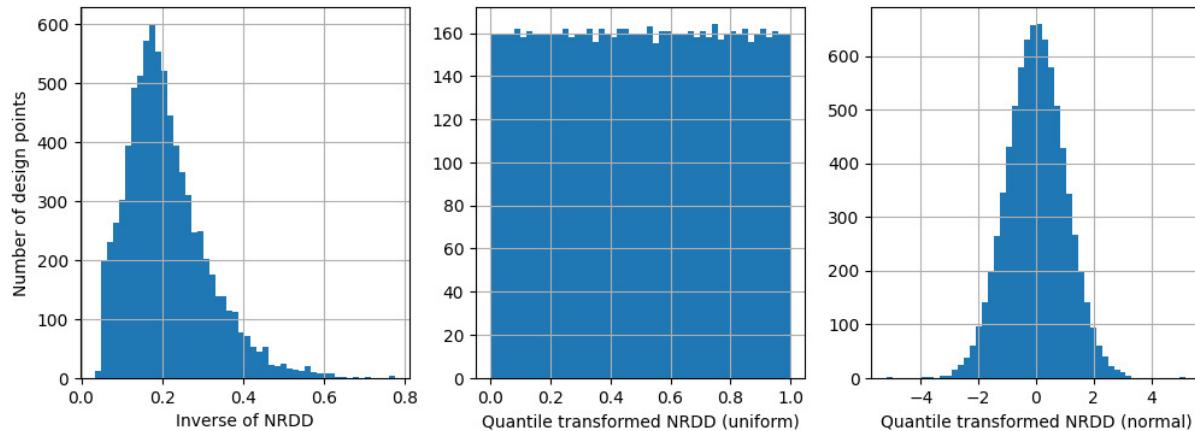
3.2.1.2.3 NRDD as target variable (long_disp_nrdd)

	Mean of score R ²	Standard dev. of R ²
MLPRegressor	0.676	0.011
XGBoostRegressor	0.604	0.027

Implementing a transformation on the target variable NRDD proved extremely promising. Below is an overview of the cross-validation results obtained by using different transformers before fitting the LightGradientBoosting regressor:

Transformer	Mean of score R ²	Standard dev. of R ²
Inverse	0.73	0.01
Logarithm	0.73	0.02
PowerTransformer	0.74	0.02
QuantileTransformer	0.76	0.02

The QuantileTransformer uses a technique for transforming features to follow a uniform (or Gaussian) distribution. The quantile function is the inverse of the cumulative probability distribution function (CDF). By changing the probability distribution, this transformation reduces the impact of outliers and facilitates the identification of relevant patterns in the data.



The prediction of the Normalized Relative Displacement Deviation shows a great potential for optimization. We decided to use this criterion for the following reasons:

- The NRDD is a single value to predict,
- The requirement of the output displacements is to ensure an even distribution of these displacements during the tuning phase of the sonotrode design. Predicting all displacements would have been ideal but unfortunately also time consuming.

We retain three different models to predict the displacement uniformities. So far, they all have been trained using standard parameters.

Target variable	Model
NRDD	MLPRegressor
Inverse of NRDD	LightGBMRegressor
QuantileTransformer of NRDD	LightGBMRegressor

Evaluation technical report

Instead of investigating more estimators and other possible transformations, we decide to optimize the hyperparameters of the identified models to increase their performance.

3.2.2 Model optimization

The hyperparameters are the external configuration variables that manage the training of machine learning models. To find the optimal combination of these and improve the model's performance, a separate process of hyperparameter tuning is necessary. The common process is to select the hyperparameters to be tuned and specify a grid of acceptable ranges that would generate the acceptable values.

The most prevalent techniques are Grid search, Random search and the Bayesian hyperparameter optimization. The latter uses an efficient approach by taking into account the previous evaluations and was used to optimize the Scikit-Learn hyperparameter's algorithms.

After the optimization, the final scores are computed with the test data. This step allows evaluating how the models are performing on unseen data that were not used to train the model. Then the predicted target values will be compared to the true values.

3.2.2.1 Frequency of the longitudinal mode

The HistGradientBoosting regressor is a histogram-based Gradient Boosting regression tree, that optimizes an arbitrary loss function. Decreasing the learning rate and increasing the maximum number of iterations improved its performance.

(Parameters of the HistGradientBoosting regressor)

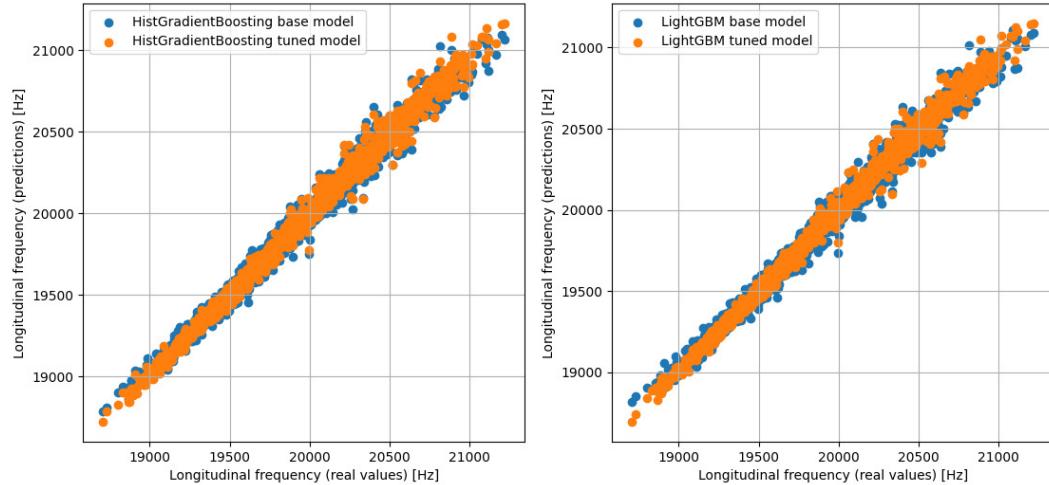
	Default	Tuned
l2_regularization	0	$8.63 \cdot 10^{-3}$
learning_rate	0.1	$5.76 \cdot 10^{-2}$
loss	squared error	squared error
max_depth	None	4
max_iter	100	4906
min_samples_leaf	20	150

(Parameters of the LightGBM regressor)

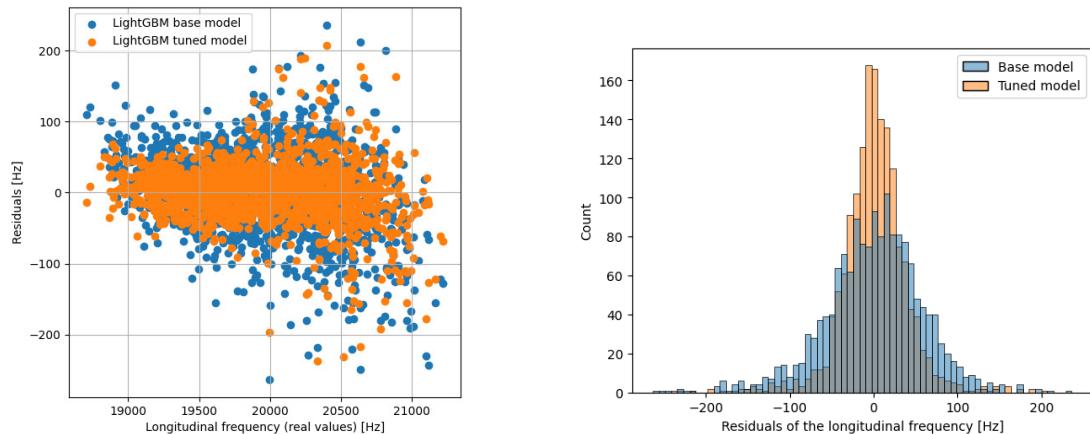
	Default	Tuned
min_child_samples	20	20
learning_rate	0.1	$4.63 \cdot 10^{-2}$
n_estimators	100	15,000
num_leaves	31	5
reg_alpha	0	1.0
reg_lambda	0	1.0

Evaluation technical report

The comparison between the real values and the predictions confirms the high performance of both HistGradientBoosting & LightGBM regressors with tuned hyperparameters:



Representation and distribution of the residuals



The final score shows that both models are equivalent in terms of performance:

	Final score R ²
HistGradientBoosting regressor	0.992
LightGBM regressor	0.993

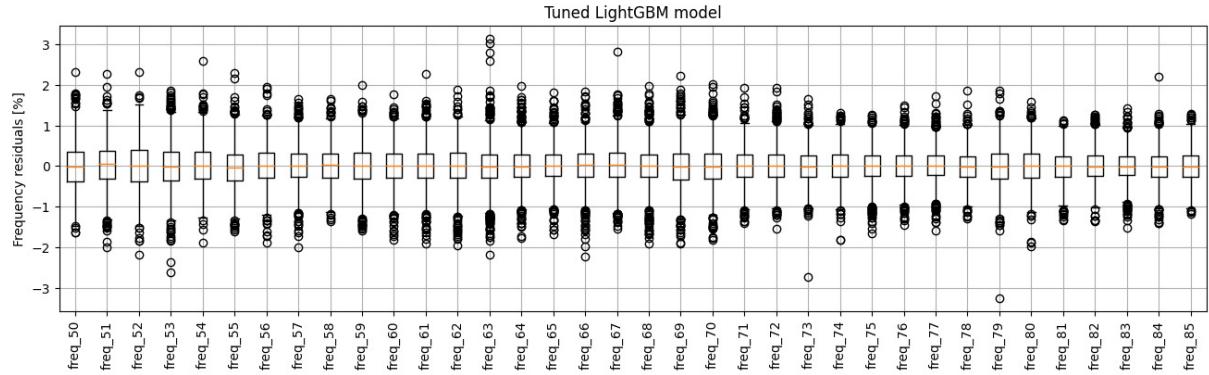
3.2.2.2 All frequencies

We did not compare the different models for the prediction of all frequencies. We assume that the models, that performed well for the longitudinal frequency will have similar performance on other frequencies.

The hyperparameters tuning was performed on the LightGBM start model alone. The MultiOutputRegressor wrapper will replace the step containing the estimator and will calculate the R² score for each frequency equally.

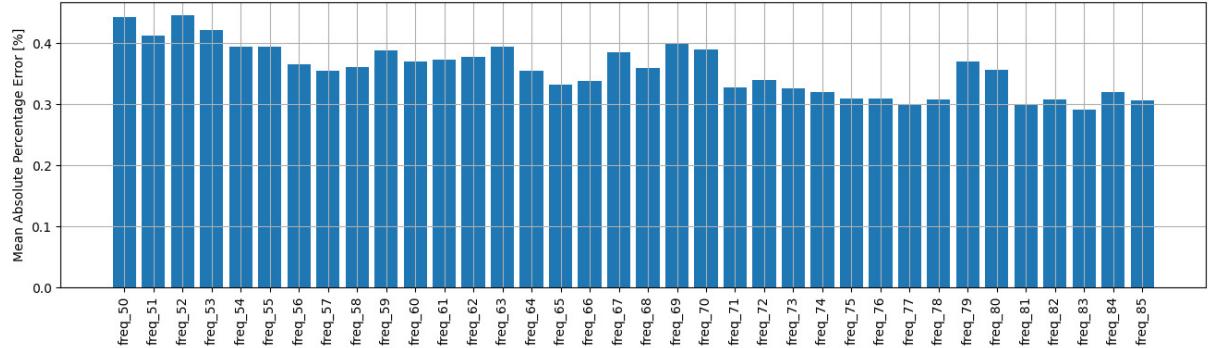
Evaluation technical report

Instead of representing true and predicted values, we display the relative differences (residuals) between the predictions and the actual values for each frequency. This provides a better overview:



The residuals relative to the frequencies are all below 1.5% (outliers excluded), and the final score of 0.9976 is very impressive in terms of the coefficient of determination.

The Mean Absolute Percentage Error (MAPE) measures the typical deviation, expressed as a proportion, between predicted and actual figures. This statistic is particularly suitable for analyzing targets with varying frequency levels, such as those ranging from 13 to 27 kHz.



The following graph displays the Mean Absolute Errors (MAE) for each slot class. The position of the longitudinal mode is highly dependent on the slot class, and the frequencies around this mode receive particular attention from the sonotrode designer due to potential interactions.



The MAE increases with the slot classes, and therefore with the complexity of the geometry. A mean error of 100 Hz could be at the limit of a good prediction in the case of a very close neighbor mode.

Evaluation technical report

3.2.2.3 Uniformity of the output displacements

By examining the various tuned estimators and transformations of the target variable, we observe that the LightGBM regressor, which predicts Quantile-transformed NRDD values, performs slightly better. However, we can notice in the last graph that some bad predictions sticking around 1.3 for real NRDD values below 5.0. This suggests that the R² metric may not be suitable for this optimization.

Regressor	Target	Predictions vs Real values	Final score R ²
MLP	NRDD		0.758
LightGBM	1 / NRDD		0.754
LightGBM	Quantile transformed NRDD		0.782

Evaluation technical report

3.2.3 Dense Neural Network with Keras

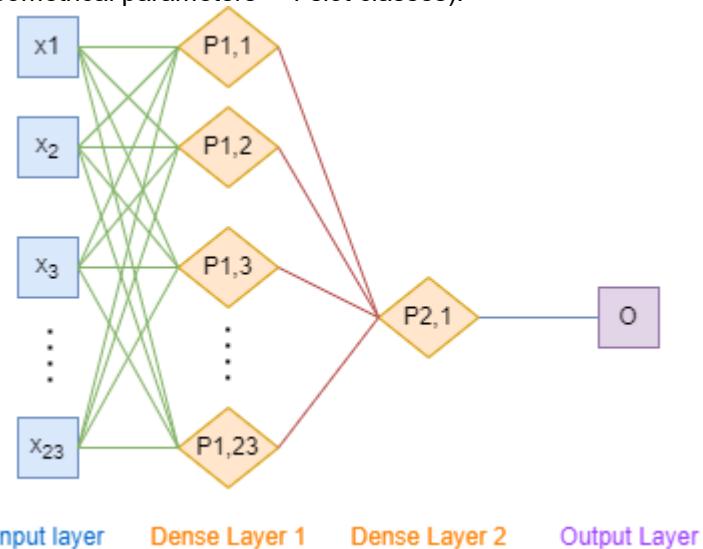
Since the modelling of the output uniformity did not yield satisfactory results, we decided to try another method using deep learning. This technique uses a layered structure, based on artificial neural networks.

In this section, we explore a sequential neural network model with two layers to predict the uniformities of output displacements. The steps are as follows:

- 1) Create a Keras baseline model
- 2) Identify the optimal transformation of the target variable.
- 3) Tune the neural network's topology
- 4) Visualize the training process history

3.2.3.1 Baseline model

The first layer has as much neurons as the number of explanatory variables (27 variables: 23 geometrical parameters + 4 slot classes).



The Adam optimization algorithm is used and a mean squared error (MSE) loss function is optimized. The weight initialization is identical for both layers and corresponds to a normal distribution. The model uses an activation function (controls the non-linearity) for the first hidden layer only, since the output layer should be giving the predicted numerical value directly without transformation. As before, we will use the same metric R^2 to evaluate the model's performance.

3.2.3.2 Transformation of the NRDD

As previously made, we will apply different transformations to the NRDD values and the cross-validation scoring will highlight the most appropriate transformation for the target variable.

Transformation	Mean score R^2	Standard deviation R^2
(No transformation)	0.569	0.060
Inverse	0.036	0.194
Logarithm	0.450	0.029
PowerTransformer	0.592	0.036
QuantileTransformer	0.501	0.029
Root square	0.509	0.025

Evaluation technical report

The PowerTransformer provides the best mean score of cross-validation for this specific model and is selected for the next optimization steps. Like the QuantileTransformer, this transformation converts numerical data into a Gaussian distribution. It is particularly useful when data has a highly skewed distribution.

3.2.3.3 Tuning of the neural network topology

The structure of a neural network model is the biggest lever for improving its performance (number of layers and neurons, activation functions, etc.). However, many parameters must be set, and training can be very slow.

