Definition of the context and of the scope

The task of the ultrasonic sonotrode is to transfer the mechanical vibration energy to the welding part and assure a constant high welding quality. This part has many vibrating modes: each mode has a eigenfrequency and an eigenshape. The first longitudinal mode in particular is used to drive the vibration and has the following requirements:

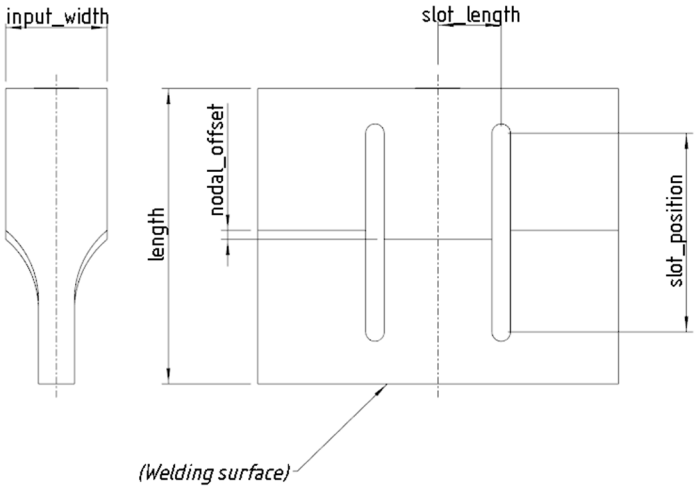
- its eigenfrequency should be around the driving frequency
- the displacement of the eigenshape at the welding surface (output surface) should be as uniform as possible
- it should be isolated from other modes

The modal properties of the sonotrode are simulated with the software Ansys. They depend on the material properties (supposed constant for that study) and the geometry. The geometry is usually tuned by the engineer in an iterative way and simulated until the requirements of the modal properties are filled.

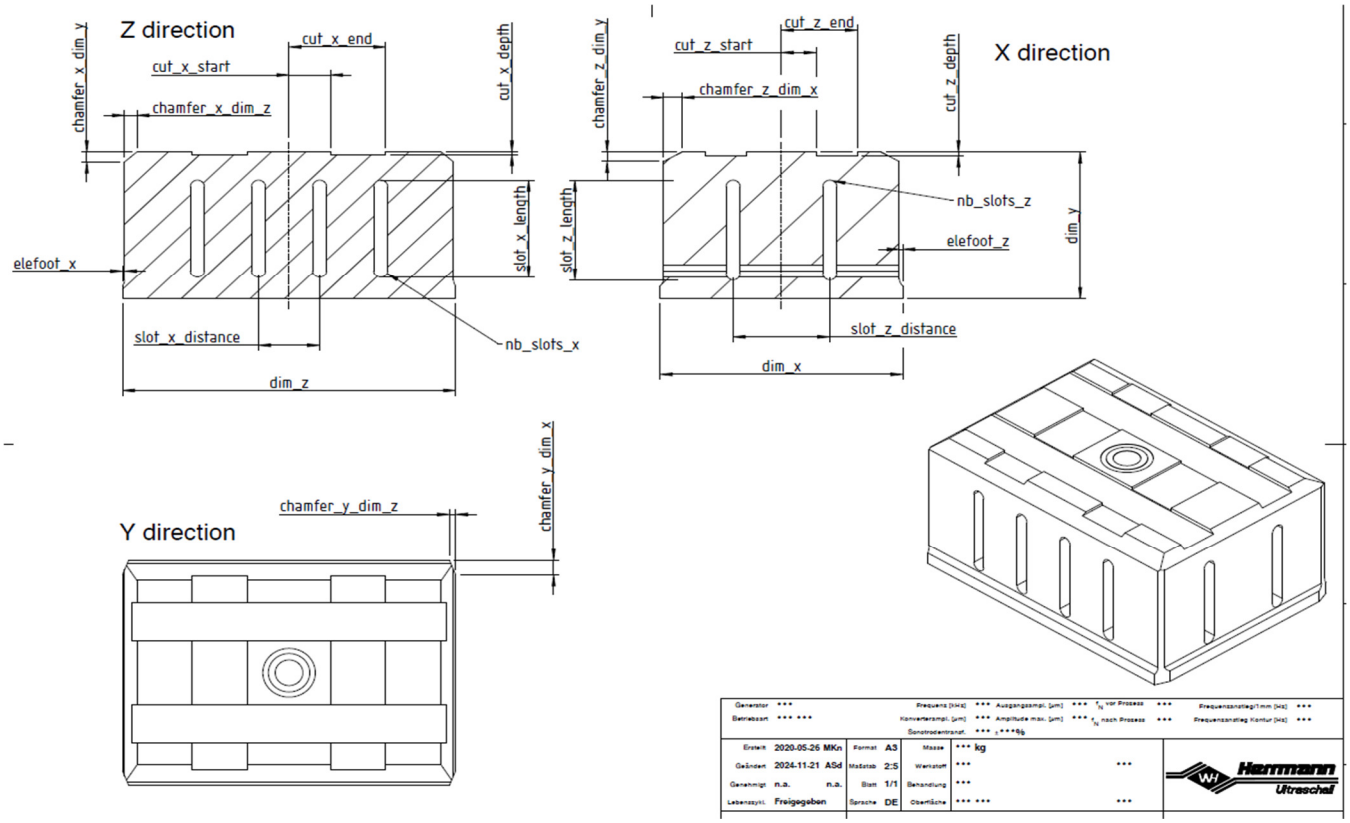
The geometry of the sonotrode is described for this project between 2 and 23 parameters, depending on the sonotrode type and complexity. The data were generated from the simulation results of many hundreds of design points, classified in three families:

