EFFICIENCY OF INTENSITY MEASURES FOR SEISMIC RESPONSE PREDICTION IN CLT BUILDINGS VIA DATA SCIENCE METHODS

Conference Paper · August 2021

CITATIONS
0

READS
54

4 authors, including:

Sebastian Aedo-Maluje
University School for Advanced Studies IUSS Pavia
3 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Christian Málaga-Chuquitaype
Imperial College London
96 PUBLICATIONS **779** CITATIONS
SEE PROFILE

Jorge Macedo
Georgia Institute of Technology
70 PUBLICATIONS **285** CITATIONS
SEE PROFILE

WCTE 2020
WORLD CONFERENCE ON TIMBER ENGINEERING
SANTIAGO, CHILE | 24 - 27 AUGUST

EFFICIENCY OF INTENSITY MEASURES FOR SEISMIC RESPONSE PREDICTION IN CLT BUILDINGS VIA DATA SCIENCE METHODS

Sebastián Aedo Maluje[1], Christian Málaga-Chuquitaype[2], Jorge Macedo Escudero[3], Farahnaz Soleimani[4]

ABSTRACT: This paper investigates the influence of different ground-motion intensity measures on the seismic response of cross-laminated timber (CLT) buildings using advanced data science methods. To this end, 9 CLT buildings with different structural arrangements, such as the number of storeys and behaviour factors, are modelled and subjected to 1654 acceleration records through nonlinear response–history analysis. From these analyses, inter–storey drifts, floor accelerations, and storey shears are determined and used as engineering demand parameters (EDPs). Initially, regular regression analyses are conducted on the assembled EDPs to explore preliminary relationships between the EDPs and several intensity measures. Next, machine learning techniques are applied, using robust algorithms such as stepwise selection and Lasso analyses, where 50% of the data is utilised for training purposes and the remaining 50% for model validation. Eventually, the intensity measures which can properly predict the seismic response of each EDP are identified.

KEYWORDS: Cross-laminated timber, machine learning, engineering demand parameter, ground-motion intensity measures.

1 INTRODUCTION

Advanced data science methodologies like machine learning techniques need more investigation, particularly in the field of structural engineering. In fact, with the amount of data collected nowadays not only from numerical analyses but also from structural monitoring, the prediction of structural and seismic responses can be significantly improved by finding new relationships and patterns that have not been thoroughly investigated yet.

Throughout the last years, different research projects have developed initiatives to promote the design and construction of multi-storey cross-laminated timber buildings. One of the main precursor studies in this topic is the Construction System Fiemme project (SOFIE) [1], in which a series of full-scale experiments were conducted on a 7-storey CLT building. The results were compared to a 3D model, emphasising the high stiffness in-plane and load bearing capacity of the CLT walls and the key role played by the connectors in the seismic response of CLT buildings.

Afterwards, several numerical studies on the seismic response of CLT multi-storey buildings have been performed. For example, Pozza et. al. [2, 3] compared and tested the seismic behaviour of four different typologies of CLT walls and used these results to calibrate three-dimensional numerical models of a 3-storey CLT building, evaluating the influence of these CLT walls on the behaviour factor of the structure. Thereafter, they investigated the seismic behaviour of different kinds of multi-storey CLT buildings paying special attention to their behaviour factor, modelling 24 two-dimensional buildings, with different characteristics, such as the number of storeys, plan size, and the level of fragmentation in CLT panels. It was concluded that the seismic response was correlated with the structural design of the CLT (provisions, methods, etc.), level of fragmentation, and the slenderness of the CLT wall.

Demirci et al. [5, 6] examined the seismic response of multi-storey CLT buildings employing numerical models with various structural arrangements and payed special attention to the influence of the ground-motion frequency content on the inelastic displacement demands of those buildings. The main remarks were related to the significant influence of the period ratio (T_1/T_m) and the modification factor. As a consequence of the long-period in structures, which is common in taller CLT buildings, significant drift demands were expected, particularly in CLT buildings with a higher level of fragmentation, corresponding to more flexible structures. Similarly, CLT structures with larger joint densities were observed to experience more energy dissipation capacity. Due to these findings, the period ratio was proposed as a parameter in the seismic response of CLT buildings. In addition, the shear demands and accelerations were assessed [6]. It was observed that the shear demand increases as the period ratio increases up to a maximum

[1]Sebastián Aedo Maluje, Department of Civil and Environmental Engineering, Imperial College London, United Kingdom, sebastian.aedo-maluje18@alumni.imperial.ac.uk
[2]Christian Málaga-Chuquitaype, Department of Civil and Environmental Engineering, Imperial College London, United Kingdom, c.malaga@imperial.ac.uk
[3] Jorge Macedo Escudero, School of Civil and Environmental Engineering, GeorgiaTech, USA, jorge.macedo@gatech.edu
[4] Farahnaz Soleimani, School of Civil and Environmental Engineering, GeorgiaTech, USA, soleimani@gatech.edu

value and, then, after this upper limit, the demand decreases. The shear demand was also related to the behaviour factor *q* and panel fragmentation level *m*, and lower shear demands were observed for higher values of *q* and *m*. In terms of accelerations, whilst smaller values of the period ratio (T_1/T_m < 1.0) are associated with lower accelerations, higher values of the period ratio and, also, an increasing number of floors reflected an increasing value of acceleration. In both studies, Demirci et al. [4, 5], a relationship between inter-storey drift, acceleration and storey shear as a function of the period ratio, the behaviour factor, the level of panel fragmentation and the slenderness of the CLT wall was proposed with the period ratio featuring prominently in them.

Employing advanced data science methodologies, this paper deals with the prediction of engineering demand parameters (EDPs) of multi-storey timber buildings subjected to earthquake loading. Particularly, the intensity measures (IMs) with the highest influence on the EDPs are detected. As such, this paper combines two emerging areas of research: timber engineering and data science.

2 METHODOLOGY

2.1 BUILDING MODELS AND SEISMIC DESIGN

A tall CLT structure representing a hotel designed according to Eurocodes is taken as a case study. In order to generate the required database, three different heights (*n* = 6, 8, and 12 storeys, where *n* stands for the number of floors over the ground floor) and behaviour factors (q = 1.5, 2.0, and 3.0) are examined. This results in nine combinations for building features. Building upon the research of Demirci [4, 5] and taking into account the industry's interest for using long panel configurations, a level of panel fragmentation equal to 0 is considered herein, ie. *m* = 0 (where *m* is the number of vertical joint lines), which means each cross-laminated timber (CLT) wall is modelled as one piece with a length equal to 8.5 m without any internal cut, as shown in Figure 1, where an 8 storey building is illustrated.

