Database Guidelines
ML-powered tool for assisted design of buildings

This document details the database for the project ***ML-powered tool for assisted design of buildings***. The database contains information regarding 201 timber (wood) buildings designed with a light-frame structural system. The buildings were designed as part of a previous project aimed at evaluating the earthquake performance of light-frame timber buildings.

The goal of the project is to explore the feasibility that, by providing some easily available input information about the buildings, it is possible to develop an ML-powered tool that is capable of providing a first estimate of the structural design. Such a tool will be of great benefit to architects and engineers given the fact that the structural design of buildings is a long resource-intensive process done mostly manually nowadays by experienced personnel. This document will guide you through the information available in the database to facilitate the achievement of such a goal.

A normal building project can be summarized into four main phases. First, some basic information about the building is decided, such as how many stories (floors) it will have, in which location it will be placed, which materials it will use, on top of which type of sold it will be built, among others. For this project, we will call this information Type A. Second, an architect designs the shape and spaces of the building in conversation with the client and based on the information Type A. The results of this architectural work are the distribution of the rooms and the spatial placement of the structural walls. We will call this information Type B. Third, an engineer is hired to do the technical design of the structural walls of each floor of the building. The structural walls are responsible for keeping the building solid, and rigid, and for avoiding its collapse. They are also responsible for keeping the building safe during an earthquake or a hurricane. Therefore, the work of the engineer when designing the structural walls is to decide which are going to be the specifications of the walls, their dimensions, and which kind of materials will they use, among others. Such specifications will also depend on the material of the walls; for example, they will be different if you are using concrete versus if you are using timber. For this project, the information about the structural walls (their specifications) will be called Type C. Fourth, once all the information is available, the engineer can employ information Type A, B, and C to perform an evaluation of the safety of the building and use it to make better decisions. The information resulting from this evaluation will be called in this project Type D.

The database of the project has information Type A, B, C, and D for 201 buildings. Now that the meaning of this information is clearer, the goal of the project is to develop an ML tool that using information Type A and B as input, it can provide reliable estimates for information Type C and D. The motivation behind this is that obtaining information Type C and D is time-consuming and very expensive, so there is an important scientific motivation to find ways to speed up this process.

Attached to this document, you will receive eight Excel files with information about each of the 201 buildings. The files are called *"Design_#_##.xlsx"*. Every Excel file contains the information of a given subset of buildings. The reasoning behind the subsets will be explained in the next paragraphs.

Once you open the Excel file, you will find the information of the buildings organized by columns. The header of the columns will tell you mostly information Type A of a given building. For example, if you open the file "Design_C_HD.xlsx", the header for the first building shows:

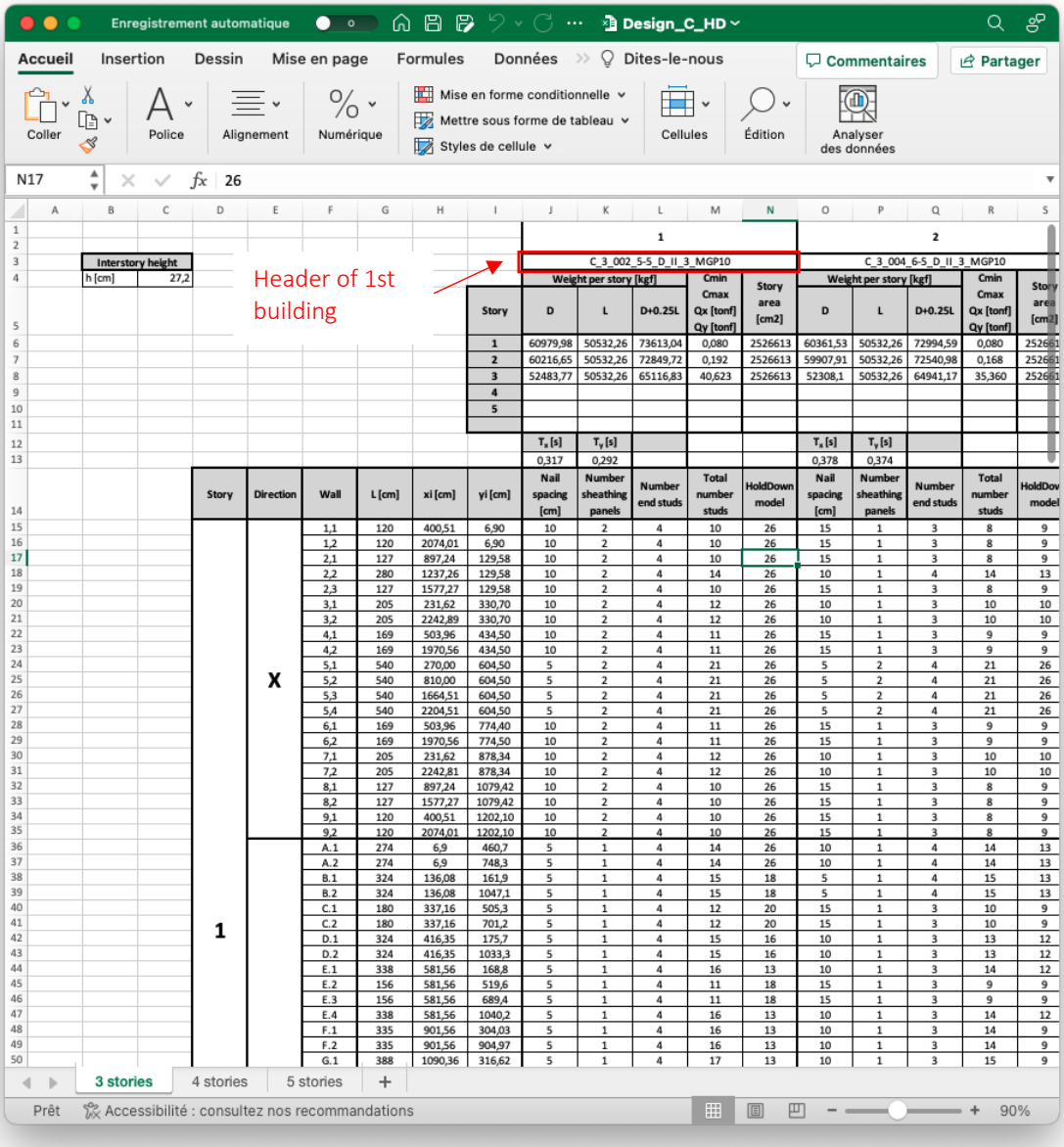


Figure 1. Header in Excel file "Design_C_HD.xlsx".

C_3_002_5-5_D_II_3_MGP10_HD. The meaning of the header is as follows:

1) *C*: architectural archetype "C".
2) *3*: three stories or floors.
3) *002*: maximum allowable drift equal to 0.002H.
4) *5-5*: R factor equal to 5.5.
5) *D*: soil class "D".
6) *II*: occupancy category "II".
7) *3*: seismic zone "3".
8) *MGP10*: timber quality used in the design.
9) *HD*: the connection system used is "hold-downs".

The parameter 1) tells you which architectural design the building uses. For the project, we have 4 architectural configurations, named C, D, P, and Q. For illustration purposes, you can see all these four architectural designs in Figure 2. Also, you can find the CAD files (.dwg) for each of

them in the files attached to this document. If you wish to open those files on your computer, you can use AutoCAD, which has free educational licenses for EPFL students.

Parameter 2) says how many stories or floors the building has.

Parameters 3) and 4) say the design factors of the building. You can ignore them for this project.

Parameter 5) says the soil type on which the building will be placed. For a quick reference, a soil type A is a good quality soil, while a soil type D is a bad soil and riskier for the building. Therefore, a building on soil type D should have "stronger" structural walls to not collapse.

Parameter 6) says the intended use of the building. Since it is the same for all the 201 buildings in the database, you can ignore this parameter for this project.

Parameter 7) says the seismic zone where the building will be placed. It depends on the location and city. For reference, a Zone 3 is considered riskier than a Zone 1 regarding earthquakes. Therefore, a building in Zone 3 should have "stronger" structural walls so as to not collapse.

Parameter 8) says the quality of the timber (wood) used to build the building. Since it is the same for all the 201 buildings in the database, you can ignore this parameter for this project.

Parameter 9) says the type of connections that the structural walls use in the building. For this project, we have only two types: HD which stands for "hold-down", and ATS which stands for "anchor tiedown system". I will explain this in more detail later.

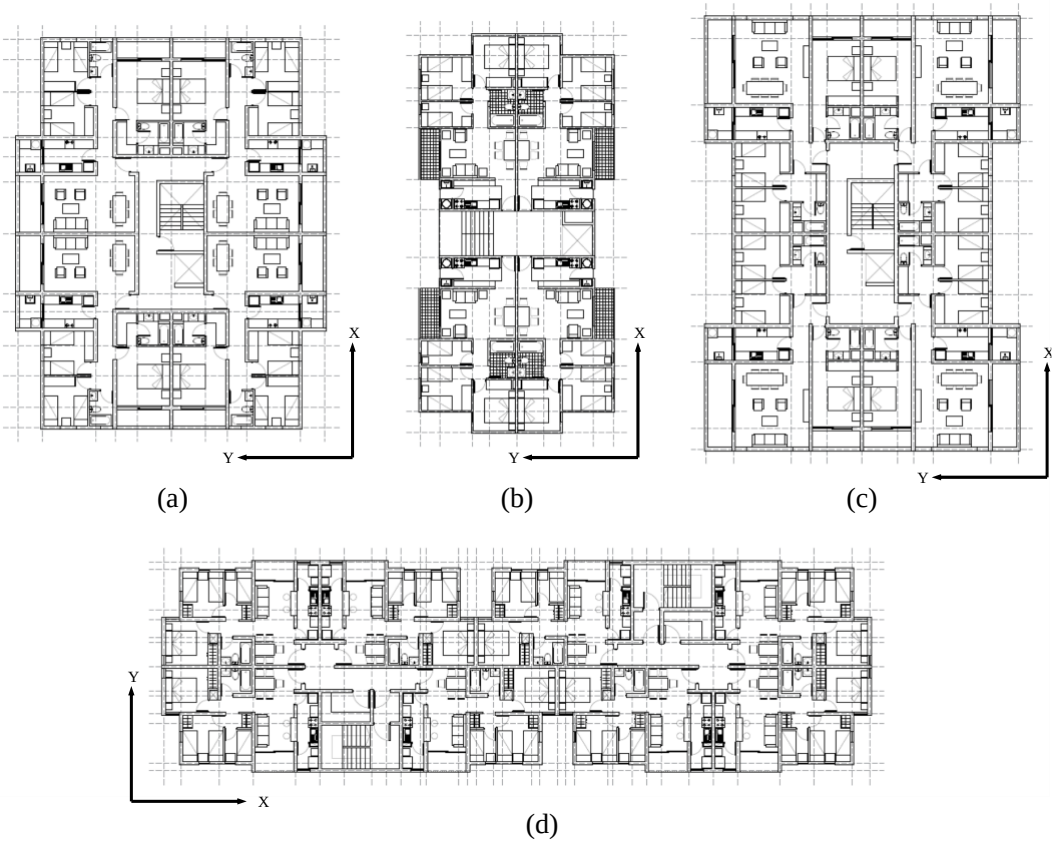


Figure 2. Architectural configurations of the buildings: (a) configuration type "Q", configuration type "P", configuration type "C", and configuration type "D".