- in the design family 01, the geometry of a blade sonotrode is changed according to 2 parameters only (slot position and slot length, see next schema)

- in the design family 02, the same blade sonotrode is described geometrically by 5 parameters instead of 2 parameters



- the block sonotrode of the design family 03 is parametrized by 23 parameters



The longitudinal mode is identified in the Ansys post-script with the MAC (Modal Assurance Criterion), which compares the eigenshapes with a reference vector.

The objective of the project is predict the FEM results (eigenfrequencies and eigenshapes) with Machine Learning models. These predictions should permit to:

- calculate the eigenshapes of each mode
- estimate the frequencies of each mode
- automatically determine the number of the first longitudinal mode (optional requirement)

A greater precision is required particularly in the vicinity of the first longitudinal mode (direct neighbor modes, for example the 5 modes below and 5 modes above).

Discovering of the dataset and its structure

(Data Audit)

For the creation of the data samples, many hundred of design points were randomly generated in arbitrarly defined ranges and calculated in FEM. The design points are listed in Excel for each family and exported as CSV files

design-points-01.csv

design-points-01.csv

```
dp_no;a;b
0;88;28
1;90.534;33.583
2;86.65;28.343
3;93.091;27.078
4;88.949;31.178
5;89.663;36.253
6;90.788;26.921
7;88.881;36.578
8;87.245;36.852
9;78.002;33.364
10;90.583;23.115
11;87.952;35.029
12;88.278;32.043
13;96.328;35.316
14;95.725;34.452
15;80.062;23.781
16;87.685;29.677
17;82.448;32.146
18;95.615;25.965
19;95.899;36.333
```

	Number of lines in the table : 601								
# 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.* *For quantitative variables, detail the distribution (basic descriptive statistics)*	*Free text*
1	dp_no	feature	Design point number	yes	int64	0,00%	quantitative	between 0 and 600	
2	a	feature	Slot length	yes	float	0,00%	quantitative	between 22 and 38	
3	b	feature	Slot position	yes	float	0,00%	quantitative	between 78 and 98	