To model the building, CLT panels of C24 timber class are selected, where the following characteristic properties are used: bending strength equal to 24.0 MPa, tensile strengths parallel and perpendicular to the grain are 16.5 MPa and 0.12 MPa, respectively, the compressive strengths parallel and perpendicular to the grain are 21.0 MPa and 2.7 MPa, respectively, shear strength is 2.7 MPa, and the rolling shear strength is 1.2 MPa, Young's modulus parallel to the grain is assumed to be 12,000 MPa, the shear modulus is 690 MPa, and the rolling shear modulus is 50 MPa, and the density mean value is 420.0 kg/m^3, with a Poisson's ratio of 0.25. Based on these properties, the design value has been calculated using Clause 2.4.3 of Eurocode 5 [6], where the modification factor is equal to 1.1 for instantaneous action such as seismic loads, and the partial factor for a material property is equal to 1.25 for glued laminated timber and 1.0 for accidental combinations, such as seismic loads. Permanent loads coming from the CLT slabs, CLT walls, and finishings were considered. For superimposed loads, the category of use A was taken into account, as defined by Tables 6.1 and 6.2 of Eurocode 1 [8] for a hotel.

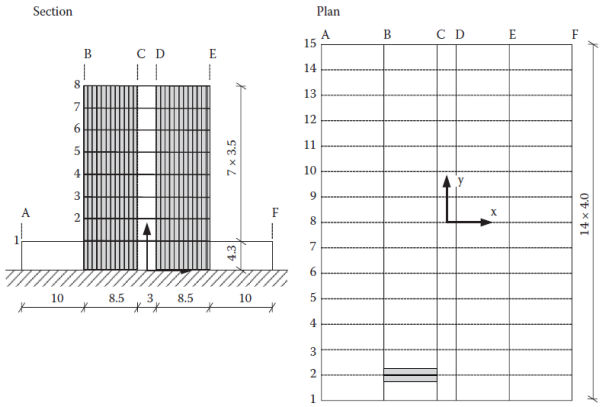

Figure 1: *Building layout based on 8 storeys. Ref.: Málaga-Chuquitaype & Elghazouli [6].*

According to Clause 3.4.2(2) of Eurocode 8 [9], the load combination considered in the seismic analysis is based on the 100% of the total dead load and 30% of the total imposed load. Seismic analysis is performed only in the x-direction, where the following parameters are considered: importance class II, according to Table 4.3 Eurocode 8 [9]; type 1 elastic response spectra, which is related to high seismic areas with surface-wave magnitude (M_s) over 5.5; soil type C, which means a soil factor equals to 1.15 with T_B = 0.2 s, T_C = 0.6 s, and T_D = 2.0 s, where T_B, T_C, and T_D are values of the periods that define the shape of the spectrum and depend on the soil; viscous damping ratio equals to 5%; reference peak ground acceleration on type A ground equals to 3.0 m/s^2; behaviour factors 1.5, 2.0 and 3.0.

Afterwards, since the building layout fulfils the regularity requirements of Eurocode 8 [9], the lateral force method is used for the design of the CLT walls. In addition, torsional effects have been taken into account, as stated in Clause 4.3.3.2.4 of Eurocode 8 [9].

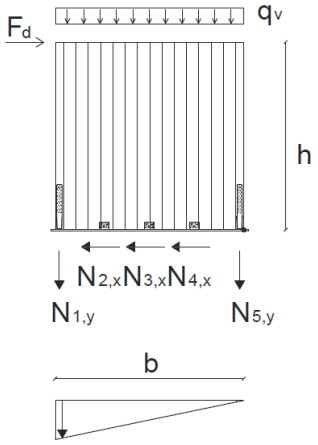

Figure 2: *Model D2 presented by Gavric and Popovski. Ref.: Gavric, I., Popovski, M. [10].*

The gravity load from each floor is calculated using the traditional procedure, ie. static equilibrium, load paths

and tributary areas. The bending moment in each single CLT wall is calculated, based on the sketch presented in Figure 2, where F_d is the lateral load and q_v is the gravity load. Following this sketch (Figure 2), static equilibrium equations can be formulated. This model was presented by Gavric and Popovski [10], and assumes that the uplift forces are fully taken by the hold-down connector, and the shear forces are taken by the shear bracket connector. From these design loads, each connector can be defined using a commercial catalogue. For this purpose, the Rothoblaas catalogue [11] was employed.

2.2 NUMERICAL MODELLING

Structural models were developed in OpenSees [12]. Each CLT building was modelled in two dimensions while considering geometric and joint nonlinearities. Figure 3 shows an idealisation of the structure, where each node is modelled with two degrees of freedom, one in the x-direction (horizontal) and one in the z-direction (vertical). A mass is assigned and distributed node by node, according to the seismic mass defined in Section 2.1 above. The support conditions were defined as a node in the base of the ground floor wall and assigned a pinned condition to be later connected to the wall through a 'link element'.

The structure is subjected to seismic loads, and as a result, a ductile mechanism of failure is expected, which is induced through the connections and their nonlinear deformations [4]. The joints play a key role in this behaviour because of the intrinsic brittle mechanism of failure of timber. Therefore, CLT panels are modelled by means of an elastic and isotropic material using properties described in Section 2.1. On the other hand, connections are assigned a hysteretic material model.

Six different kinds of connectors are employed, one shear bracket connector and five hold-down connectors. To model the hysteretic behaviour of these connectors, information from the literature and Rothoblaas catalogue [4, 11, 13] is used.

The CLT walls are modelled as linear elastic quad elements. In the case of the connectors, a two-node link element is defined with zero-length for both shear brackets and hold-downs. Even though the behaviour of each connector is based on the model D2 (Figure 2) proposed by Gavric & Popovski [10], each link element has two degrees of freedom by definition, which means the response in both directions has to be defined. Therefore, the horizontal response in the hold-down connector and the vertical response in the angle bracket connector are modelled with high values of deformation and low values in terms of forces, resulting in a model that does not transfer significant forces in these directions.

Since OpenSees restricts the definition of the geometric transformation of Quad elements to be linear, a leaning column was modelled next to each CLT wall along the height of the structure to simulate global geometric nonlinearities. To this end, two computational domains and a corotational transformation are employed for the leaning column. To model the boundary condition, a pinned connection is used in the base of the column, and equal lateral displacements are defined at each level between the CLT panels and the column through EdofMP constrains, which imposes an equal displacement at each level between both elements.