We decided on the GridSearch technique, which exhaustively tries all possible combinations of the selected set of parameters. To limit the number of evaluations, we divide the optimization into multiple tuning steps. Each step uses the best parameters from the previous one.

Tuning step	Best parameters	Score R ²
Batch size and number of epochs	Batch size: 1000 Number of epochs: 100	0.806
Learning rate of the Adam optimizer	Learning rate: 4.3 10 ⁻²	0.811
Number of neurons of the hidden layer	N: 25	0.817
Weight initialization and activation function	Activation: "Softmax" Weight initialization: "normal"	0.822

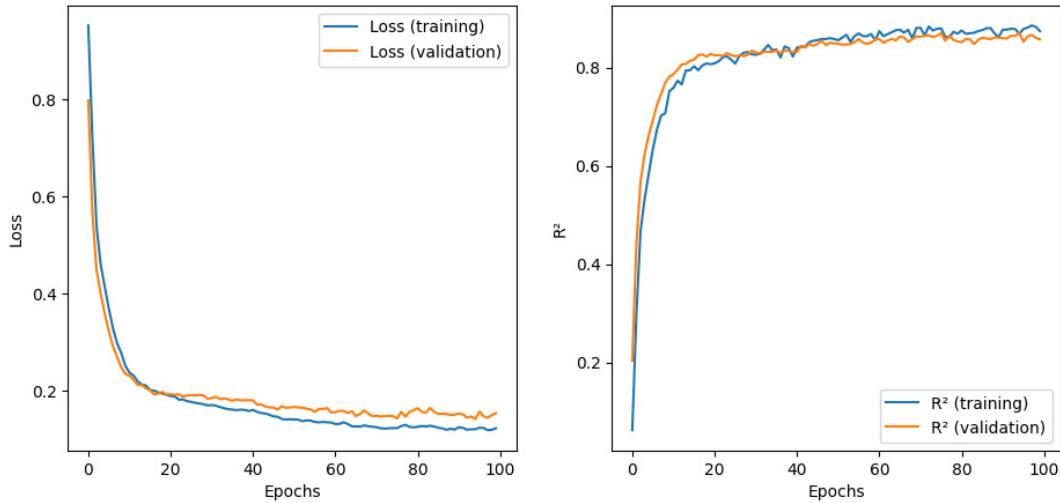
```
Model: "tuned_model"
=====
Layer (type)          Output Shape         Param #
=====
Dense_Layer_1 (Dense)    (None, 25)           700
Dense_Layer_2 (Dense)    (None, 1)            26
=====
Total params: 726 (2.84 KB)
Trainable params: 726 (2.84 KB)
Non-trainable params: 0 (0.00 Byte)
```

With a batch size of 1000, 100 epochs, 25 neurons in the hidden layer and the Softmax activation function, we obtained the best R² score of around 0.82.

3.2.3.4 History

The history contains information on the training process, including the evolution of the R² score and the loss function for both the training and validation sets.

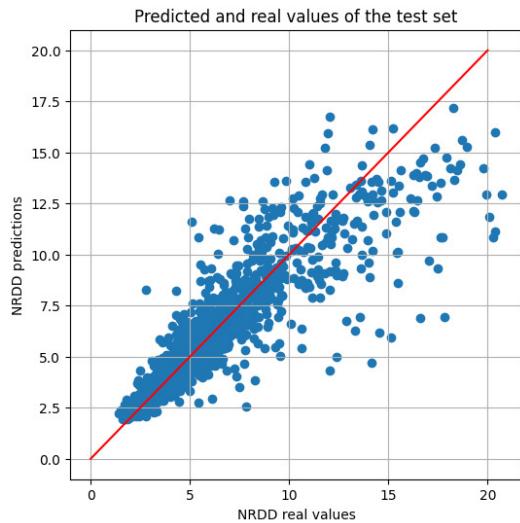
Evaluation technical report



The value of the loss function and of the evaluation metric remains stable throughout the training. The R^2 score is not significantly lower in the validation set than in the training set, indicating a slight overfitting. However, the training process can be considered a success.

3.2.3.5 Predictions on the test set

The following graph shows a comparison of predictions to actual values. The neural network model outperforms the previous machine learning models in terms of accuracy. Notably, there are two outliers in the NRDD values below 5.0. However, the predictions for higher NRDD values exhibit a larger deviation, a topic we will explore in the next section on model interpretations.



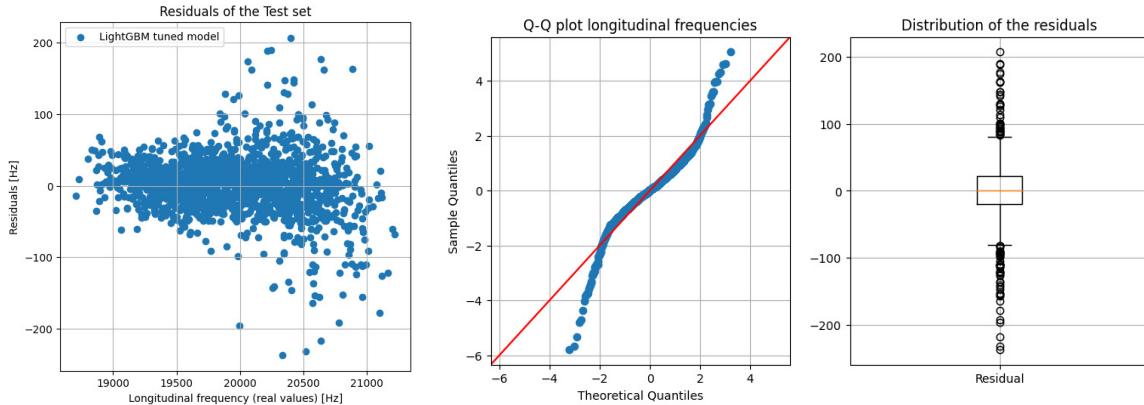
3.3 Interpretation of the results

In this section, we will determine the residuals (difference between the true target values and the predicted values) and visualize them.

Evaluation technical report

3.3.1 Prediction of the longitudinal mode frequency

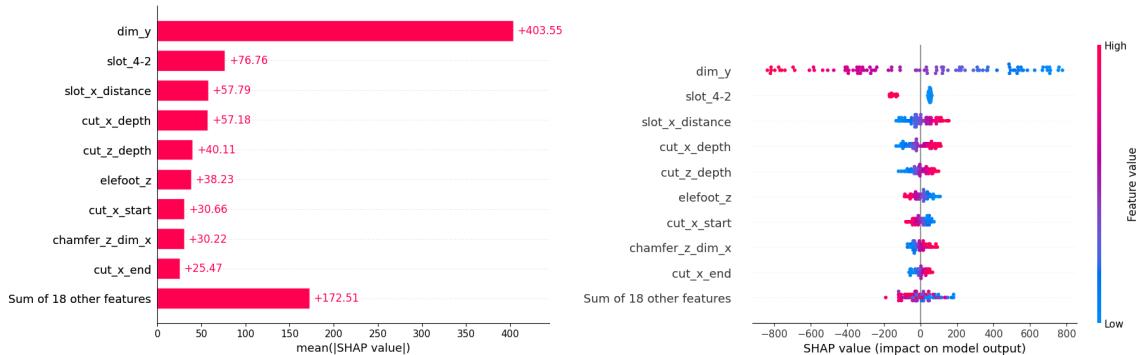
We chose the equally valuable HistGradientBoosting and LightGBM regressors for the longitudinal frequency predictions. We will only analyse the distribution of residuals calculated with the LightGBM model.



The maximum absolute error is around 240 Hz. The Q-Q plot compares the distribution of the residuals with a theoretical normal distribution: the residuals follow a normal law inside the first two quantiles only. Furthermore, several outliers are indicated by the boxplot.

3.3.2 SHAP values

The SHAP (SHapley Additive exPlanations) values provide an objective explanation of how each feature affects the model's predictions. Like in a game, they measure the contribution of each player to the final outcome.