design-points-02.csv

design-points-02.csv

```
dp_no;a;b;c;d;e
100;88;28;131;4;45
101;92.946;23.481;128.68;5.8024;42.584
102;89.685;34.334;128.39;3.4915;41.655
103;88.366;27.754;130.79;3.1933;44.697
104;84.941;25.859;129.41;6.5444;43.605
105;91.61;30.096;133.27;0.93856;47.203
106;88.522;37.607;131.98;2.2184;46.604
107;87.596;28.59;129.71;0.16035;45.376
108;84.169;23.843;130.26;2.7667;41.819
109;89.937;37.847;129.89;6.7239;42.442
110;90.219;28.283;129.8;3.5597;41.719
111;90.07;30.646;130.55;3.934;44.008
112;89.48;24.678;130.87;6.556;46.729
113;87.58;29.347;130.29;0.9313;43.431
114;92.239;36.859;133.8;6.5516;45.619
115;85.352;36.747;131.3;6.8388;45.319
116;84.526;31.48;128.23;5.1279;48.648
117;88.28;33.114;132.03;0.96187;48.95
118;90.696;22.763;131.58;4.1647;48.211
119;84.6;34.651;128.48;4.1679;48.997
120;89.124;22.16;133.48;4.9;41.801
121;89.105;30.838;128.93;3.2797;48.287
122;91.741;33.978;131.77;2.6716;42.098
```

	Number of lines in the table : 2001								
# 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.* *For quantitative variables, detail the distribution (basic descriptive statistics)*	*Free text*
1	dp_no	feature	Design point number	yes	int64	0,00%	quantitative	between 100 and 2100	
2	a	feature	Slot length	yes	float	0,00%	quantitative	between 83 and 93	
3	b	feature	Slot position	yes	float	0,00%	quantitative	between 22 and 38	
4	c	feature	Length	yes	float	0,00%	quantitative	between 128 and 134	
5	d	feature	Nodal offset	yes	float	0,00%	quantitative	between 0,1 and 7	
6	e	feature	Input width	yes	float	0,00%	quantitative	between 40 and 50	

design-points-03.csv

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_dim_y;chamfer_x_dim_z;chamfer_y_dim_x;chamfer_y_dim_z;chamfer_z_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
1000;175;120;240;4;77;48;2;77;58;12;6;8;8;6;12;0.1;0.1;30;70;0.1;25;55;0.1
1001;175.06;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;28.51;97.91;0.88177;25.293;66.06;1.9945
1002;173.45;122.77;238.13;4;81.168;45.859;2;78.924;57.208;11.718;9.9347;9.6338;1.057;11.558;6.0339;1.3697;3.8841;40.496;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;21.081;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;32.485;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;1.6037;37.137;60.277;2.9584
1009;174.25;121.85;241.47;4;74.147;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;55.02;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;43.688;93.3;0.15764;27.245;54.411;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;39.426;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
```

	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.* *For quantitative variables, detail the distribution (basic descriptive statistics)*	*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	

The FEM results are exported for each design family in one main file and 3 subdirectories:

./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

- the „main file“ **mode-ident-full-0x.dat** contains the overview of the design point results, one line per design point

mode-ident-full-03.dat

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

The columns names given as output from Ansys should be renamed for better understanding and displayed only once. This contains the mode number of the identified longitudinal mode in Ansys. The other features are actually redundant, since they can be picked from the „mode files“.

	Number of lines in the table : (familly 01: 601) (familly 02: 2001) (familly 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...* *If "object", develop.*	*in %*		*For categorical variables with less than 10 categories, list all categories.* *For quantitative variables, detail the distribution (basic descriptive statistics)*	*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)

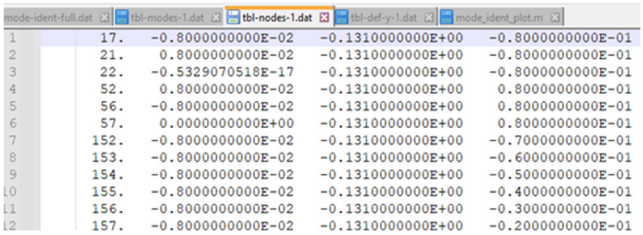
- for <u>each</u> design point, 3 different files are written in the corresponding subdirectories:
    - the „mode files“ tbl-modes-(dp).dat contain the list of the first calculated modes

node-ident-full.dat | tbl-modes-1.dat | tbl-nodes-1.dat | tbl-def-y-1.dat | mode_ident_plot.m