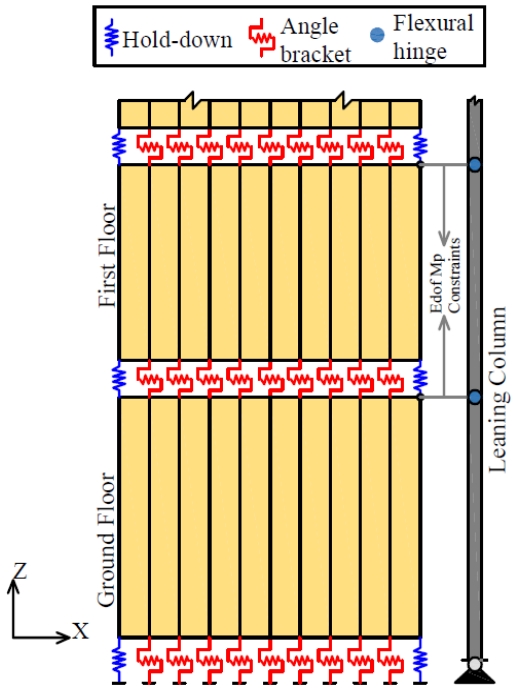

Figure 3: *Idealisation of the structure. Adapted and modified from Demirci et al. [5].*

The gravity load is determined according to the values calculated and shown in Section 2.1 and is applied at the column in each level as a point load. On the other hand, lateral loads are defined as a uniform excitation load pattern, which reflects a seismic base motion through a time series path in the x-direction.

Due to the nature of the seismic records, a nonlinear response-history analysis is performed, assuming a damping ratio of 5% of the critical value. A set of 1654 seismic records coming from 51 earthquake events is selected from the NGA–West1 ground motion database, resulting in 14,886 nonlinear dynamic analyses. The analyses were performed in the DesignSafe-CI platform [14]. In addition, the energy increment test is used as the convergence criteria, with a tolerance equals to $1 \cdot 10^{-6}$ and 25 iterations, reaching an overall convergence of 91.86%, ie. 13,675 analyses. The Newton – Raphson algorithm is selected to solve the nonlinear problem and the Newmark method is used to integrate each seismic record with $\gamma = 0.5$ and $\beta = 0.25$, where γ and β are parameters of the method related to the variation of acceleration, stability of the model and accuracy. To investigate the seismic response of timber buildings, acceleration, drifts and storey shear values are recorded at each level.

3 RESULTS AND DISCUSSION

3.1 INITIAL ASSESSMENT OF THE SEISMIC RESPONSE ESTIMATION USING REGRESSION ANALYSIS

This study aims to identify the relationship between fourteen ground motion intensity measure parameters (IMs) and six engineering demand parameters (EDPs). In this context, the following IMs are selected: the peak ground acceleration (PGA), peak ground velocity (PGV), the peak ground displacement (PGD), the significant duration between 5% and 75% of the total Arias intensity ($D_{5\text{-}75}$), the significant duration between 5% and 95% of the total Arias intensity ($D_{5\text{-}95}$), the mean period of the ground (T_m), the total Arias intensity (I_a), the spectral acceleration (S_a) at periods of T = 0.2s, T = 0.3s, T = T_1, T = 1.0s, T = 3.0s, and T = 5.0s, and the ratio between the fundamental building period and the mean period of the ground motion (T_1/T_m). In terms of the seismic responses, the following EDPs are chosen: the maximum inter-storey drift (ID_{max}), peak acceleration (Acc_{max}) and peak shear force (S_{max}) along the height of the building, the maximum inter-storey drift (ID_{roof}) and peak acceleration (Acc_{roof}) at the top of the building, and the maximum base shear at the bottom of the building (S_{base}).

After cleaning the data obtained from the models that do not reach the convergence criteria and parameters with lower variability along the different seismic records, the following functional form in equation (1) is used to perform an initial assessment of the seismic response estimation using regression analysis:

$$\ln EDP_k = \beta_0 + \sum_j \beta_j \cdot \ln(IM_j) + \sum_j \alpha_j \cdot (\ln(IM_j))^2 \quad (1)$$

where EDP_k is the k-th (k=1,2,…,6) engineering demand parameter, IM_j is the j-th (j=1,2,…,14) intensity measure, β_o is the intercept, and β_j and α_j are the j-th estimated coefficients related to the j-th intensity measure.

A regression analysis is executed on each EDP, considering only one intensity measure at a time, resulting in 840 analyses in total. Then, from each regression, the R^2 value is extracted, symbolising how well one engineering demand parameter is explained by one or more intensity measures.

Figures 4, 5, and 6 show a summary of each linear model analysed using the inter-storey drift, acceleration, and storey shear as EDPs, respectively. In each graph, twenty R^2 values per IM are illustrated together with their corresponding error bars, which indicate mean and mean ±1 standard deviation values. In Figure 4, the higher R^2 values correspond to PGD and $Sa_{(T = 5.0\ s)}$ with mean values of 0.832 and 0.758, followed by $Sa_{(T = 3.0\ s)}$ and PGV with mean values of 0.331 and 0.487, respectively. In contrast, the lower R^2 values relate to $D_{5\text{-}75}$ and $D_{5\text{-}95}$, with mean values of 0.030 and 0.032, followed by $Sa_{(T = 0.2\ s)}$ and $Sa_{(T = 0.3\ s)}$, with mean values of 0.052 and 0.067, respectively. In other words, if we used the peak ground displacement as a predictor, the inter-storey drift would be explained by PGD with 83.2% certainty. A higher value of the peak ground displacement is related to structure with longer periods, as expected. However, higher values associated with pseudo-acceleration response at higher values of the period (e.g. $Sa_{(T=3.0s)}$ and $Sa_{(T=5.0s)}$) are not usual and can be due to significant dispersion of data.