Now, the reason there are eight Excel files is that the buildings have been grouped by the architectural configurations and the connection types. Since there are 4 architectural configurations and 2 connection types, there are eight Excel files. Besides, inside each Excel file,

the buildings have been grouped by the number of stories. Therefore, you will be able to find several sheets inside each Excel file; in some cases, they have 3, 4, and 5 stories, and in some only 5 stories. It depends on each case.

In summary, the information Type A that is relevant to extract for every building from the headers in the Excel files are:

2) Number of stories or floors.
5) Soil type.
7) Seismic zone.
9) *Connection* system.

Those four parameters constitute the information Type A that will be used as input for training the ML model. Next, information Type B can be found inside the Excel files, as Figure 3 shows.

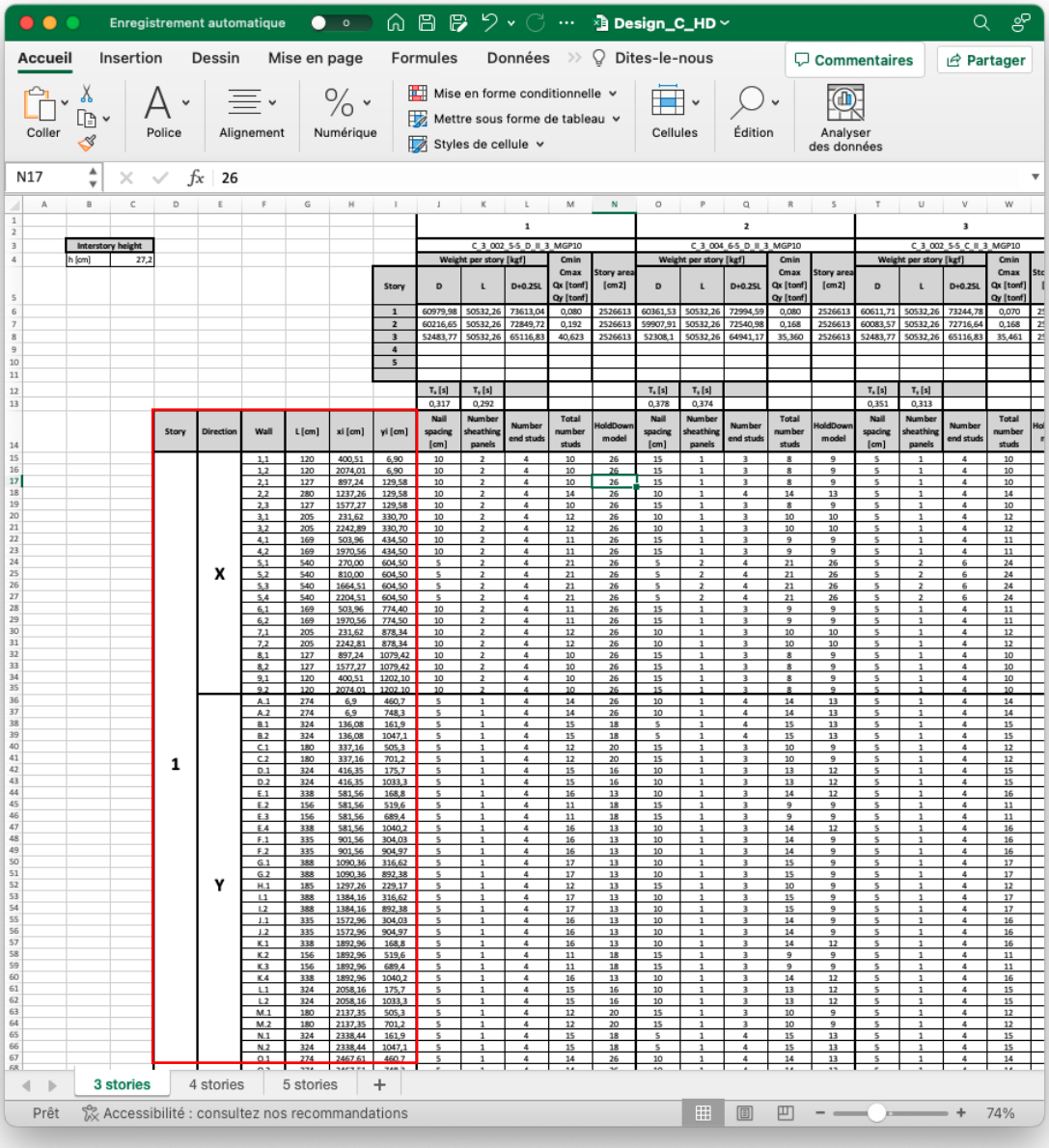


Figure 3. Architectural information Type B.

Information Type B tells you the spatial location and dimensions of each of the structural walls in the building. The first column **Story** tells in which story or floor the wall is located. The second column **Direction** tells if the wall is oriented in the X or Y direction regarding the main orientation of the building. You can see the orientation in Figure 2. The third column **Wall** tells you the ID of each wall. You can see that the walls in the X direction have a different ID compared to the ones in the Y direction. You can graphically see this ID in the CAD (.dwg) files that are provided; I recommend you to check those files so that you will understand this concept better. The fourth column **L [cm]** tells you the length of the walls, either in the X or Y direction. Finally, the columns fifth and sixth **xi [cm]** and **yi [cm]** tell you the X and Y location of the wall, measured from the lower left corner of the full building plan to the geometric center of the wall.