```
# ModeNr      Freq        RelAbw      MeanUey      MeanUay      ModeTran      ModeMAC
     1.    17309.306     2821.42607       -0.001       -0.000       0.70246   0.0000012937
     2.    17580.286        7.08428        0.111        0.278       2.49783   0.0859419390
     3.    18278.168    17119.96784       -0.000        0.000       0.75937   0.0000001039
     4.    18976.973    11685.73962        0.000       -0.000       3.07165   0.0000001360
     5.    19552.164     1644.27622        0.001       -0.001       0.63710   0.0000195102
     6.    20020.662        0.23651       -0.617        1.370       2.22040   0.9957468188
     7.    21281.712     3568.40663        0.001       -0.001       1.28042   0.0000009409
     8.    21406.993     5090.73601       -0.001        0.000       0.39451   0.0000062592
     9.    21443.076     3719.00771       -0.000        0.000       0.94309   0.0000030271
    10.    23677.840       46.43709       -0.811       -0.075       0.09296   0.3383284693
    11.    23803.344     3276.26027       -0.005       -0.000       0.09003   0.0002784554
    12.    24782.837    10254.30642       -0.001       -0.000       0.22393   0.0000040893
    13.    25422.144    29611.24334       -0.001        0.000       0.08344   0.0000028454
```

	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
		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.* *For quantitative variables, detail the distribution (basic descriptive statistics)*	*Free text*
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

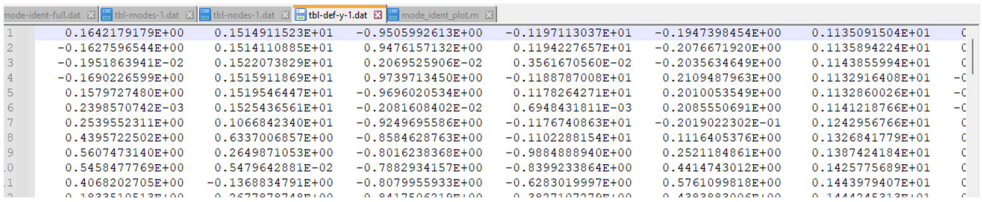
- the „node files“ tbl-nodes-(dp).dat contain the nodes of the output surface only, with coordinates



```
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
		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.* *For quantitative variables, detail the distribution (basic descriptive statistics)*	*Free text*
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	

- the „deformation files“ tbl-def-(dp).dat contain the displacement of the nodes located on the output surface. The displacement are given for each mode and in the longitudinal direction only (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   -0.1947398454E+00    0.1135091504E+01  0
2   -0.1627596544E+00    0.1514110885E+01    0.9476157132E+00    0.1194227657E+01   -0.2076671920E+00    0.1135894224E+01  0
3   -0.1951863941E-02    0.1522073829E+01    0.2069525906E-02    0.3561670560E-02   -0.2035634649E+00    0.1143855994E+01  0
4   -0.1690226599E+00    0.1515911869E+01    0.9739713450E+00   -0.1188787008E+01    0.2109487963E+00    0.1132916408E+01 -0
5    0.1579727480E+00    0.1519546447E+01   -0.9696020534E+00    0.1178264271E+01    0.2010053549E+00    0.1132860026E+01 -0
6    0.2398570742E-03    0.1525436561E+01   -0.2081608402E-02    0.6948431811E-03    0.2085550691E+00    0.1141218766E+01 -0
7    0.2539552311E+00    0.1066842340E+01   -0.9249695586E+00   -0.1176740863E+01   -0.2019022302E-01    0.1242956766E+01  0
8    0.4395722502E+00    0.6337006857E+00   -0.8584628763E+00   -0.1102288154E+01    0.1116405376E+00    0.1326841779E+01  0
9    0.5607473140E+00    0.2649871053E+00   -0.8016238368E+00   -0.9884888940E+00    0.2521184861E+00    0.1387424184E+01  0
10   0.5458477769E+00    0.5479642881E-02   -0.7882934157E+00   -0.8399233864E+00    0.4414743012E+00    0.1425775689E+01  0
11   0.4068202705E+00   -0.1368834791E+00   -0.8079955933E+00   -0.6283019997E+00    0.5761099818E+00    0.1443979407E+01  0
```

	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
		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.* *For quantitative variables, detail the distribution (basic descriptive statistics)*	*Free text*
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		

Data consolidation

* Rassembler toutes les données d'un seul coup pour éviter de revenir aux fichiers bruts:
* Extraire les paramètres géométriques des fichiers CSV
* Collecter les résultats de simulation FEM (fréquences et déplacements)
* Créer un dictionnaire ou DataFrame centralisé avec tous les design points

All the data are collected in the dictionary `all_data` to avoid to deal with raw data again.

all_data				
	'01'			
		'parameters'		
		'identification'		
		'design_data'		
			0	
				'modes'
				'nodes'
				'defs'
			1	
				'modes'
				'nodes'
				'defs'
			…	
			600	
				'modes'
				'nodes'
				'defs'
	'02'			
		'parameters'		
		'identification'		
		'design_data'		
			100	
				'modes'
				'nodes'
				'defs'
			101	
				'modes'
				'nodes'
				'defs'
			…	
			2037	
				'modes'
				'nodes'
				'defs'
	'03'			
		'parameters'		
		'identification'		
		'design_data'		
			1000	
				'modes'
				'nodes'
				'defs'
			1001	
				'modes'
				'nodes'
				'defs'
			…	
			11999	
				'modes'
				'nodes'
				'defs'

- The first key is the design family
- The key „parameters“ contains the list of design points with geometrical parameters. The list of design points is read from „design-points-xx.csv“ and is different for each family
- The key „identification“ contains the number of the longitudinal mode from „mode-ident-full-xx.csv“
- The key „design_data“ contains the data of all design points: modes, nodes and deformations

DataFrames for the design family „02“