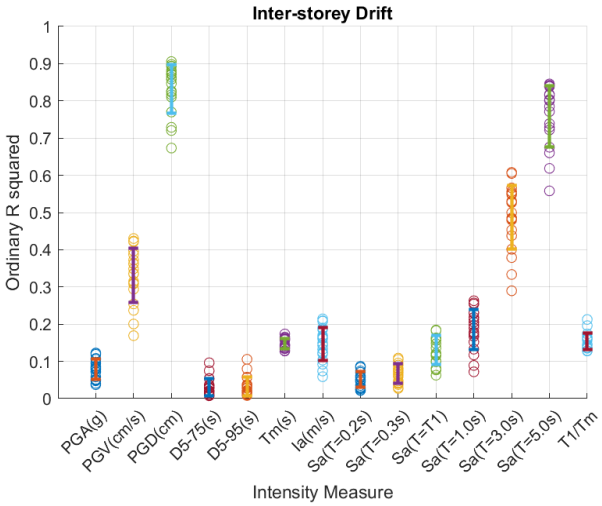

Figure 4: *Inter-storey drift – R^2 value vs each individual intensity measure.*

In Figure 5, the higher R^2 values are reached by PGA and $Sa_{(T = 0.2\ s)}$ with mean values of 0.993 and 0.930, followed by $Sa_{(T = 0.3\ s)}$ and I_a with mean values of 0.929 and 0.929, respectively. In contrast, lower R^2 values are presented by T_m and T_1/T_m, with mean values of 0.161 and 0.157, followed by PGD and $Sa_{(T = 5.0\ s)}$, with mean values of 0.240 and 0.276, respectively. In other words, if we used the peak ground acceleration as a variable, the acceleration would predict the seismic response with 99.3% certainty. As expected, the acceleration is related to the peak ground acceleration in both Acc_{max} and Acc_{roof}. Since the PGA is a high – frequency parameter, it should be correlated with the Arias intensity, which is confirmed in Figure 5. Similarly, the values of the pseudo-acceleration response at higher frequencies, or lower periods, are connected to Arias intensity and PGA as well.

Furthermore, in Figure 6 the higher R^2 values are presented by PGA and $Sa_{(T = 0.3\ s)}$, with mean values of 0.985 and 0.926, followed by $Sa_{(T = 0.2\ s)}$ and I_a with mean values of 0.926 and 0.921, respectively. In contrast, the lower R^2 values are presented by T_m and T_1/T_m, with mean values of 0.172 and 0.170, followed by PGD and $Sa_{(T = 5.0\ s)}$, with mean values of 0.224 and 0.262, respectively. In other words, if we used the peak ground acceleration as an input variable in the regression model, the acceleration would predict the seismic response with 98.5% certainty. As known, the acceleration has a significant relation to the storey shear. In fact, not only Eurocode 8 but also international seismic codes use parameters related to the acceleration to express the shear force under seismic loads, hence, it is predictable that Figures 5 and 6 are related.

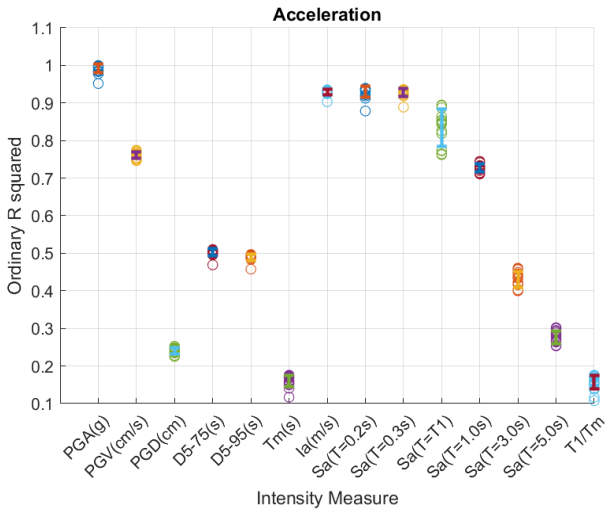

Figure 5: *Acceleration – R^2 value vs each individual intensity measure.*

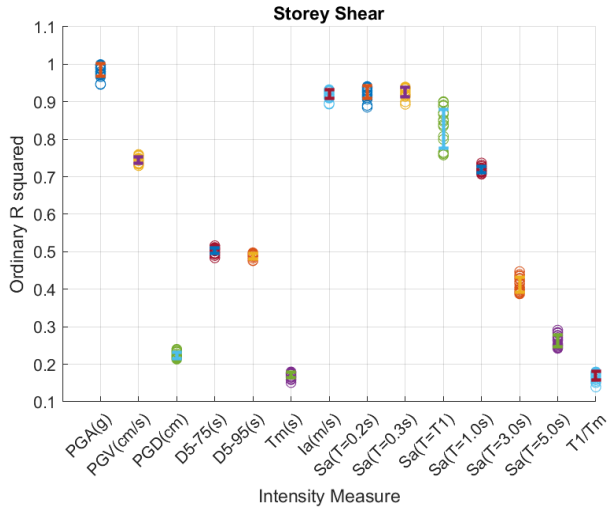

Figure 6: *Storey Shear – R^2 value vs each individual intensity measure.*

Analysing each parameter graphically, a substantial dispersion is observed in the inter-storey drift. This observed dispersion could have consequences in the final model, not only in terms of fitting but also in the aspect of R^2 values, standard deviations, and errors. In comparison with the inter-storey drift, accelerations and storey shear present a better correlation in terms of fitting and dispersion. At this point, a good fit is expected in the model for both accelerations and storey shears, as reflected by their R^2 values, standard deviations, and errors.

3.2 ASSESSMENT OF THE SEISMIC RESPONSE ESTIMATION USING MACHINE LEARNING APPROACHES

Machine learning is the area from Artificial intelligence and Computing Science, which, using statistical tools and algorithms, learns from the training data. With time, as more experience (data) is obtained by the model, its capacity to predict the response will improve. The three main machine learning approaches can be identified as supervised learning, unsupervised learning, and reinforcement learning. In this context, supervised learning is selected herein, where the predictive model has a clear aim. That means given an input (data), a model is sought for a particular and likely output.

As a part of the supervised learning, the database is split into two subsets. The first subset is utilised in the training of the predictive model, which means only a part of the original database is used to build the model, while the second subset is unknown to the model and is used to validate. The percentages of training and testing data are determined based on experience; however, the literature proposes a minimum value of 50% for training data [15]. For each set, the related error and accuracy are calculated. According to James [15], the training (tr_{error}) and test error (ts_{error}) rates are computed by equations (2) and (3) below:

$$tr_{error} = \frac{1}{n} \cdot \sum_{i=1}^{n} I(y_i \neq \hat{y}_i) \quad (2)$$

$$ts_{error} = Ave\left(I(y_0 \neq \hat{y}_0)\right) \quad (3)$$

where n is the sample size; y_i is the i-th output value from the training data; $\hat{y}_i$ is the i-th predicted output value; $I(y_i \neq \hat{y}_i)$ is an indicator, which is equal to 1 if $y_i \neq \hat{y}_i$ and 0 otherwise; y_0 is the 0-th output value from the testing data; $\hat{y}_0$ is the 0-th predicted output value, using the calculated model; $I(y_i \neq \hat{y}_i)$ is an indicator, which is equal to 1 if $y_i \neq \hat{y}_i$ and 0 otherwise; and Ave is the average of this indicator. Consequently, the accuracy of prediction is calculated as one minus this error.