From the bar plot (left) we can see the biggest impact of the oscillation length (`dim_y`). The beeswarm plot on the right also visually summarizes the importance of features across multiple predictions. However, the vertical spread at each feature provides information about the density of data points:

- As the length value increases, its SHAP value decreases, as expected.
- The second most important feature is the slot distribution "4-2" (with values of 0 or 1). If the class is active, the longitudinal frequency is about 250 Hz lower. The influence is the greatest for that slot distribution because the total number of slots is the highest (6). More mass is removed in the nodal plane and the stiffness is decreasing, causing a frequency drop.
- The chamfers' widths and cut depths raise the frequencies because some mass is removed at the antinodes (points with maximum displacements).

Evaluation technical report

3.3.3 Focus on the best and worst predictions

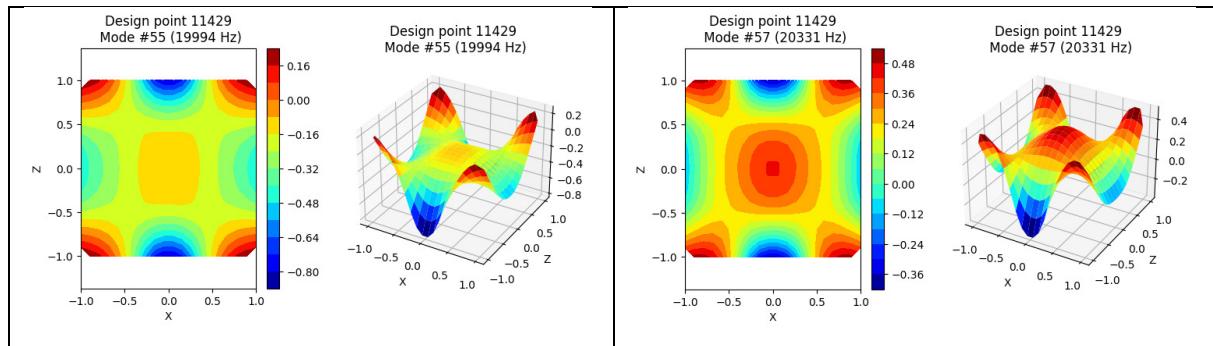
We will take a particular look on the best and worst predictions of the LightGBM regressor.

	Worst prediction	Best prediction
Residual	237 Hz	0.02 Hz
Predicted frequency	20094 Hz	19492.73 Hz
True frequency	20331 Hz	19492.75 Hz
Design point	DP 11429	DP 1189

The design point 11429 is particularly interesting because the predicted frequency is close to the nominal frequency (20 kHz) and the measured value is higher (20331 kHz). In order to understand why the prediction fails for that design point, we look at the different modes (neighbor modes) around the longitudinal mode:

dp_no	mode_no	freq	mode_mac
564025	11429	54	19321.948
564026	11429	55	19993.723
564027	11429	56	20109.459
564028	11429	57	20330.910
564029	11429	58	20736.378
564030	11429	59	20895.412

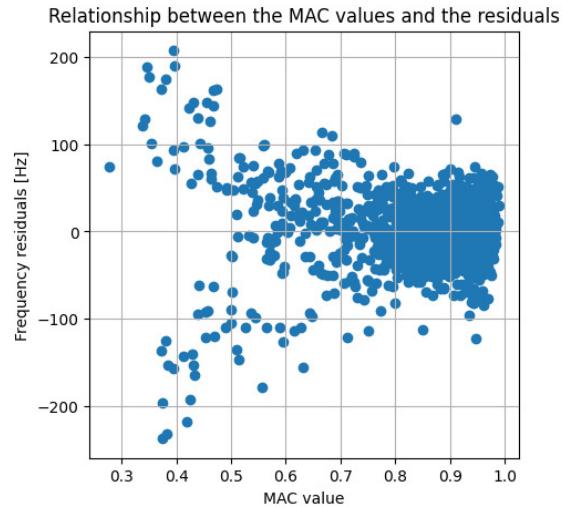
We observe that the MAC values of modes 55 and 57 are both significantly below the maximum value of 1.0, and they are quite similar. This suggests that these two modes may be overlapping in their identification of the longitudinal mode. In this scenario, the longitudinal mode is ambiguous and experiences degeneration [3]:



The deformed shapes show that both modes 55 and 57 have significant longitudinal component. If mode 55 had been selected as the longitudinal mode, the residual would have been much lower.

The visualization of the residuals as a function of the MAC values confirms the fact, that the residuals are smaller for the design points with well-identified longitudinal modes (with high values of MAC values):

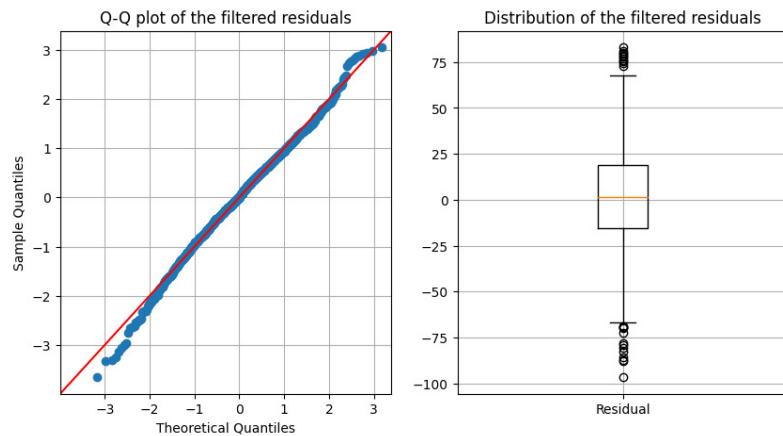
Evaluation technical report



3.3.4 Filtering of the data set

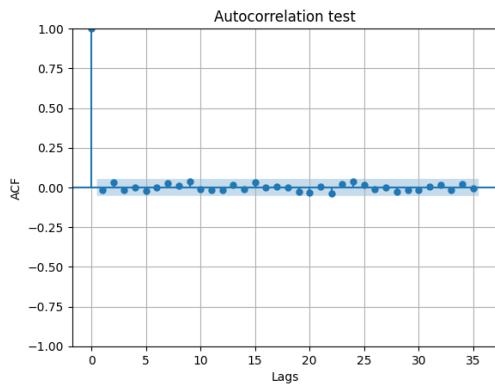
According to the previous results, the residuals are increasing drastically for designs with poor MAC values. By filtering these points, the model can focus on designs with clearly identified modes, providing more accurate predictions. By eliminating all design points with MAC values below 0.75, we see that 85% of the data would be kept for the modelling.

With the filtered dataset, the distribution of the residuals is now normally distributed and still has a mean of zero. The mean absolute error (MAE: arithmetic average of the absolute errors) is 21 Hz, which is an excellent value.



The autocorrelation test-analyzes the residuals using the ACF (autocorrelation function). All autocorrelation points fall within the 5% confidence interval. Therefore, we can conclude that the residuals are not correlated, normally distributed, and centred on zero: the white noise assumption can be considered as validated.