```
all_data['02']['parameters']
```
✓ 0.0s

dp_no	slot_length	slot_position	length	nodal_offset	input_width
100	88.000	28.000	131.00	4.0000	45.000
101	92.946	23.481	128.68	5.8024	42.584
102	89.685	34.334	128.39	3.4915	41.655
103	88.366	27.754	130.79	3.1933	44.697
104	84.941	25.859	129.41	6.5444	43.605
...	...	...	...	...	...
2096	89.049	34.855	130.71	5.3875	42.877
2097	83.298	23.885	129.67	4.4879	43.661
2098	91.117	22.654	128.16	4.4353	45.137
2099	92.787	32.873	131.42	6.2983	49.404
2100	84.422	37.653	132.11	6.4033	48.979

2001 rows × 5 columns

```
all_data['02']['identification']
```
✓ 0.0s

dp_no	mode_no	freq	rel_dev	mean_in_disp	mean_out_disp	mode_gain	mode_mac
100	22	20020.243	0.23964	-0.617	1.369	2.21909	0.982752
101	22	20332.648	0.14075	-0.601	1.412	2.35204	0.997157
102	22	20612.574	0.22144	-0.654	1.381	2.11211	0.986601
103	22	20116.753	0.24645	-0.626	1.373	2.19367	0.985702
104	22	20090.275	0.25445	-0.625	1.363	2.17982	0.979134
...	...	...	...	...	...	...	...
2033	22	20486.995	0.24134	-0.645	1.374	2.13160	0.985980
2034	22	20592.676	0.30209	-0.666	1.360	2.04317	0.976012
2035	21	20137.804	0.30639	-0.667	1.351	2.02572	0.979389
2036	22	20142.861	0.27606	-0.634	1.357	2.14067	0.975680
2037	22	20347.944	0.36470	-0.698	1.348	1.93091	0.976085

1842 rows × 7 columns

DataFrame examples for the design point 100

```
all_data['02']['design_data'][100]['modes']
```
✓ 0.0s

mode_no	freq	rel_dev	mean_in_disp	mean_out_disp	mode_gain	mode_mac
17	17320.098	1781.78359	0.000	0.001	1.36485	8.700000e-09
18	17583.907	7.06817	0.112	0.278	2.49409	1.395820e-02
19	18281.803	12735.07533	0.000	0.000	0.39242	3.890000e-08
20	18983.452	13628.92285	-0.000	-0.000	0.48817	3.350000e-08
21	19563.566	3139.22192	-0.001	0.000	0.28737	6.557900e-06
22	20020.243	0.23964	-0.617	1.369	2.21909	9.827523e-01
23	21284.509	1124.94459	-0.003	0.003	0.87854	1.362670e-05
24	21409.196	2329.67715	0.000	0.001	3.64296	1.994000e-07
25	21442.497	4529.43903	-0.001	0.000	0.27033	8.678000e-07
26	23685.305	46.63335	-0.810	-0.075	0.09280	3.209751e-01
27	23818.721	1899.44851	0.009	0.001	0.09324	4.724910e-05
28	24778.454	21500.81486	0.001	0.000	0.19117	8.460000e-08
29	25422.004	72145.51359	-0.001	0.000	0.03943	8.208000e-07
30	26706.666	32.54787	0.223	-0.087	0.38764	2.759145e-02