Two different methods are employed to assess the seismic response model. The first is called forward and backward stepwise selection and is related to the subset selection approach, which identifies the best subset of predictors that can predict the seismic response of CLT structures. The second method is called Lasso and is associated with the shrinkage or regularisation approach, where all the predictors are used in the model, but they are also regulated by a parameter, which can make the predictor equal to zero, improving (reducing) the variance of the model. In both cases, the main objective is to select the best group of predictors to estimate the seismic response of CLT structures.

Regarding forward stepwise selection, the data is imported and normalised using the lognormal space, which is divided into input and output. The training and testing samples are defined, starting with 50% each. Then, using the functional form given by equation (1), the forward selection is performed, which uses the deviance as adding criterion in the model. That means the deviance of a model with more parameters is smaller than the deviance of a model with fewer variables. Hence, the algorithm checks in each iteration the difference between the deviance from one model to another. In this process, the inverse of the chi-square cumulative distribution function (CDF) is assumed by the algorithm as the maximum difference of deviance. To this end, one degree of freedom and 0.95 as the

corresponding probability is considered in the chi-square inverse CDF. Then, the regression is performed on each EDP, where the training and test error rates are calculated as well as the accuracy of the prediction. A total of 90 regression models are developed and evaluated.

Figure 7 displays the evolution of the R^2 value for each predicted EDP as a function of the number of variables corresponding to IMs. It is observed that inter-storey drifts (ID_{max} and ID_{roof}) present a similar path as accelerations (Acc_{max} and Acc_{roof}) and storey shears (S_{max} and S_{base}). In the case of the accelerations and storey shears, the aggregation effect of adding more than two predictors in the R^2 value is not significant. For instance, considering two predictors, the acceleration R^2 values are 0.9831 and 0.9989 for Acc_{max} and Acc_{roof}; and the storey shear R^2 values are 0.9714 and 0.9952 for S_{max} and S_{base}, respectively, and these values would not change noticeably if we added three or more variables to the model.

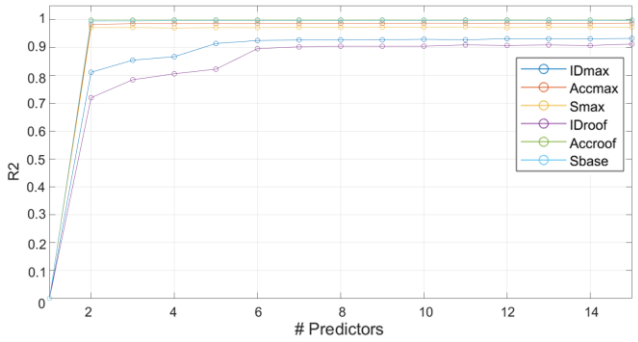

Figure 7: *Forward selection: R^2 value vs number of predictors for each EDP.*

In comparison to accelerations and storey shears, inter-storey drifts present R^2 values of 0.8114 and 0.7202 for ID_{max} and ID_{roof} with two predictors and increase to over 0.9 when using five and seven predictors. Regarding the standard deviation and test error, the values are 0.96 and 0.99; and 8.69% and 11.41% for ID_{max} and ID_{roof}, respectively, using two predictors. Whereas the standard deviation is kept almost constant after two predictors, the test error is decreased to 5.82% and 6.71% in both ID_{max} and ID_{roof}, using five and six predictors. For reasons of simplicity and since the dispersion of the prediction plateaus after two predictors, two predictors are considered in the model.

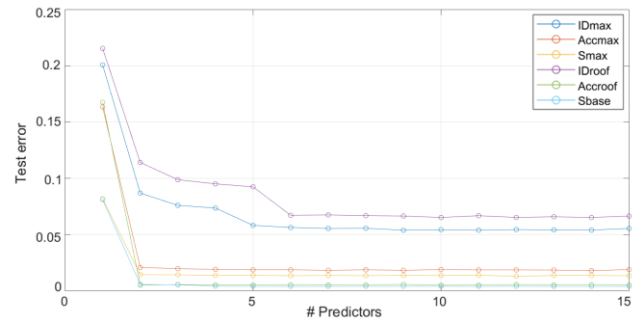

Figure 8: *Forward selection: Testing error vs number of predictors for each EDP.*

This scenario is also supported by Figure 8, which illustrates how the test error decreases if the number of predictors is increased. Likewise, it is observed that inter-storey drifts (ID_{max} and ID_{roof}) present a similar path as well as accelerations (Acc_{max} and Acc_{roof}) and storey shears (S_{max} and S_{base}), using forward stepwise selection. For example, considering two predictors, the test errors are 2.08% and 0.52% for Acc_{max} and Acc_{roof}, while the errors are 1.43% and 0.57% for S_{max} and S_{base}, respectively. Therefore, two parameters are used to build the model.

Table 1 shows the two selected predictors for each EDP, the R^2 values, and the standard deviations of the models, using forward stepwise selection. Although some intensity measures are repeated in Table 1 and Section 3.1, such as PGD in inter-storey drifts and PGA in both accelerations and storey shears, the aggregation effects in combination with other intensity measures lead to the incorporation of other predictors in the model. This is explained because when the second iteration starts, the algorithm evaluates the best fitting of the model in function of the results from the first iteration, with the first intensity measure (IM_1) already added. In fact, in terms of methods, this is one of the disadvantages of the forward selection procedure, which has to be contrasted with other methods and evaluated by experimented professionals in the field.

Table 1: *Variable selection, R^2 values and standard deviation using forward stepwise selection.*

EDP_i	IM_1	IM_2	R^2	σ
ID_{max}	PGV	PGD	0.8114	0.96
Acc_{max}	PGA	T_1/T_m	0.9831	1.11
S_{max}	PGA	$S_{a(T=1.0\ s)}$	0.9714	1.16
ID_{roof}	PGV	PGD	0.7202	0.99
Acc_{roof}	PGA	PGV	0.9989	1.16
S_{base}	PGA	T_1/T_m	0.9952	1.15

Whilst the inter-storey drift models present smaller R^2 values due to the dispersion of the inter-storey drift data, the accelerations and storey shears models present a better correlation. In addition, the calculated R^2 values in the first approach in Section 3.1 and Table 1 are dissimilar; however, they can be explained by the different data sub-sets employed where the initial approach performed the regression using 100% of the data instead of the 50% of training data that was assumed in the forward stepwise selection.

Turning now to the backward stepwise selection, the same algorithm is used but with one feature changed. This refers to the kind of stepwise selection applied. In the same way, 90 regression models are calculated and evaluated.