Evaluation technical report



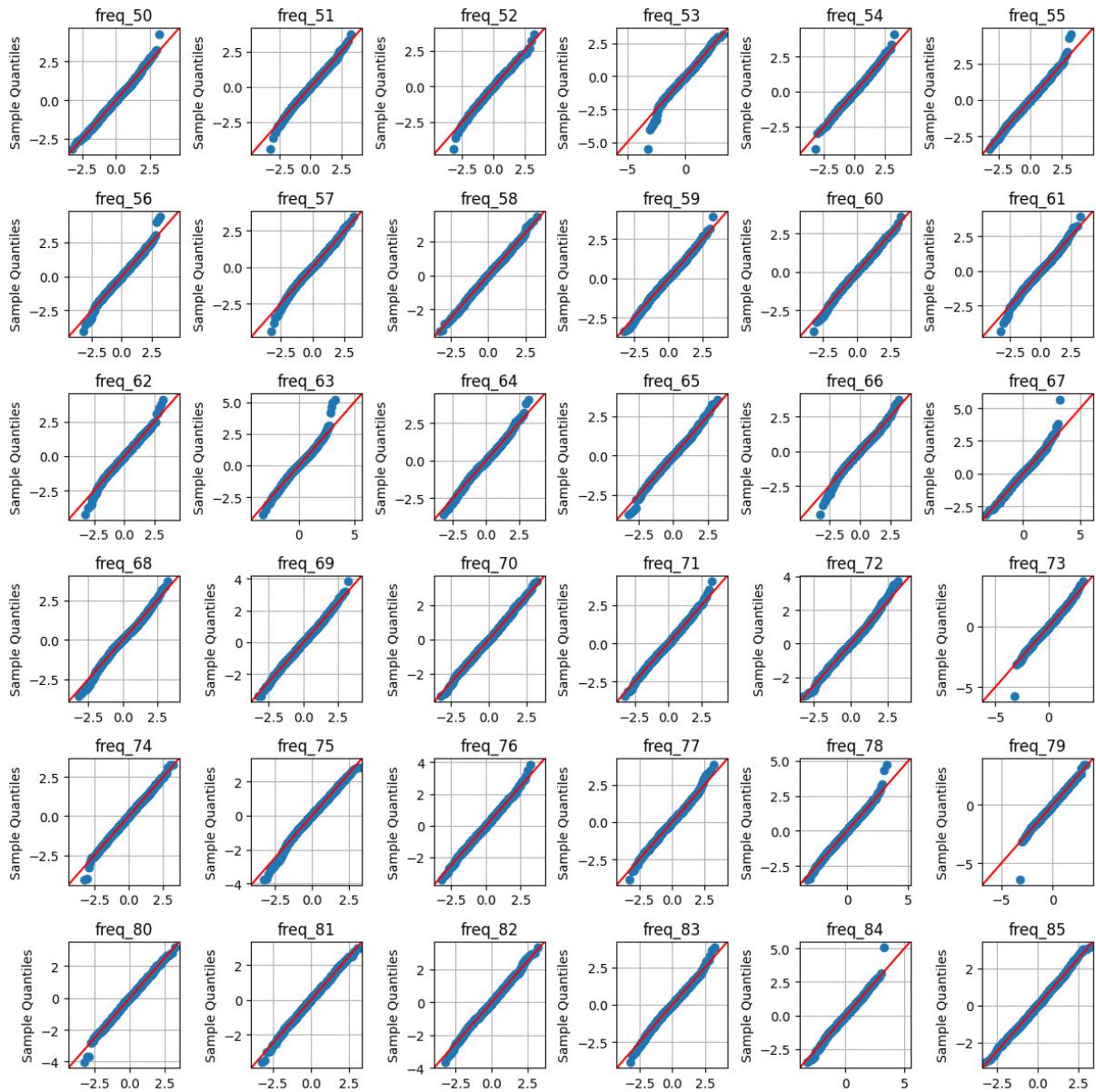
The process of filtering the data set before training the model significantly enhances its predictions. It is assumed that the residuals consist of independent, random values.

3.3.5 Prediction of all frequencies

In the previous section, we observed that the mean absolute errors for all frequencies ranged from 50 to 100 Hz, varying by slot class. The distribution of residuals across all frequencies is more complex than that for a single frequency. Nevertheless, the following multivariate plot suggests that most frequencies and design points follow a normal, zero-centred distribution.

This study does not delve into the causes of the outliers due to a lack of time. However, it is hypothesized that they may be the result of interactions between the modes, based on the longitudinal frequency prediction findings.

Evaluation technical report



3.3.6 Prediction of the longitudinal mode's displacements

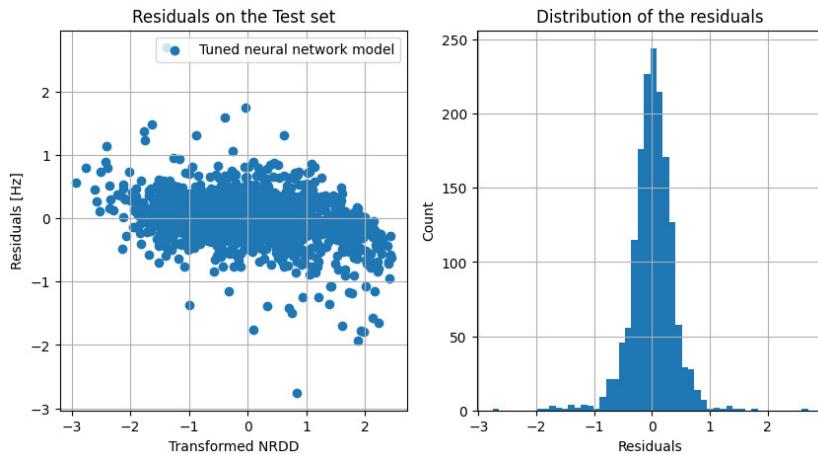
The uniformity of displacement has been modelled using many estimators. In this section, which focuses on interpreting the model's results, we will:

- Analyze the neural network model's residuals (with the highest R² value)
- Look at the worst and best predictions
- Perform a SHAP analysis to understand how the variables affect the predictions.

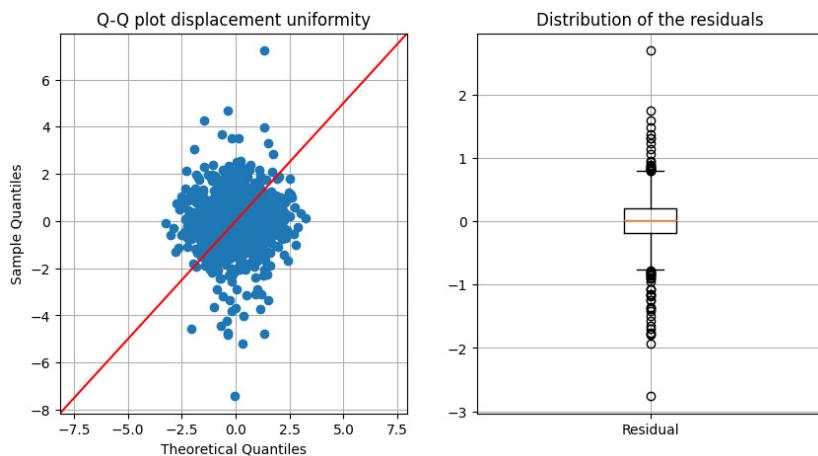
3.3.6.1 Residuals of the transformed target variable

The Keras neural network model was developed to predict the transformation of the uniformity criterion for displacement (NRDD) converted by the PowerTransformer. The following graphs show the distribution of the raw residuals calculated using the transformed values of NRDD:

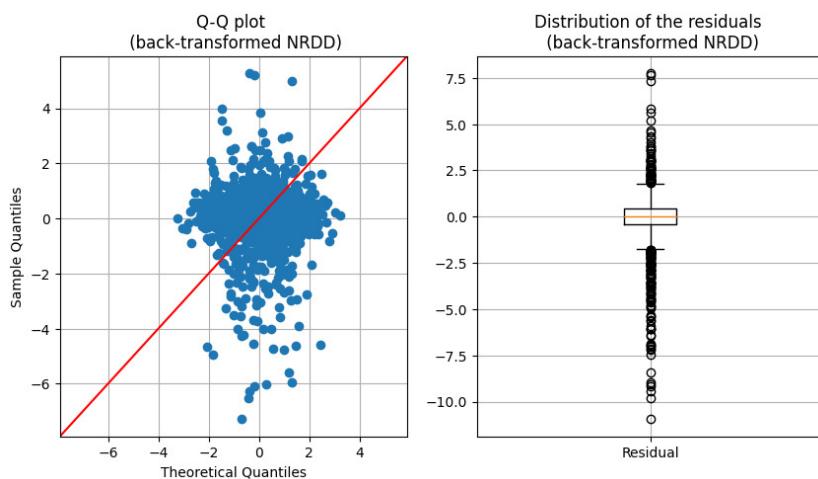
Evaluation technical report



We notice in the Q-Q plot that the distribution does not follow a normal law, but that the residuals are centred on zero:



If the residuals are calculated using the back-transformed values of the NRDD (for a matter of understanding), we observe a similar pattern:



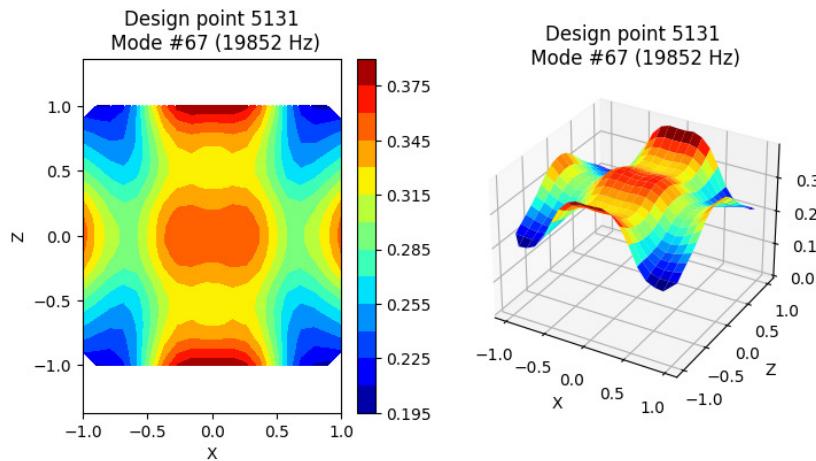
Evaluation technical report

3.3.6.2 Best and worst predictions

In the same way as the longitudinal frequency, the design points with both best and worst predictions will be analyzed:

	Worst prediction	Best prediction
Position of the longitudinal mode	66	67
Predicted NRDD	2.14	3.31
True NRDD	7.85	3.31
Design point	DP 5961	DP 5131

The residual of the best prediction is almost zero. The following figure represents the displacements of mode 67:



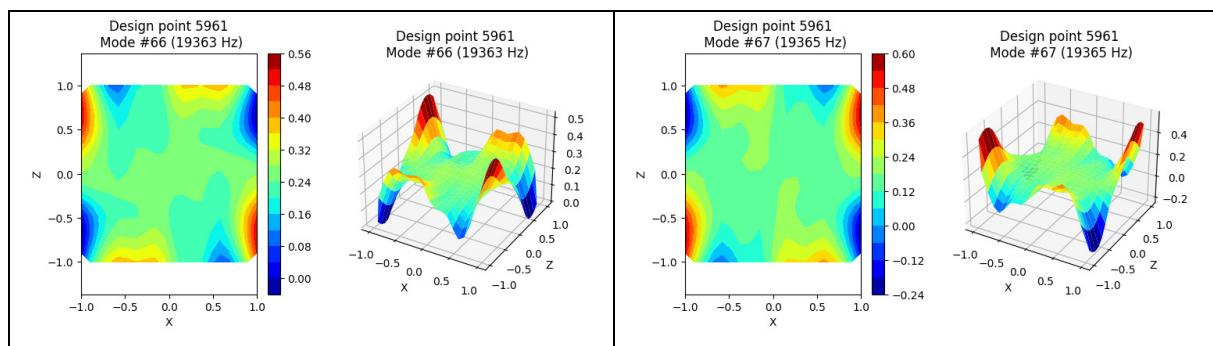
The worst prediction generated a large residual for the design point DP 5961. The frequency of the modes immediately below and above the longitudinal mode #66 reveals that the frequency gap to the next mode #67 is very small (1.8 Hz).

The plots of the mode shapes highlight a typical case of a modal interaction, where two modes swap their properties. The longitudinal mode receives part of its neighbor's displacement distribution, resulting in a poor uniformity of its displacement.

```

freq_long      19363.449
freq_delta_below -486.686
freq_delta_above   1.813

```

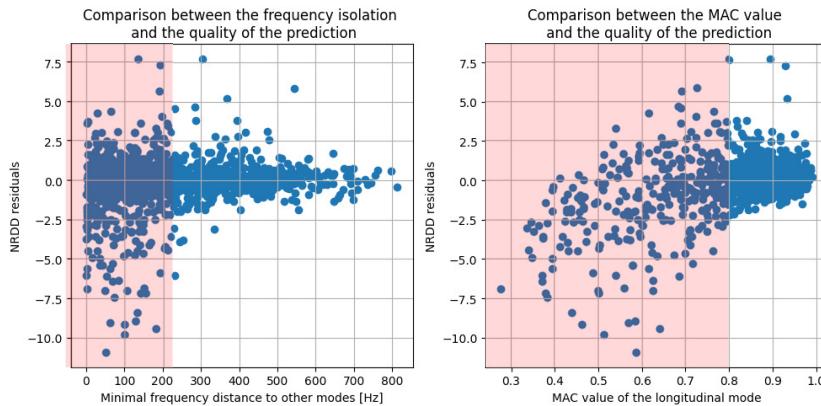


Evaluation technical report

3.3.6.3 Filtering the data set

As we did before, we are wondering if filtering the data set before training would improve the model's performance:

The minimum frequency distance corresponds to the smallest frequency gap between the longitudinal mode and its two neighbors (minimum value between `freq_delta_below` and `freq_delta_above`). The left diagram represents the residuals as a function of this value.



When the longitudinal frequency is isolated from other modes, the residuals become smaller. The MAC values tend towards one.

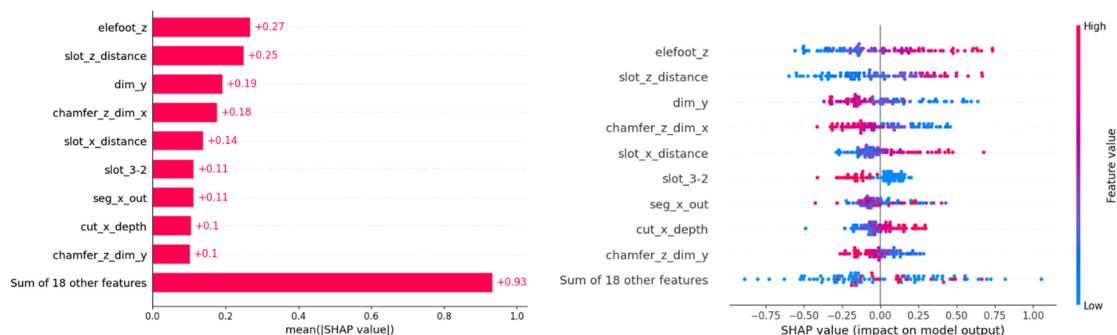
As well as for the longitudinal frequency prediction, it would make sense to keep the design points with the well-identified longitudinal mode to achieve the best accuracy. This statement favors the design strategy for the sonotrodes: the longitudinal mode should be as isolated as possible from other modes, which is part of the goals.

If we decide to filter the design points, we can do so by keeping:

- The design points with minimum frequency distances *above* 250 Hz, which would keep about 35% of the data set.
- The design points with MAC values *above* 0.8, which would keep about 80% of the data set and significantly improve the model's precision.

3.3.6.4 SHAP values

The left graph illustrates the most significant factors influencing the displacements uniformity. The beeswarm plot indicates the direction of these influences, thanks to the sign:



Basically the `elefoot_z` parameter is counterproductive. Higher values increase the NRDD, which in turn worsens uniformity. To address this, the slot distances `slot_z_distance` and `slot_x_distance` should be reduced, while the chamfers should be increased. These parameters are typically adjusted in practice to improve the sonotrode's output displacement uniformity.

Evaluation technical report

Furthermore, we can see that the slot class '3-2' is the best slot distribution for this problem. It leads to the lowest NRDD values when active (Feature value = 1).

4 Description of the work completed

4.1 Difficulties / eases encountered during the project

On one hand, the lack of idea exchange was a significant challenge during the project, since I was working alone without a team. It would have been helpful to have discussed different viewpoints before starting to explore the subject. The timeline for progress limited the number of approaches I could try.

On the other hand, my previous experience designing sonotrodes made it easier to define the objectives and interpret the results. I could conduct intensive computations directly and remotely on my workstation and use the efficient hardware and software available in the company. I generated the data sets based on the design points and cleaned them up briefly at the beginning.

Determining the variables to be predicted in the case of output displacements took longer than expected, as there were many possibilities to reduce the field of displacements and I had to test and evaluate the performance of various models.

4.2 Assessment and Continuation of the project

4.2.1 Assessment

We can conclude that the objectives of the project and the performance of the model are consistent: frequency isolation and unambiguous identification of the longitudinal mode are simultaneously required for the design of the sonotrode and for obtaining an accurate prediction of the frequencies and the uniformity of displacement.

Only in rare circumstances are modal interactions thoroughly examined. The proposed models may not be suitable for this type of problem. From a scientific perspective, the lack of a literature review on related topics is concerning. The bibliography could have served as a solid starting point for model development.