There are two additional important Type B information to be considered; please check Figure 4. The column **D+0.25L** tells you the total weight calculated for each story, taking into account everything that the building will have to bear: furniture, people, and walls, among others. The column **Story area** tells you the area of the floor; please note that this area will be the same in all the buildings that use the same architectural configuration.

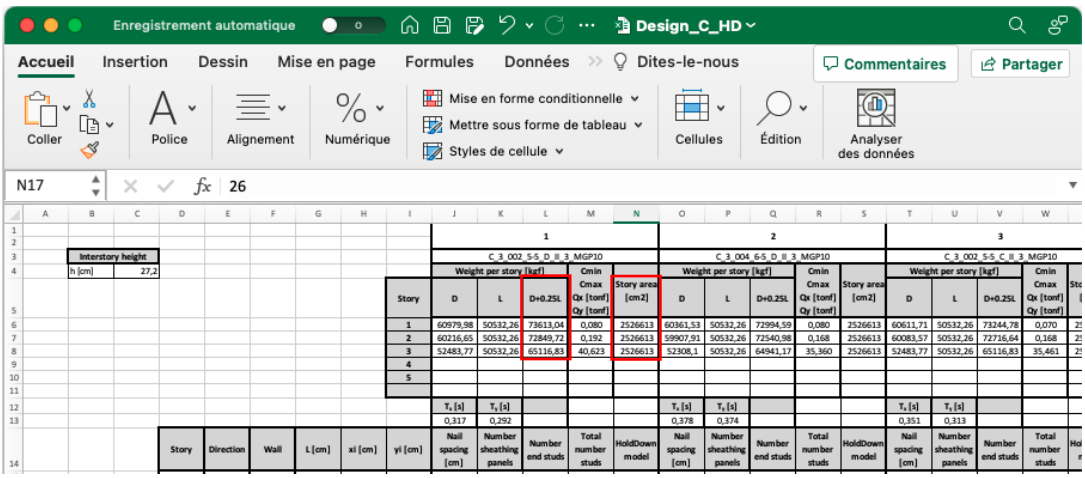
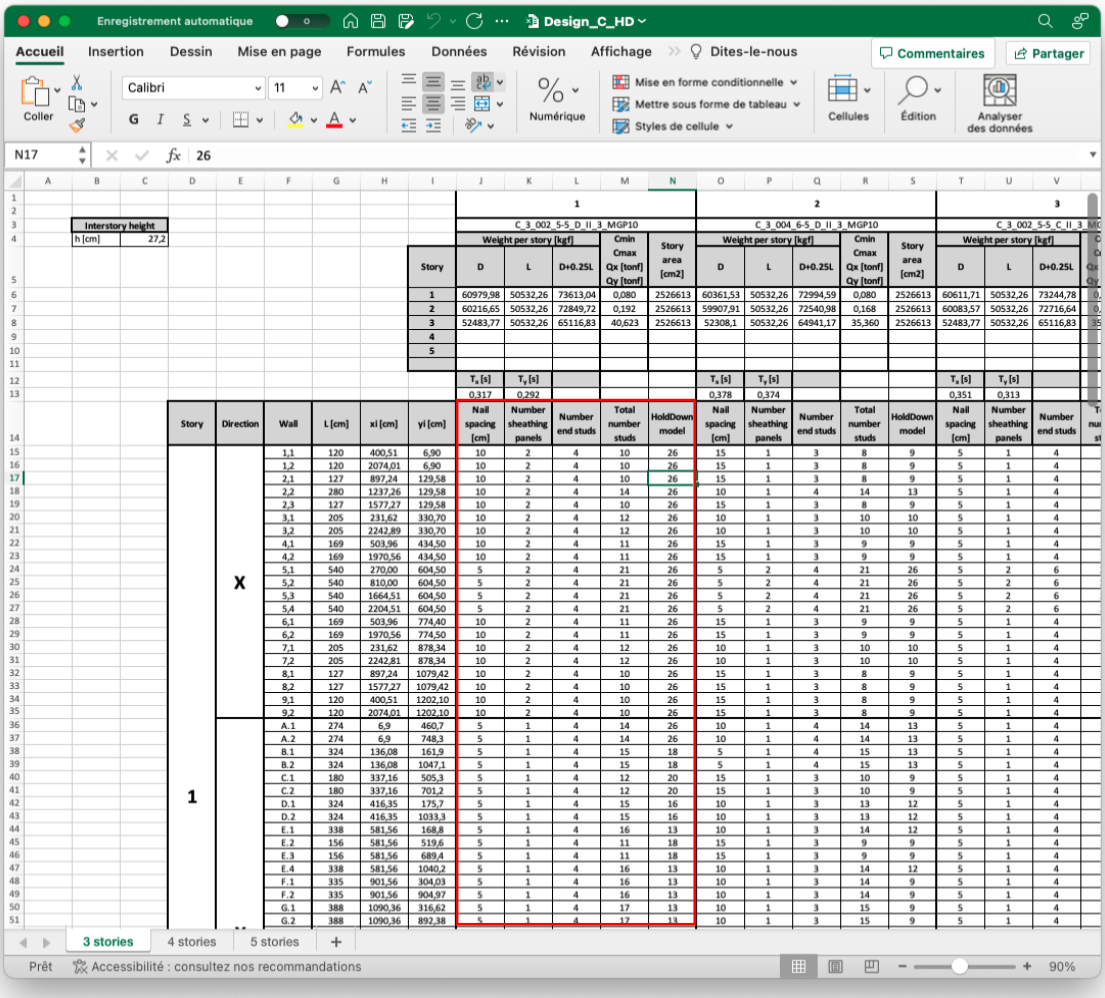


Figure 4. Information about the weight and the area of the buildings.