```
all_data['02']['design_data'][100]['nodes']
```
✓ 0.0s

node_no	x_coord	y_coord	z_coord
17	-8.000000e-03	-0.131	-0.08000
21	8.000000e-03	-0.131	-0.08000
22	-1.421085e-17	-0.131	-0.08000
52	8.000000e-03	-0.131	0.08000
56	-8.000000e-03	-0.131	0.08000
...	...	...	...
6490	-1.421085e-17	-0.131	-0.03000
6493	-1.421085e-17	-0.131	-0.04000
6496	-1.421085e-17	-0.131	-0.05000
6499	-1.421085e-17	-0.131	-0.06000
6502	-1.421085e-17	-0.131	-0.06875

169 rows × 3 columns

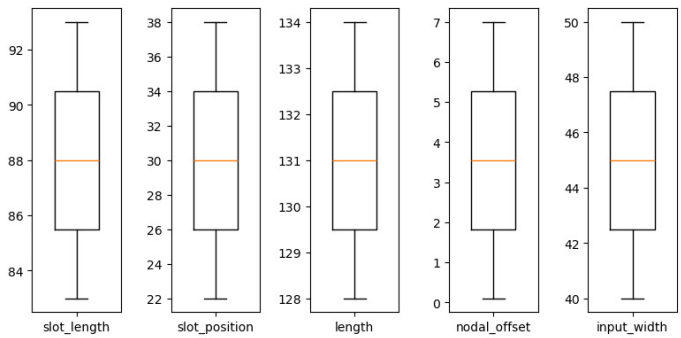
```
all_data['02']['design_data'][100]['defs']
```
✓ 0.0s

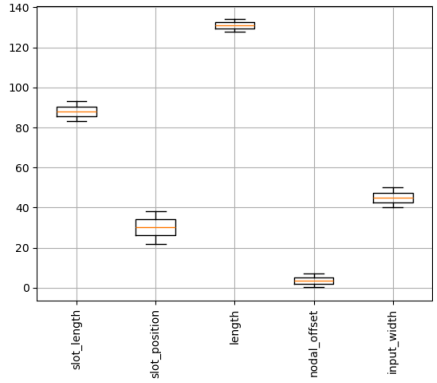
node_no	mode17	mode18	mode19	mode20	mode21	mode22	mode23	mode24	mode25	mode26	mode27	mode28	mode29	mode30
17	0.178134	1.514982	-0.956845	1.192229	-0.192502	1.137949	-1.600252	0.664140	-0.506705	-1.643926	0.645856	-1.051676	1.785693	-1.758488
21	-0.164502	1.513457	0.958615	-1.190917	-0.203059	1.138911	-1.595547	-0.651818	0.517927	-1.656907	-0.611877	1.065444	1.775005	-1.754014
22	0.004870	1.521893	0.003497	-0.002984	-0.200180	1.146847	-1.611992	0.002866	0.006171	-1.668829	0.012822	0.008951	1.800610	-1.776967
52	-0.166853	1.520378	0.960370	1.191760	0.208533	1.129018	1.596931	-0.640892	-0.547249	-1.651740	0.661819	1.056239	-1.781715	-1.730997
56	0.178136	1.519131	-0.961542	-1.197334	0.197584	1.128877	1.604093	0.643608	0.532246	-1.667674	-0.624725	-1.062915	-1.777240	-1.735823
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6490	-0.000351	-0.136564	0.000484	-0.003108	0.554375	1.455124	-1.028655	0.000999	0.005024	0.569401	-0.003684	-0.004406	-1.189925	0.631154
6493	0.000169	0.017152	0.000223	-0.002633	0.439560	1.439467	-1.431099	0.001320	0.006929	-0.207784	0.003503	-0.002177	-1.015964	0.988304