Figure 9 shows R^2 values for each predicted EDP, using different numbers of variables for IMs. The results are similar to those observed in the forward selection. The inter-storey drifts (ID_{max} and ID_{roof}) present a similar pattern to the accelerations (Acc_{max} and Acc_{roof}) and storey shears (S_{max} and S_{base}). Regarding the accelerations and storey shears, the aggregation effect is

not significant after two predictors. For instance, considering two predictors, the acceleration R^2 values are 0.9833 and 0.9988 for Acc_{max} and Acc_{roof}; and the storey shear R^2 values are 0.9715 and 0.9952 for S_{max} and S_{base}, respectively.

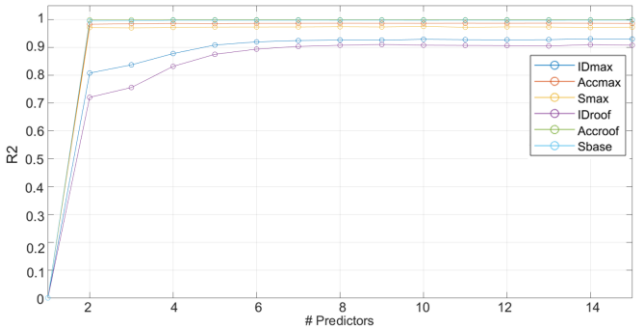

Figure 9: *Backward selection: R^2 value vs number of predictors for each EDP.*

In comparison with accelerations and storey shears, inter-storey drifts present R^2 values of 0.8080 and 0.7207 for ID_{max} and ID_{roof} with two predictors, which increase to over 0.9 when using five and seven predictors. Regarding the standard deviation and test error (Figure 10), the values are 0.95 and 0.99 for ID_{max} and ID_{roof}; and 8.71% and 11.47% for ID_{max} and ID_{roof}, respectively, using two predictors. While the standard deviation is kept almost constant after two predictors, the test error is reduced to 5.99% and 6.74% in both ID_{max} and ID_{roof}, using five and seven predictors. For the sake of simplicity and because the dispersion of the prediction becomes constant, two predictors are considered in the model.

Similarly to the forward selection, Figure 10 illustrates insignificant impact on the test error. The test errors are 2.10% and 0.51% for Acc_{max} and Acc_{roof}, while the errors are 1.42% and 0.56% for S_{max} and S_{base}, respectively, using two predictors. As before, two variables are utilised to build our model.

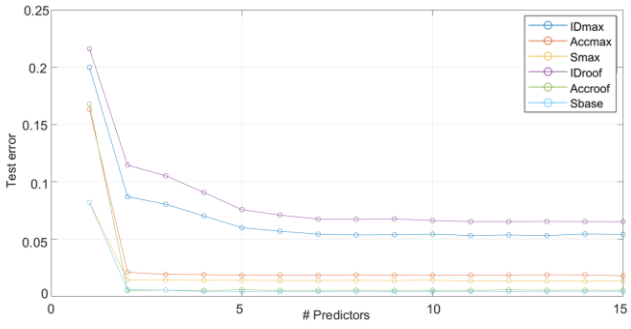

Figure 10: *Backward selection: Testing error vs number of predictors for each EDP.*

Table 2 shows the two chosen predictors for each EDP, the R^2 values, and the standard deviations of the models, using backward stepwise selection. In this case, there are several similarities between Section 3.1, Table 1 and Table 2. For instance, the first intensity measure (IM_1) in each model with backward selection is used in the forward selection model and is also supported by Figures 4, 5, and 6. In addition, it is interesting to note that the second intensity measure in the backward selection is related to the ratio between the fundamental period and the mean period (T_1/T_m) in four EDPs. In other words, although the R^2 values for this variable are small in the first approach in Section 3.1, the aggregation effect of this variable in the backward selection is better in comparison with the other parameters.

Table 2: *Variable selection, R^2 values and standard deviation using backward stepwise selection.*

EDP_i	IM_1	IM_2	R^2	σ
ID_{max}	PGD	T_1/T_m	0.8376	0.97
Acc_{max}	PGA	T_1/T_m	0.9854	1.12
S_{max}	PGA	T_m	0.9708	1.15
ID_{roof}	PGD	T_1/T_m	0.7606	1.02
Acc_{roof}	PGA	PGV	0.9990	1.15
S_{base}	PGA	T_1/T_m	0.9956	1.15

In comparison with the stepwise selection, the following functional form (equation 4) is adopted by the Lasso algorithm:

$$\ln EDP_k = \beta_0 + \sum_j \beta_j \cdot \ln(IM_j) \qquad (4)$$

where EDP_k is the k-th (k=1,2,…,6) engineering demand parameter, IM_j is the j-th (i=1,2,…,14) intensity measure, β_o is the intercept, and β_j is the j-th estimated coefficient related to the j-th intensity measure. In terms of importing and preparing the data, the same procedure is applied. Then, a total of 6 regression models are calculated and evaluated.

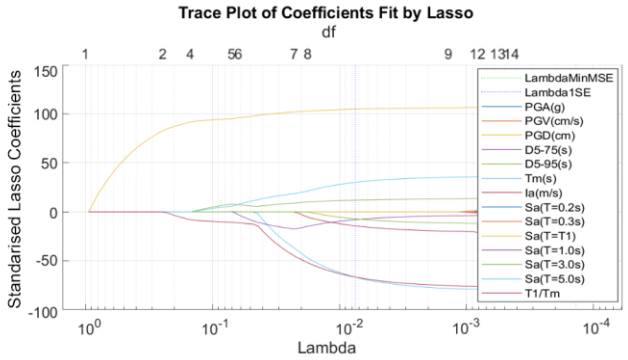

Figure 11: *Maximum Inter-storey Drift along the height, using Lasso algorithm.*

Figures 11 and 12 summarise the results from Lasso as applied to the inter-storey drifts along the building height (ID_{max}) and at the roof level (ID_{roof}), respectively. In the y-axis, the estimated coefficients are shown with respect to different values of the tuning parameter λ. The entrance order of each predictor, according to the λ value is also presented in Figures 11 and 12. The selection of this parameter (λ) is key because this reduces the mean squared error of the model. In the graph, the tuning parameter is drawn, considering the minimum value and the minimum plus one standard error. In the case of ID_{max}, PGD and T_m are selected as the first and second

intensity measures in the model, which are the same identified variables for ID_{roof}.