In light of the project's objectives, it is clear that the desired outcomes have been met. The models are precise in particular instances that align with the goals.

4.2.2 Next steps

If more time was available, we could have tested other evaluation metrics or combined them with the R^2 coefficient, depending on the target variable(s). For example, the Mean Squared Error (MSE) measures the average of squared residuals and provides an idea of the magnitude of the errors in the model's predictions.

Predicting every node displacement, as well as the longitudinal mode and all other modes, would significantly enhance our ability to predict modal interactions. The deformed shape's eigenvectors are forming a basis, allowing any new displacement vector to be decomposed into this basis [4]. We may predict the main components of the rotation matrix, giving an approximation of the entire transformed vector.

The developed models can fit into the process of designing a sonotrode especially in the case of critical geometries where the modal properties are not favorable. Tuning sonotrodes with specific dimensions can be time-consuming, but predictions can save simulation time.

Evaluation technical report

In a later project, we could develop an optimizer, working in conjunction with the prediction models to produce a design proposal without human intervention.

5 Bibliography

[1] M Pastor, M Binda, T Harčarik - Procedia Engineering, 2012 - Elsevier - Modal Assurance Criterion
<https://www.sciencedirect.com/science/article/pii/S1877705812046140>

[2] Cardoni, A., Lim, F.C.N. , Lucas, M. and Cartmell, M.P. (2002) Modal interaction in ultrasonic systems

[3] D. Afolabi, Modal interaction in linear dynamic systems near degenerates modes, National Aeronautics and Space Administration (1991)

[4] T. Yang, J. H. Griffin, A normalized modal eigenvalues approach for resolving modal interaction, Journal of Engineering for Gas Turbines and Power 119, 647-650 (1997)

[5] R2 Score: Linear Regression

<https://medium.com/@deependra.verma00/r2-score-linear-regression-e095a1188e87>

6 Appendices

6.1 Data Audit

design-points-03.csv

contains the list of design points with geometric parameters

```
design-points-03.csv
```

dp_no	dim_x	dim_y	dim_z	nb_slots_x	slot_x_length	slot_x_distance	nb_slots_z	slot_z_length	slot_z_distance	chamfer_x	chamfer_y	chamfer_z_dim_x	chamfer_y_dim_x	chamfer_z_dim_y	elefoot_x	elefoot_z	cut_x_start	cut_x_end	cut_x_depth	cut_z_start	cut_z_end	cut_z_depth						
1	1000;175;120;240;4;77;58;12;6;8;6;12;0;1;0;1;30;70;0;1;25;55;0;1	1001;175;0;6;121;95;240;31;4;75;993;43;235;2;78;219;59;508;15;638;12;074;1;8138;5;4657;12;201;6;0551;4;2614;4;1238;2	8;51;97;91;0;88177;25;293;66;06;1;9945	1002;173;45;122;77;238;13;4;81;168;45;059;2;78;924;57;208;11;718;9;9347;9;6338;1;057;11;558;6;0339;1;3697;3;8841;40	4;496;79;356;4;859;26;997;56;481;4;803	1003;175;85;116;71;238;1;4;77;532;45;26;2;80;099;49;215;2;4722;9;9912;3;814;7;5228;16;019;3;6131;3;4375;1;3356;21;0	81;83;599;1;2178;22;264;51;204;0;81847	1004;173;51;116;37;238;3;4;82;428;51;007;2;74;757;50;84;12;108;12;838;1;8759;10;539;4;7831;4;0512;0;86571;4;0132;53	.546;63;529;4;165;37;379;58;304;1;0282	1005;175;6;118;72;241;04;4;76;576;45;981;2;78;773;64;702;15;87;11;024;9;8682;9;7962;10;076;11;955;2;8653;0;97684;36	.388;89;119;0;64304;27;381;51;502;0;74029	1006;175;95;122;34;240;71;4;75;052;54;94;2;74;005;58;988;11;089;17;156;5;0753;7;4036;6;2974;8;3918;2;8905;3;9462;56	.666;65;926;4;1891;27;802;67;867;2;8444	1007;173;88;123;57;241;79;4;79;669;42;284;2;81;424;56;839;3;3621;16;032;0;34361;3;5168;5;3254;14;112;4;742;1;1019;3	2;485;68;136;2;2424;36;054;46;83;0;22284	1008;176;66;117;240;41;4;78;671;50;954;2;80;038;53;146;2;1469;7;3374;5;795;7;2382;4;6128;10;242;3;0949;4;531;56;882	;73;258;1;6;0307;37;137;60;277;2;9584	1009;174;25;121;85;241;47;4;74;14;74;48;019;2;78;355;52;106;7;7334;4;4537;7;8336;2;403;3;0494;6;9991;3;5452;4;6008;54	.761;97;437;0;90816;34;603;56;925;2;3412	1010;176;41;122;73;241;85;4;80;978;51;195;2;76;951;62;821;14;201;16;242;2;4103;10;396;16;099;6;8716;4;2357;1;4741;5	5;02;86;209;0;13012;20;144;61;138;3;9149	1011;176;11;115;17;239;33;4;78;575;41;628;2;75;245;61;219;15;557;18;293;11;215;7;2941;5;2916;1;0484;1;5866;4;4783;4	3;688;93;3;0;15764;27;245;54;41;1;9081	1012;173;39;120;19;241;46;4;77;772;46;408;2;74;156;63;737;5;944;8;254;3;0831;5;9404;17;698;13;754;3;0039;0;78393;45	.341;89;361;1;1755;41;464;48;223;4;7387	1013;177;122;85;240;14;4;75;18;43;702;2;79;249;64;084;0;48529;4;3806;4;7133;10;588;9;3344;9;0846;1;5223;1;8237;39;4	26;61;866;2;4344;32;497;68;207;3;6977	1014;173;74;124;74;241;55;4;80;948;43;493;2;75;229;65;086;12;728;0;24759;5;8537;5;7447;13;128;1;5153;3;334;4;103;27

mode-ident-full-03.dat

contains the overview of the design point results and results of the longitudinal mode's identification

Evaluation technical report

#	DPNr	ModeNr	Freq	RelAbw	MeanUey	MeanUay	ModeTran	ModeMAC
1	1000.	78.	19600.060	0.29446	-0.312	0.305	0.97904	0.9808981981
2	1001.	77.	19217.965	0.92217	-0.371	0.280	0.75609	0.9217838992
3	1002.	79.	19386.656	0.95697	-0.386	0.293	0.75940	0.9234288892
4	1003.	80.	20087.459	0.51795	-0.337	0.321	0.95450	0.9713867639
5	1004.	79.	20209.177	0.79699	-0.351	0.319	0.90825	0.9482598256
6	1005.	79.	19813.899	0.38915	-0.341	0.309	0.90749	0.9790681962
7	1006.	78.	19401.945	1.26503	0.417	-0.244	0.58502	0.8268708981

The column names in the ANSYS output have been redefined for easier comprehension. They are only displayed once. The column "ModeNr" represents the longitudinal mode number identified in Ansys. The other features are not necessary, since they can be retrieved from the "mode files" (see below).