6496	0.000945	0.277782	0.000915	-0.001395	0.256727	1.401412	-1.643680	0.004338	0.005262	-0.784987	0.010155	-0.002843	-0.507529	0.774297
6499	0.003291	0.645520	0.003366	-0.003843	0.110891	1.340071	-1.702040	0.003209	0.007009	-1.157741	0.011776	0.003905	0.178190	0.133147
6502	0.002839	1.019648	0.003133	-0.003278	-0.007537	1.267131	-1.680533	0.002836	0.005721	-1.391340	0.012091	0.003994	0.857435	-0.640908

169 rows × 14 columns

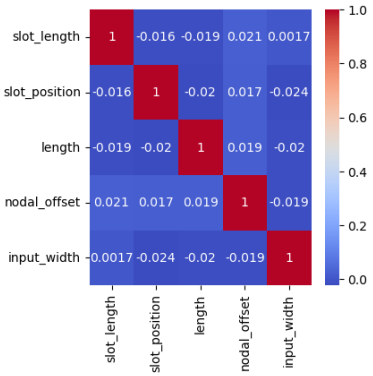
Data exploration

Distribution of the geometrical parameters





Exploration of the correlation between the parameters



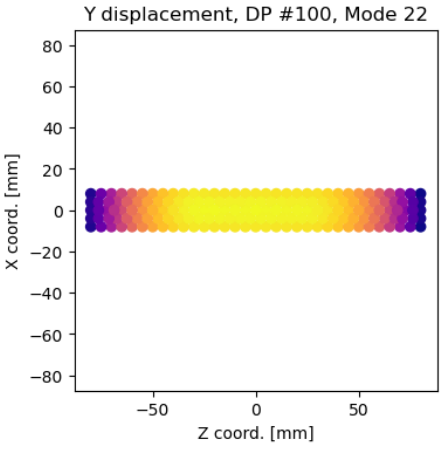
➔ The design points were independantly generated with a normal distribution

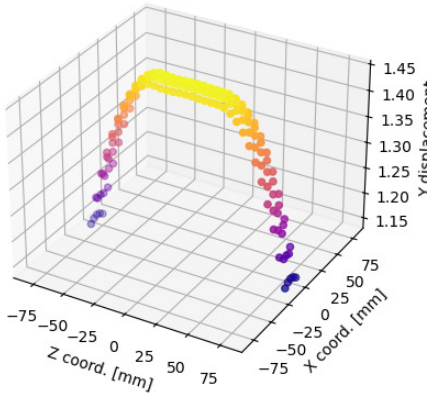
Visualize the range of the geometrical parameters

	slot_length	slot_position	length	nodal_offset	input_width
count	2001.000000	2001.000000	2001.000000	2001.000000	2001.000000
mean	87.999969	29.998999	131.000075	3.550235	44.999991
std	2.886781	4.619135	1.731980	1.991881	2.886762
min	83.000000	22.007000	128.000000	0.102900	40.002000
25%	85.502000	26.000000	129.500000	1.826100	42.501000
50%	88.000000	29.994000	131.000000	3.550000	45.000000
75%	90.497000	33.992000	132.500000	5.272300	47.496000
max	92.998000	37.993000	134.000000	6.999300	49.999000

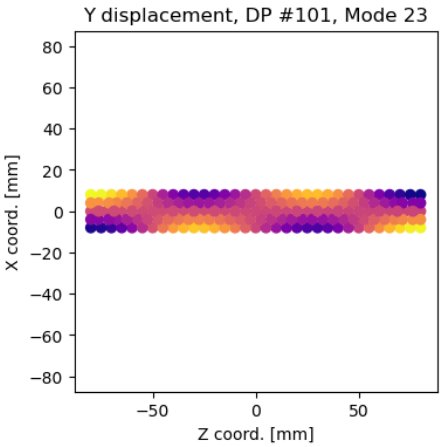
Graphical representations (at least 5)

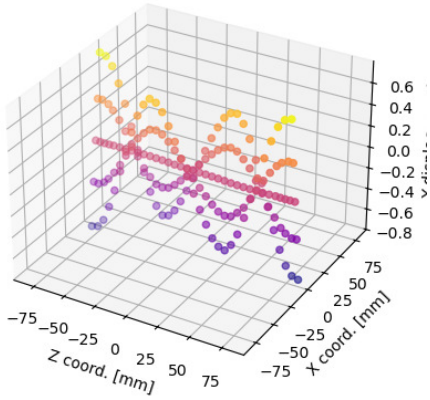
Uniformity of the displacement of the output surface for some longitudinal modes



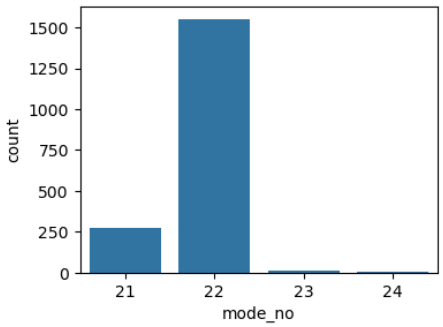


Visualization of the deformation shape for some modes (3D)



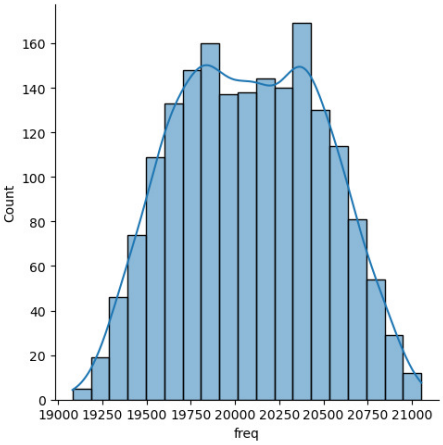


Distribution of the number of the longitudinal mode

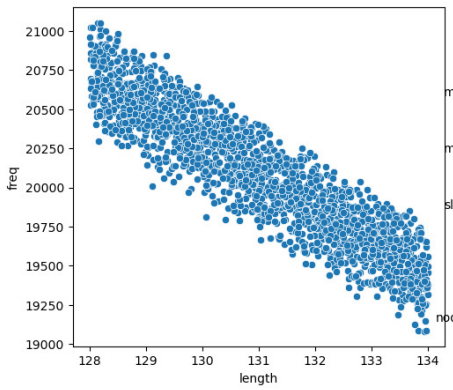


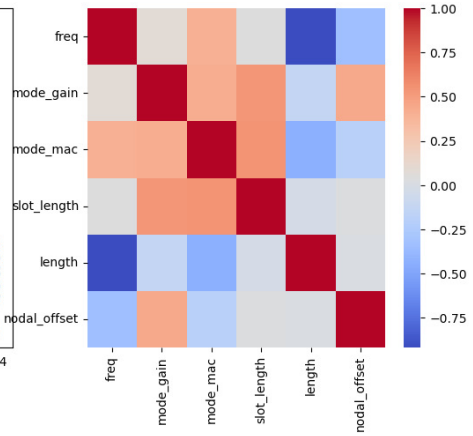