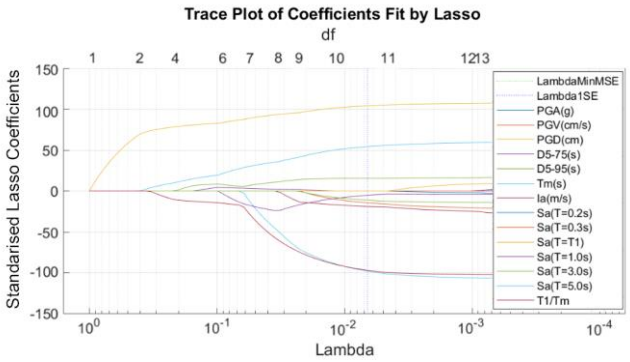

Figure 12: *Inter-storey Drift (Roof), using Lasso algorithm.*

On the other hand, Figure 13 shows the case of Acc_{max}, choosing PGA and I_a as the first and second intensity measures in the model, respectively. Similarly, Figure 14 displays the case of Acc_{roof}, where PGA and PGV are selected as the first and second intensity measures in the model.

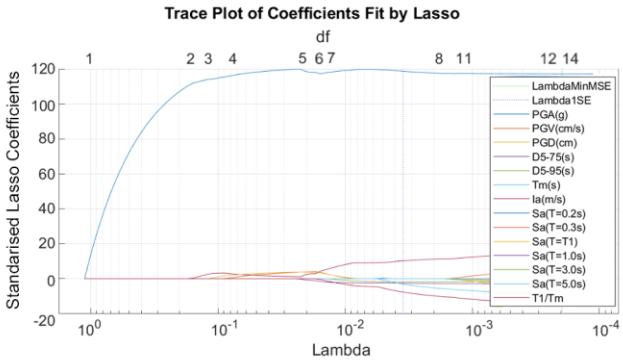

Figure 13: *Maximum Storey Acceleration along the height, using Lasso algorithm.*

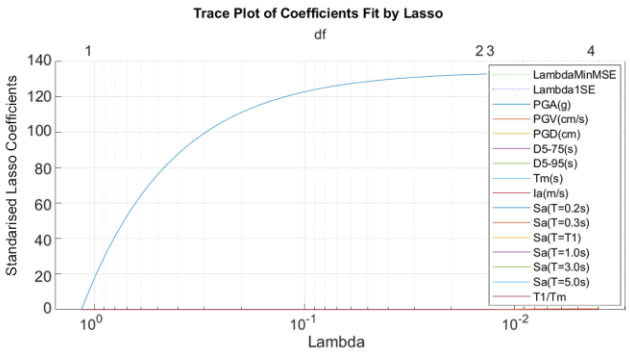

Figure 14: *Acceleration (Roof), using Lasso algorithm.*

According to Figures 15 and 16, in the case of S_{max}, PGA and I_a are selected as the first and second intensity measures in the model. Likewise, in the case of S_{base}, PGA and T_1/T_m are selected as the first and second intensity measures in the model.

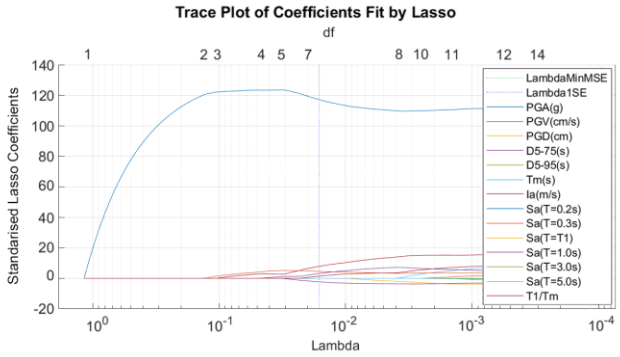

Figure 15: *Maximum Storey Shear along the height, using Lasso algorithm.*

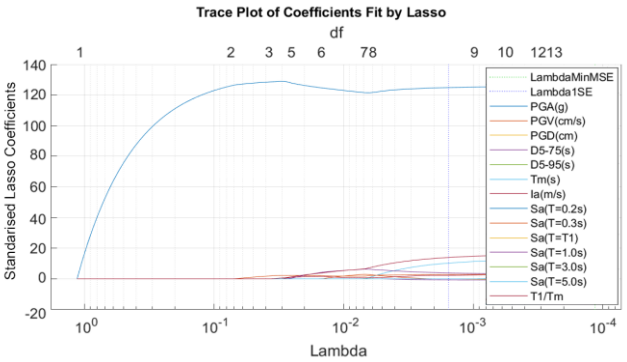

Figure 16: *Base Shear, using Lasso algorithm.*

Table 3 summarises the variables selection by each method, where several patterns are repeated for each EDP. As seen below, the inter-storey drifts (ID_{max} and ID_{roof}) have a common IM among the three methods, which is the peak ground displacement (PGD). Also, the mean period is used in two of the three methods, which could be selected as a good parameter to predict the seismic response of CLT buildings.

Table 3: *Summary of the variable selection.*

EDP	Forward		Backward		Lasso	
	IM_1	IM_2	IM_1	IM_2	IM_1	IM_2
ID_{max}	PGV	PGD	PGD	T_1/T_m	PGD	T_m
Acc_{max}	PGA	T_1/T_m	PGA	T_1/T_m	PGA	I_a
S_{max}	PGA	$S_{a(T=1.0\ s)}$	PGA	T_m	PGA	I_a
ID_{roof}	PGV	PGD	PGD	T_1/T_m	PGD	T_m
Acc_{roof}	PGA	PGV	PGA	PGV	PGA	PGV
S_{base}	PGA	T_1/T_m	PGA	T_1/T_m	PGA	T_1/T_m

In the case of acceleration, on one hand, the maximum acceleration considers the peak ground acceleration (PGA) as the first predictor. However, one variable is repeated, which is the ratio between the fundamental and mean period (T_1/T_m). Even though Lasso regression identifies the Arias intensity as the second parameter, in Figure 13 the algorithm is close to select T_1/T_m, confirming that this predictor might be a good second parameter. On the other hand, for the acceleration at the roof level, PGA is identified by the three methodologies together with PGV as the first and the second variables, respectively.

The storey shear is governed by the PGA as the first predictor in both maximum and base storey shear, which is consistent with current seismic codes. For the second parameter, whereas there is an agreement on the use of T_1/T_m in the base shear, the second variable is not clearly in agreement.

Table 4 summarises the training and test errors for individual EDPs and the techniques used. As seen in Lasso, both errors are slightly similar in each situation, e.g. the training and testing error of the maximum storey shear is 1.4% in both cases. In fact, since the testing error is the obtained error of the model tested with an unseen data, this is a good indicator that the model can predict the seismic response of our data quite well, for each EDP. In addition, with respect to the values of error, they are consistent with the other inferences from the first variable selection in Section 3.1 and stepwise selection.