Now that we have discussed all information Type A and B, we can check information Type C, i.e., the details about the structural walls. You can see where this information is located inside of the Excel files in Figure 5. Besides, for you to understand the physical meaning of the information, please check Figure 6. The column **Nail Spacing** tells how much every nail should be spaced from each other in the wall. The less the spacing, the more nails you have, and therefore, the stronger the wall. The **Number sheathing panels** column tells if you need to install sheathing panels on both sides of the wall. If you install panels on the two sides, the wall will be stronger. The column **Number end studs** tells how many studs you need to install on the left and right corners of the wall, as you can see in Figure 6. The more studs you have, the stronger the wall. The column **Total number studs** tell how many studs you have in total in the wall, counting inner and end studs. The more studs, the stronger the wall. The column **HoldDown model** tells the type of connection that you need to use for the wall. Note that for the buildings that use the ATS system, that column header will say **ATS diameter**. This is because the HD system and the ATS system are different and require different information. To understand this, check Figure 7. If you would like additional information about the different models of connection systems and their meaning, please check the files "Details_HoldDown.xlsx" and "Details_ATS.xlsx". However, you do not need to understand this information in detail to carry out the project.



Interstory height	
h [cm]	27,2

	1					2					3				
	C_3_002_5-5_D_II_3_MGP10					C_3_004_6-5_D_II_3_MGP10					C_3_002_5-5_C_II_3_MG				
	Weight per story [kgf]			Cmin Cmax Qx [tonf] Qy [tonf]	Story area [cm2]	Weight per story [kgf]			Cmin Cmax Qx [tonf] Qy [tonf]	Story area [cm2]	Weight per story [kgf]				
Story	D	L	D+0.25L			D	L	D+0.25L			D	L	D+0.25L		
1	60979,98	50532,26	73613,04	0,080	2526613	60361,53	50532,26	72994,59	0,080	2526613	60611,71	50532,26	73244,78		
2	60216,65	50532,26	72849,72	0,192	2526613	59907,91	50532,26	72540,98	0,168	2526613	60083,57	50532,26	72716,64		
3	52483,77	50532,26	65116,83	40,623	2526613	52308,1	50532,26	64941,17	35,360	2526613	52483,77	50532,26	65116,83		
4															
5															
	T_x [s]	T_y [s]				T_x [s]	T_y [s]				T_x [s]	T_y [s]			
	0,317	0,292				0,378	0,374				0,351	0,313			

Story	Direction	Wall	L [cm]	xi [cm]	yi [cm]	Nail spacing [cm]	Number sheathing panels	Number end studs	Total number studs	HoldDown model	Nail spacing [cm]	Number sheathing panels	Number end studs	Total number studs	HoldDown model	Nail spacing [cm]	Number sheathing panels	Number end studs
1	X	1,1	120	400,51	6,90	10	2	4	10	26	15	1	3	8	9	5	1	4
		1,2	120	2074,01	6,90	10	2	4	10	26	15	1	3	8	9	5	1	4
		2,1	127	897,24	129,58	10	2	4	10	26	15	1	3	8	9	5	1	4
		2,2	280	1237,26	129,58	10	2	4	14	26	10	1	4	14	13	5	1	4
		2,3	127	1577,27	129,58	10	2	4	10	26	15	1	3	8	9	5	1	4
		3,1	205	231,62	330,70	10	2	4	12	26	10	1	3	10	10	5	1	4
		3,2	205	2242,89	330,70	10	2	4	12	26	10	1	3	10	10	5	1	4
		4,1	169	503,96	434,50	10	2	4	11	26	15	1	3	9	9	5	1	4
		4,2	169	1970,56	434,50	10	2	4	11	26	15	1	3	9	9	5	1	4
		5,1	540	270,00	604,50	5	2	4	21	26	5	2	4	21	26	5	2	6
		5,2	540	810,00	604,50	5	2	4	21	26	5	2	4	21	26	5	2	6
		5,3	540	1664,51	604,50	5	2	4	21	26	5	2	4	21	26	5	2	6
		5,4	540	2204,51	604,50	5	2	4	21	26	5	2	4	21	26	5	2	6
		6,1	169	503,96	774,40	10	2	4	11	26	15	1	3	9	9	5	1	4
		6,2	169	1970,56	774,50	10	2	4	11	26	15	1	3	9	9	5	1	4
		7,1	205	231,62	878,34	10	2	4	12	26	10	1	3	10	10	5	1	4
		7,2	205	2242,81	878,34	10	2	4	12	26	10	1	3	10	10	5	1	4
		8,1	127	897,24	1079,42	10	2	4	10	26	15	1	3	8	9	5	1	4
		8,2	127	1577,27	1079,42	10	2	4	10	26	15	1	3	8	9	5	1	4
		9,1	120	400,51	1202,10	10	2	4	10	26	15	1	3	8	9	5	1	4
		9,2	120	2074,01	1202,10	10	2	4	10	26	15	1	3	8	9	5	1	4
		A.1	274	6,9	460,7	5	1	4	14	26	10	1	4	14	13	5	1	4
		A.2	274	6,9	748,3	5	1	4	14	26	10	1	4	14	13	5	1	4
		B.1	324	136,08	161,9	5	1	4	15	18	5	1	4	15	13	5	1	4
		B.2	324	136,08	1047,1	5	1	4	15	18	5	1	4	15	13	5	1	4
		C.1	180	337,16	505,3	5	1	4	12	20	15	1	3	10	9	5	1	4
		C.2	180	337,16	701,2	5	1	4	12	20	15	1	3	10	9	5	1	4
		D.1	324	416,35	175,7	5	1	4	15	16	10	1	3	13	12	5	1	4
		D.2	324	416,35	1033,3	5	1	4	15	16	10	1	3	13	12	5	1	4
		E.1	338	581,56	168,8	5	1	4	16	13	10	1	3	14	12	5	1	4
		E.2	156	581,56	519,6	5	1	4	11	18	15	1	3	9	9	5	1	4
		E.3	156	581,56	689,4	5	1	4	11	18	15	1	3	9	9	5	1	4
		E.4	338	581,56	1040,2	5	1	4	16	13	10	1	3	14	12	5	1	4
		F.1	335	901,56	304,03	5	1	4	16	13	10	1	3	14	9	5	1	4
		F.2	335	901,56	904,97	5	1	4	16	13	10	1	3	14	9	5	1	4
		G.1	388	1090,36	316,62	5	1	4	17	13	10	1	3	15	9	5	1	4
		G.2	388	1090,36	892,38	5	1	4	17	13	10	1	3	15	9	5	1	4