```
mode_no
22    1553
21     275
23      11
24       3
Name: count, dtype: int64
```

Distribution of the longitudinal frequencies around the nominal frequency



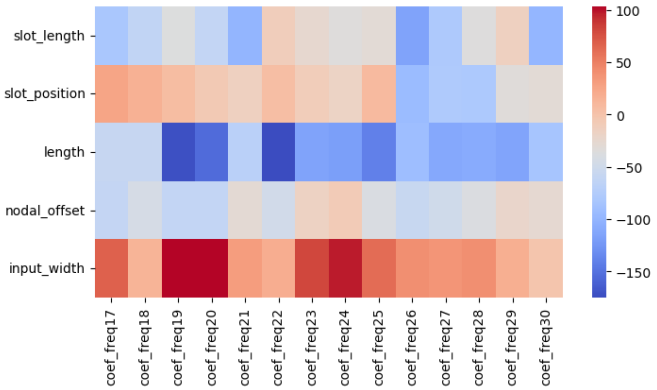
Sensivity of the geometrical parameters in the longitudinal frequency





	coefficient
slot_length	4.800572
slot_position	10.911957
length	-214.973229
nodal_offset	-67.114187
input_width	25.224254

Sensitivity of the geometrical parameters in all frequencies



Identification of the longitudinal mode

Elaborate a method to identify manually the first longitudinal mode
Validate this method for many design points

Recalculate the MAC value for each mode and compare with the values retrieved from Ansys