Table 4: *Summary of the training (tr) and test (ts) error.*

EDP	Forward		Backward		Lasso	
	tr	ts	tr	ts	tr	ts
	Error	Error	Error	Error	Error	Error
ID_{max}	5.7%	5.8%	8.1%	8.1%	5.6%	5.5%
Acc_{max}	2.0%	2.1%	2.0%	1.9%	1.9%	1.9%
S_{max}	1.4%	1.4%	1.4%	1.4%	1.4%	1.4%
ID_{roof}	6.8%	6.7%	10.6%	10.6%	6.8%	6.7%
Acc_{roof}	0.6%	0.5%	0.5%	0.5%	0.5%	0.6%
S_{base}	0.5%	0.6%	0.5%	0.5%	0.4%	0.4%

Regarding the global errors obtained in the three techniques applied; the errors are comparable for accelerations and storey shears, not only between the training and test errors but also between the three methodologies employed. By contrast, the overall error in the inter-storey drift is higher, which can be explained by the larger dispersion of the data of inter-storey drifts. Also, it is observed that, in backward selection, the error is greater than when forward selection or the Lasso are used for the same EDP. Although there are differences between EDP errors, in all cases, the training error and test error are very similar, which means that the model proposed could properly predict the seismic response under new data.

4 CONCLUSIONS

This paper has applied machine learning techniques to assess the efficiency of different ground-motion intensity measures as seismic response predictors for CLT buildings. A direct relationship between storey shear and the PGA was observed, which is consistent with current codified guidance, where the design storey shear and acceleration are related to the PGA. In addition, the inter-storey drift was found to be related to the PGD, which usually applies only in the case of long period structures.

Importantly, a clear pattern was identified in the inter-storey drift (both the maximum value along the height and the maximum drift at the roof level) where a trend is observed among the three machine learning methodologies adopted (backward and forward selection and Lasso) where the PGD, period ratio (T_1/T_m), and mean period (T_m) are identified as the preferred intensity measures to model the inter-storey drift. Furthermore, a pattern was recognized in the acceleration in both the maximum floor acceleration and the peak acceleration at the roof level, where the PGA emerges as the first predictor. However, for the second parameter, there is a tendency to use the T_1/T_m and PGV in the maximum acceleration along the height and the maximum acceleration at the roof level, respectively. Similarly, a pattern was noticed in the storey shear values for both the maximum shear along the height and the maximum value at the base, where a trend is seen once the PGA and T_1/T_m are considered as the intensity measures to model the storey shear. These conclusions support those presented by Demirci, et al. [4, 5] on the basis of more conventional statistical approaches.

The procedure used in this study to select the input variables in the regression models can be replicated to predict not only the seismic response in CLT buildings but also in other kinds of structures. However, in terms of the number of variables, the models have demonstrated to be quite sensitive to the order of adding or removing variables, mainly in the forward and backward selection techniques. In this sense, it is important to consider their validation and cross-validation as steps to choose the optimal model. In the current research, and as a consequence of the variable selection performed, a predictive model can be defined from each of the three machine learning methodologies used, however, those models have not been tested and require further validation.

ACKNOWLEDGEMENT

This study was supported by the Advanced Human Capital Program of the Chilean National Commission for Scientific and Technological Research (CONICYT PFCHA/MAGISTER BECAS CHILE/2017-73180396) whose support is gratefully acknowledged.

REFERENCES

[1] Dujic, B., Strus, K., Zarnic, R., Ceccotti, A.: Prediction of dynamic response of a 7-storey massive Xlam wooden building tested on a shaking table. In: 11th World Conference on Timber Engineering (WCTE), 3450-3457, 2010.

[2] Pozza, L., Scotta, R., Trutalli, D., Polastri, A.: Behaviour factor for innovative massive timber shear walls. Bulletin of Earthquake Engineering, 13(11):3449-3469, 2015.

[3] Pozza, L., Trutalli, D.: An analytical formulation of q-factor for mid-rise CLT buildings based on parametric numerical analyses. Bulletin of Earthquake Engineering. 15(5):2015-2033, 2017.

[4] Demirci, C., Málaga-Chuquitaype, C., Macorini, L.: Seismic drift demands in multi-storey cross-laminated timber buildings. Earthquake Engineering & Structural Dynamics, 47(4):1014-1031, 2018.

[5] Demirci, C., Málaga-Chuquitaype, C., Macorini, L.: Seismic shear and acceleration demands in multi-storey cross-laminated timber buildings. Engineering Structures, 198:109467, 2019.
[6] Subcommittee B/525/5: BS EN 1995-1-1:2004+A2:2014 - Eurocode 5: Design of timber structures. General. Common rules and rules for buildings. BSI Standards Limited, 2014.
[7] C. Málaga-Chuquitaype and A. Y. Elghazouli. Chapter 8: Design of timber structures. In A. Y. Elghazouli, editor, *Seismic Design to Eurocode 8*, pages 213–233. CRC Press, Taylor & Francis Group, Boca Raton, FL, USA, 2016.
[8] Subcommittee B/525/1: BS EN 1991-1-1:2002 - Eurocode 1. Actions on structures. General actions. Densities, self-weight, imposed loads for buildings. BSI Standards Limited, 2010.
[9] Subcommittee, B/525/8: BS EN 1998-1:2004+A1:2013 - Eurocode 8: Design of structures for earthquake resistance. General rules, seismic actions and rules for buildings. BSI Standards Limited, 2013.
[10] Gavric, I., Popovski, M.: Design models for CLT shearwalls and assemblies based on connection properties. In: 47th Meeting - International Network on Timber Engineering Research (INTER), 269-280, 2014.
[11] Rotho Blass srl.: Wood connectors and timber plates. Rothoblass, 2015.
[12] McKenna F.: OpenSees: A framework for earthquake engineering simulation. Computing in Science and Engineering, 13(4):58-66, 2011.
[13] Gavric, I., Fragiacomo, M., Ceccotti, A.: Cyclic Behavior of CLT Wall Systems: Experimental Tests and Analytical Prediction Models. Journal of Structural Engineering, 141(11):4015034, 2015.
[14] Rathje, E., Dawson, C., Padgett, J.E., Pinelli, J.-P., Stanzione, D., Adair, A., Arduino, P., Brandenberg, S.J., Cockerill, T., Dey, C., Esteva, M., Haan, Jr.F.L., Hanlon, M., Kareem, A., Lowes, L., Mock, S., Mosqueda, G.: DesignSafe: A New Cyberinfrastructure for Natural Hazards Engineering. ASCE Natural Hazards Review. 1527-6996, 2017.
[15] James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning: with Applications in R. Springer New York Heidelberg Dordrecht London, 2013.