Figure 5. Information Type C for a given building inside the Excel file.

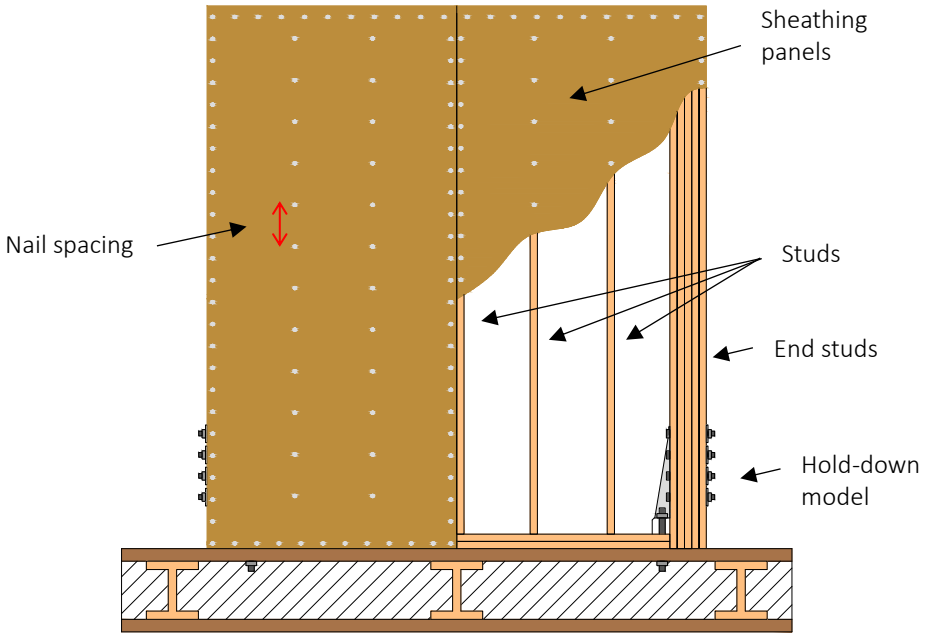


Figure 6. Physical meaning of information Type C for structural walls.

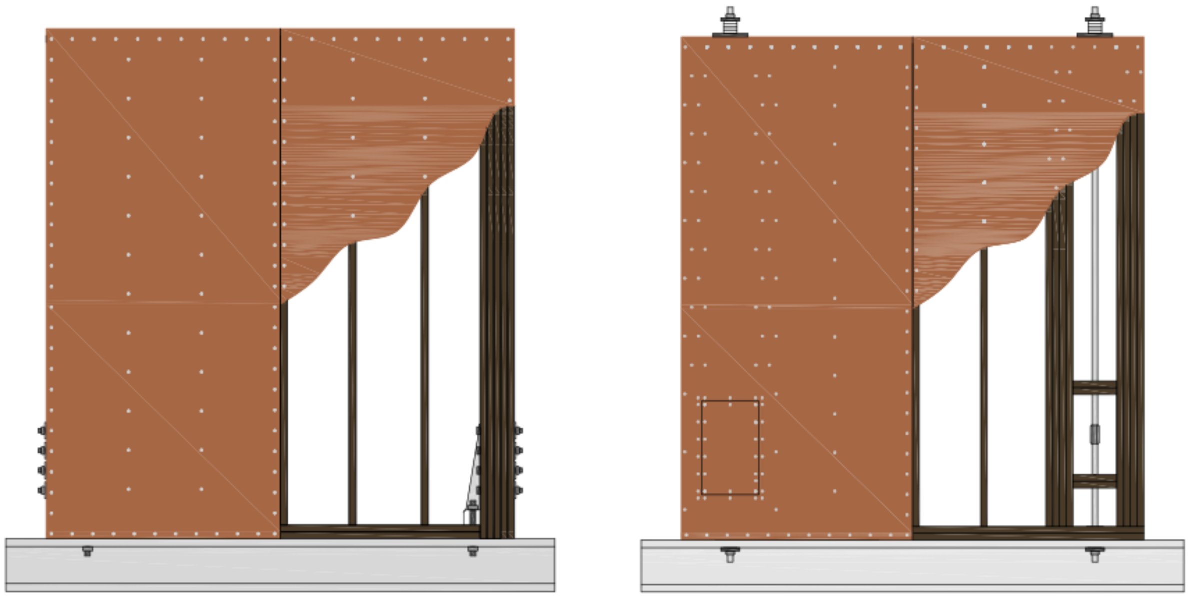


Figure 7. Difference between a connection system HD (left) and a systems ATS (right).

The last Type C information that is relevant for this project is the fundamental periods Tx and Ty of the building for the directions X and Y, respectively. They give information about the dynamic properties of the building and are useful to understand how the building will react during an earthquake, for example. For every building, you can find that information above the headers, as shown in Figure 8.