Number of lines in the table : (family 01: 601) (family 02: 2001) (family 03: 8001)									
# Column	Name of the Column	Variable's type	Description	Is the variable available before prediction	Variable's type	Percentage of missing values	Categorical / Quantitative	Distribution	Comments
		Is the variable a feature or the target ? (Only applicable for supervised learning projects)	What does this variable represent (from a business perspective ?)	Is this variable known before the prediction is made ? (Only applicable for supervised learning projects)	int64, float etc...	in %		For categorical variables with less than 10 categories, list all categories.	Free text
1	dp_no	feature	Design point number	yes	int64	0,00%		between 0 and 11999	Original name: DPNr
2	mode_no	target	Mode number of the longitudinal mode	no	int64	0,00%	categorical	from 1 to 100	Original name: ModeNr
3	freq	target	Frequency of the longitudinal mode	no	float	0,00%	quantitative	above 0	Original name: Freq (redundant value)
4	rel_dev	feature	Relative deviation of the displacement of the output surface	no	float	0,00%	quantitative	above 0	Original name: RelAbw (redundant value)
5	mean_in_disp	feature	Average displacement of the input displacement	no	float	0,00%	quantitative		Original name: MeanUey (redundant value)
6	mean_out_disp	feature	Average displacement of the output surface	no	float	0,00%	quantitative		Original name: MeanUay (redundant value)
7	mode_gain	feature	Gain of the long. mode (output / input displacement)	no	float	0,00%	quantitative		Original name: ModeTran (redundant value)
8	mode_mac	feature	MAC value of the long. mode	no	float	0,00%	quantitative	between 0 and 1	Original name: ModeMAC (redundant value)

3 subdirectories: modes, nodes, defs

./data-03

design-points-03.csv
mode-ident-full.dat

./modes ./nodes.dat ./defs.dat

tbl-modes-0.dat	tbl-nodes-0.dat	tbl-defs-0.dat
tbl-modes-1.dat	tbl-nodes-1.dat	tbl-defs-1.dat
tbl-modes-11999.dat	tbl-nodes-11999.dat	tbl-defs-11999.dat

Additionally, 3 files are created in the corresponding subdirectories for each design point:

tbl-modes-(dp).dat

The « mode files » contain the list of the first computed modes.

#	ModeNr	Freq	RelAbw	MeanUey	MeanUay	ModeTran	ModeMAC
1	1.	17309.306	2821.42607	-0.001	-0.000	0.70246	0.0000012937
2	2.	17580.286	7.08428	0.111	0.278	2.49783	0.0859419390
3	3.	18278.168	17119.96784	-0.000	0.000	0.75937	0.0000001039
4	4.	18976.973	11685.73962	0.000	-0.000	3.07165	0.0000001360
5	5.	19552.164	1644.27622	0.001	-0.001	0.63710	0.0000195102
6	6.	20020.662	0.23651	-0.617	1.370	2.22040	0.9957468188
7	7.	21281.712	3568.40663	0.001	-0.001	1.28042	0.0000009409
8	8.	21406.993	5090.73601	-0.001	0.000	0.39451	0.0000062592
9	9.	21443.076	3719.00771	-0.000	0.000	0.94309	0.0000030271
10	10.	23677.840	46.43709	-0.811	-0.075	0.09296	0.3383284693
11	11.	23803.344	3276.26027	-0.005	-0.000	0.09003	0.0002764554
12	12.	24782.837	10254.30642	-0.001	-0.000	0.22393	0.0000040893
13	13.	25422.144	29611.24334	-0.001	0.000	0.08344	0.0000028454

Evaluation technical report

Number of lines in the table : 100									
# Column	Name of the Column	Variable's type	Description	Is the variable available before prediction	Variable's type	Percentage of missing values	Categorical / Quantitative	Distribution	Comments
1	mode_no	target	Mode number of the longitudinal mode	no	int64	0,00%	categorical	from 1 to 100	Original name: ModeNr
2	freq	target	Frequency of the longitudinal mode	no	float	0,00%	quantitative	above 0	Original name: Freq
3	rel_dev	feature	Relative deviation of the displacement of the output surface	no	float	0,00%	quantitative	above 0	Original name: RelAbw
4	mean_in_disp	feature	Average displacement of the input displacement	no	float	0,00%	quantitative		Original name: MeanUey
5	mean_out_disp	feature	Average displacement of the output surface	no	float	0,00%	quantitative		Original name: MeanUay
6	mode_gain	feature	Gain of the long. mode (output / input displacement)	no	float	0,00%	quantitative		Original name: ModeTran
7	mode_mac	feature	MAC value of the long. mode	no	float	0,00%	quantitative	between 0 and 1	Original name: ModeMAC

tbl-nodes- (dp) .dat

The “node files” contain the node located on the output surface, along with their numbers and spatial coordinates.

mode-ident-full.dat	tbl-modes-1.dat	tbl-nodes-1.dat	tbl-def-y-1.dat	mode_ident_plot.m
1	17. -0.8000000000E-02	-0.1310000000E+00	-0.8000000000E-01	
2	21. 0.8000000000E-02	-0.1310000000E+00	-0.8000000000E-01	
3	22. -0.5329070518E-17	-0.1310000000E+00	-0.8000000000E-01	
4	52. 0.8000000000E-02	-0.1310000000E+00	0.8000000000E-01	
5	56. -0.8000000000E-02	-0.1310000000E+00	0.8000000000E-01	
6	57. 0.0000000000E+00	-0.1310000000E+00	0.8000000000E-01	
7	152. -0.8000000000E-02	-0.1310000000E+00	-0.7000000000E-01	
8	153. -0.8000000000E-02	-0.1310000000E+00	-0.6000000000E-01	
9	154. -0.8000000000E-02	-0.1310000000E+00	-0.5000000000E-01	
10	155. -0.8000000000E-02	-0.1310000000E+00	-0.4000000000E-01	
11	156. -0.8000000000E-02	-0.1310000000E+00	-0.3000000000E-01	
12	157. -0.8000000000E-02	-0.1310000000E+00	-0.2000000000E-01	

Number of lines in the table : (number of nodes at the output surface)									
# Column	Name of the Column	Variable's type	Description	Is the variable available before prediction	Variable's type	Percentage of missing values	Categorical / Quantitative	Distribution	Comments
1	node_no	feature	Node number	yes	int64	0,00%			
2	x_coord	feature	X coordinate	yes	float	0,00%	quantitative	between -dim_x/2 and +dim_x/2	
3	y_coord	feature	Y coordinate	yes	float	0,00%	quantitative	0 or -0,13327	Constant value, since the nodes are located on the output surface
4	z_coord	feature	Z coordinate	yes	float	0,00%	quantitative	between -dim_z/2 and +dim_z/2	

tbl-def- (dp) .dat

The “deformation files” contain the displacements of the nodes on the output surface for each mode, only in the longitudinal direction (Y-direction)

mode-ident-full.dat	tbl-modes-1.dat	tbl-nodes-1.dat	tbl-def-y-1.dat	mode_ident_plot.m
1	0.1642179179E+00	0.1514911523E+01	-0.9505992613E+00	-0.1197113037E+01
	-0.1627596544E+00	0.1514110885E+01	0.9476157132E+00	0.1194227657E+01
3	-0.1951863941E-02	0.1522073829E+01	0.2069525906E-02	0.3561670560E-02
	-0.1690226599E+00	0.1515911869E+01	0.9739713450E+00	-0.1188787008E+01
5	0.1579727480E+00	0.1515946447E+01	-0.9696020534E+00	0.1178264271E+01
	0.2398570742E-03	0.1525436561E+01	-0.2081609402E-02	0.6948431811E-03
7	0.2539552311E+00	0.1066842340E+01	-0.9249695506E+00	-0.1176740863E+01
	0.4395722502E+00	0.6337006857E+00	-0.8584628763E+00	-0.1102288154E+01
9	0.5607473140E+00	0.2649871053E+00	-0.8016238368E+00	-0.9884888940E+00
	0.5458477769E+00	0.5479642881E-02	-0.788293157E+00	-0.8399233864E+00
11	0.4068202705E+00	-0.1368334791E+00	-0.8079955933E+00	0.6283019997E+00
	0.1692510512E+00	0.6277067671E+00	0.8175021011E+00	0.2007197707E+00

Evaluation technical report

	Number of lines in the table : (number of nodes at the output surface)								
# Column	Name of the Column	Variable's type	Description	Is the variable available before prediction	Variable's type	Percentage of missing values	Categorical / Quantitative	Distribution	Comments
			<i>Is the variable a feature or the target ? (Only applicable for supervised learning projects)</i>	<i>What does this variable represent (from a business perspective ?)</i>	<i>Is this variable known before the prediction is made ? (Only applicable for supervised learning projects)</i>	<i>int64, float etc...</i> <i>If "object", develop.</i>	<i>in %</i>		<i>For categorical variables with less than 10 categories, list all categories.</i> <i>For quantitative variables, detail the distribution (basic descriptive statistics)</i>
1	mode1	feature	Node displacements for mode 1	no	float	0,00%	quantitative		
2	mode2	feature	Node displacements for mode 2	no	float	0,00%	quantitative		
	...								
100	mode100	feature	Node displacements for mode 100	no	float	0,00%	quantitative		