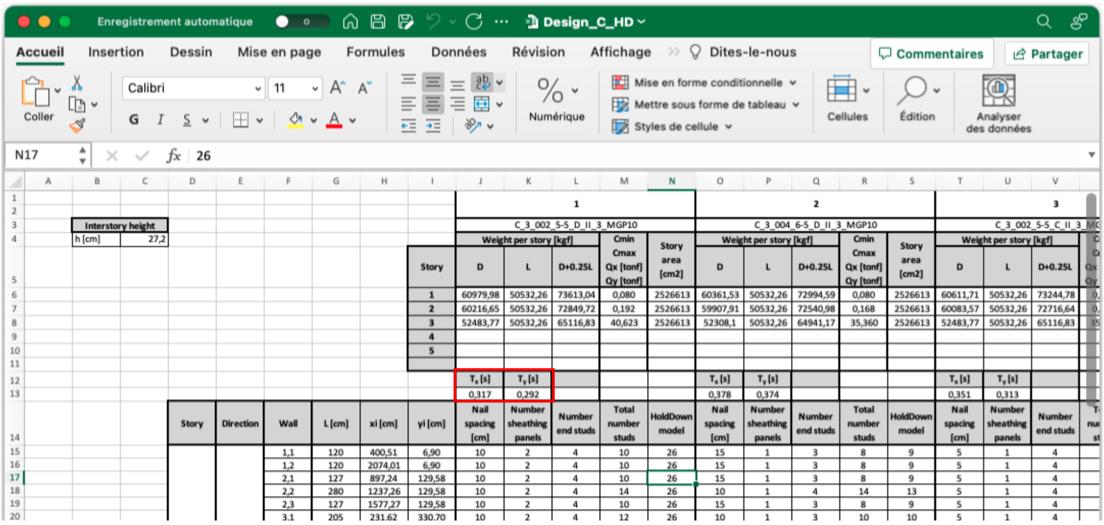
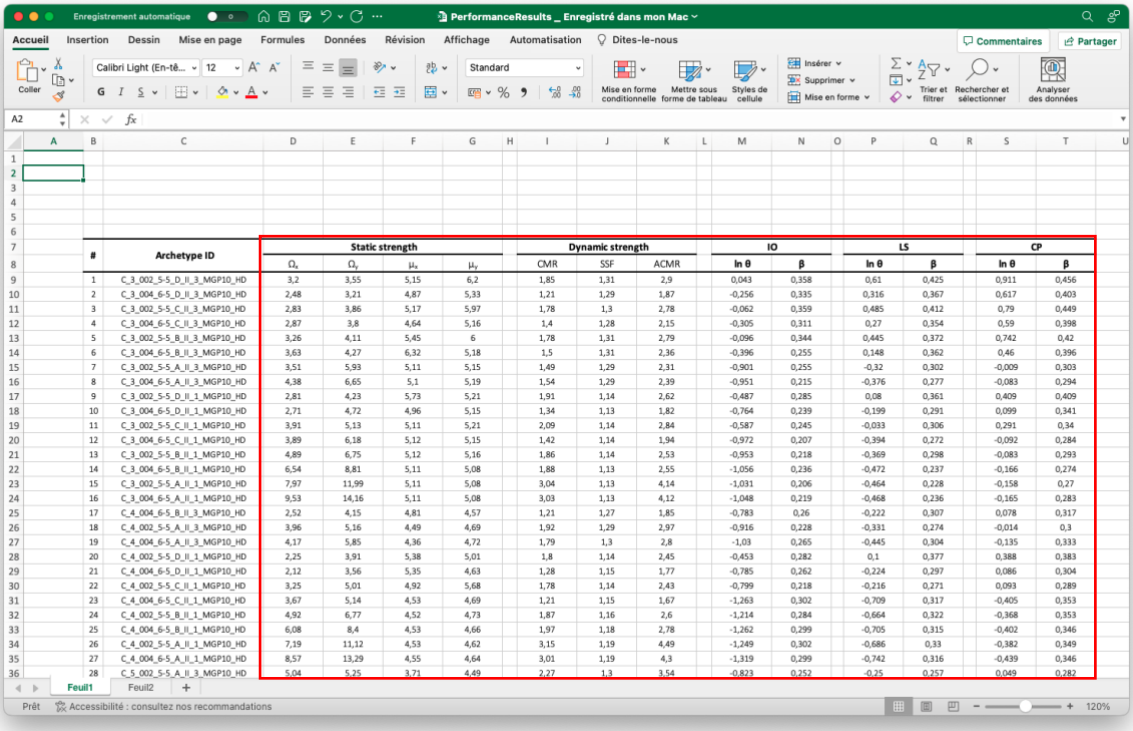


Figure 8. Fundamental periods of each building.

Finally, the last relevant information is Type D. As explained before, this information tells the user about the safety and performance of the building. It was obtained by applying mathematical models to the information Type A, B, and C, getting valuable information about it. If you would like to know more about such mathematical models, you can check this reference. However, given that this is a highly complex engineering topic, you do not need to fully understand how those values were calculated, you only need to understand that the values represent the safety of the building, and that if you compare two similar buildings, the one with the higher values is considered safer. This information is included in the Excel file "PerformanceResults_xlsx", and Figure 9 explains the relevant columns.



#	Archetype ID	Static strength				Dynamic strength			IO		LS		CP	
		Ω_x	Ω_y	μ_x	μ_y	CMR	SSF	ACMR	ln θ	β	ln θ	β	ln θ	β
1	C_3_002_5-5_D_II_3_MGP10_HD	3,2	3,55	5,15	6,2	1,85	1,31	2,9	0,043	0,358	0,61	0,425	0,911	0,456
2	C_3_004_6-5_D_II_3_MGP10_HD	2,48	3,21	4,87	5,33	1,21	1,29	1,87	-0,256	0,335	0,316	0,367	0,617	0,403
3	C_3_002_5-5_C_II_3_MGP10_HD	2,83	3,86	5,17	5,97	1,78	1,3	2,78	-0,062	0,359	0,485	0,412	0,79	0,449
4	C_3_004_6-5_C_II_3_MGP10_HD	2,87	3,8	4,64	5,16	1,4	1,28	2,15	-0,305	0,311	0,27	0,354	0,59	0,398
5	C_3_002_5-5_B_II_3_MGP10_HD	3,26	4,11	5,45	6	1,78	1,31	2,79	-0,096	0,344	0,445	0,372	0,742	0,42
6	C_3_004_6-5_B_II_3_MGP10_HD	3,63	4,27	6,32	5,18	1,5	1,31	2,36	-0,396	0,255	0,148	0,362	0,46	0,396
7	C_3_002_5-5_A_II_3_MGP10_HD	3,51	5,93	5,11	5,15	1,49	1,29	2,31	-0,901	0,255	-0,32	0,302	-0,009	0,303
8	C_3_004_6-5_A_II_3_MGP10_HD	4,38	6,65	5,1	5,19	1,54	1,29	2,39	-0,951	0,215	-0,376	0,277	-0,083	0,294
9	C_3_002_5-5_D_II_1_MGP10_HD	2,81	4,23	5,73	5,21	1,91	1,14	2,62	-0,487	0,285	0,08	0,361	0,409	0,409
10	C_3_004_6-5_D_II_1_MGP10_HD	2,71	4,72	4,96	5,15	1,34	1,13	1,82	-0,764	0,239	-0,199	0,291	0,099	0,341
11	C_3_002_5-5_C_II_1_MGP10_HD	3,91	5,13	5,11	5,21	2,09	1,14	2,84	-0,587	0,245	-0,033	0,306	0,291	0,34
12	C_3_004_6-5_C_II_1_MGP10_HD	3,89	6,18	5,12	5,15	1,42	1,14	1,94	-0,972	0,207	-0,394	0,272	-0,092	0,284
13	C_3_002_5-5_B_II_1_MGP10_HD	4,89	6,75	5,12	5,16	1,86	1,14	2,53	-0,953	0,218	-0,369	0,298	-0,083	0,293
14	C_3_004_6-5_B_II_1_MGP10_HD	6,54	8,81	5,11	5,08	1,88	1,13	2,55	-1,056	0,236	-0,472	0,237	-0,166	0,274
15	C_3_002_5-5_A_II_1_MGP10_HD	7,97	11,99	5,11	5,08	3,04	1,13	4,14	-1,031	0,206	-0,464	0,228	-0,158	0,27
16	C_3_004_6-5_A_II_1_MGP10_HD	9,53	14,16	5,11	5,08	3,03	1,13	4,12	-1,048	0,219	-0,468	0,236	-0,165	0,283
17	C_4_004_6-5_B_II_3_MGP10_HD	2,52	4,15	4,81	4,57	1,21	1,27	1,85	-0,783	0,26	-0,222	0,307	0,078	0,317
18	C_4_002_5-5_A_II_3_MGP10_HD	3,96	5,16	4,49	4,69	1,92	1,29	2,97	-0,916	0,228	-0,331	0,274	-0,014	0,3
19	C_4_004_6-5_A_II_3_MGP10_HD	4,17	5,85	4,36	4,72	1,79	1,3	2,8	-1,03	0,265	-0,445	0,304	-0,135	0,333
20	C_4_002_5-5_D_II_1_MGP10_HD	2,25	3,91	5,38	5,01	1,8	1,14	2,45	-0,453	0,282	0,1	0,377	0,388	0,383
21	C_4_004_6-5_D_II_1_MGP10_HD	2,12	3,56	5,35	4,63	1,28	1,15	1,77	-0,785	0,262	-0,224	0,297	0,086	0,304
22	C_4_002_5-5_C_II_1_MGP10_HD	3,25	5,01	4,92	5,68	1,78	1,14	2,43	-0,799	0,218	-0,216	0,271	0,093	0,289
23	C_4_004_6-5_C_II_1_MGP10_HD	3,67	5,14	4,53	4,69	1,21	1,15	1,67	-1,263	0,302	-0,709	0,317	-0,405	0,353
24	C_4_002_5-5_B_II_1_MGP10_HD	4,92	6,77	4,52	4,73	1,87	1,16	2,6	-1,214	0,284	-0,664	0,322	-0,368	0,353
25	C_4_004_6-5_B_II_1_MGP10_HD	6,08	8,4	4,53	4,66	1,97	1,18	2,78	-1,262	0,299	-0,705	0,315	-0,402	0,346
26	C_4_002_5-5_A_II_1_MGP10_HD	7,19	11,12	4,53	4,62	3,15	1,19	4,49	-1,249	0,302	-0,686	0,33	-0,382	0,349
27	C_4_004_6-5_A_II_1_MGP10_HD	8,57	13,29	4,55	4,64	3,01	1,19	4,3	-1,319	0,299	-0,742	0,316	-0,439	0,346
28	C_5_002_5-5_A_II_3_MGP10_HD	5,04	5,25	3,71	4,49	2,27	1,3	3,54	-0,823	0,252	-0,25	0,257	0,049	0,282

Figure 9. Type D information for every building.

Now that you have the nig picture of the organization of the information for the project, we can once again summarize the goal:

Employing information types A and B, the ML model should be able to forecast information C and D to speed up the design process of buildings. Given that this is still an exploratory project, your work will also consist of identifying the best ways to treat data and results. If any additional information or clarifications are required, do not hesitate to contact Xavier Estrella by email (edisson.estrellaarcos@epfl.ch).