



Politecnico
di Torino

EPFL

POLITECNICO DI TORINO

Department of Structural, Building and Geotechnical Engineering

Master of Science in Civil Engineering

Structures and infrastructures

Master thesis

Exposure modelling and seismic vulnerability assessment in Switzerland

Student

Annalisa Casciato

Supervisors

Rosario Ceravolo

Alireza Khodaverdian

Pierino Lestuzzi

1859

December 2021

ABSTRACT

Natural disasters have been always caused a danger to human life, and among these are earthquakes. Seismic risk assessment consists of the evaluation of existing buildings and their expected response in case of earthquake; exposure model of buildings has a significant role in the final results of risk calculations. With this respect, several studies, including traditional data acquisition (e.g. visual survey) or advanced methods (e.g. remote sensing and machine learning) are conducted. In recent years, advanced techniques have been developed to speed up and automatize the processes of data acquisition to data interpretation, although it is worth mentioning that the visual survey is essential to train and validate machine learning methods.

In the present study, we combined the traditional visual survey with the implementation of a deep learning model to identify building types. First, in order to understand the taxonomy of buildings in Switzerland, several cities (e.g. Neuchatel, Yverdon-Les-Bains) are studied with a virtual/physical survey. As a first outcome of the survey, city mapping schemes are obtained by classifying buildings according to the main features (i.e., construction period and height classes). Next, Random Forest (RF), a supervised learning algorithm, is applied to classify buildings into building types by exploiting all the building attributes. The RF model, trained and tested on the cities of Neuchatel and Yverdon-Les-Bains and then applied to two other Swiss cities (e.g. Solothurn and Visp), which are also visually/physically (e.g. Google street) surveyed. The decent accuracy of the results by application of the model to two cities of Solothurn and Visp with different distributions of building types showed that the robustness of the method in prediction of building types in other cities in Switzerland, paving the path for its application to whole country. Finally, to study the performance of the proposed building type detection in seismic risk assessment, the seismic damage for two different scenarios is evaluated by considering the real and predicted building exposure models. A negligible discrepancy between the estimated damages based on the real and predicted exposure models demonstrate the successfulness of the method in risk assessment with high accuracy.

TABLE OF CONTENTS

Abstract.....	III
Table of contents.....	V
List of figures.....	VII
List of tables.....	XI
1 Introduction.....	1
1.1 Natural disasters: severity and consequences.....	1
1.2 Risk assessment: different techniques.....	1
1.3 Exposure models: techniques used in the past and our method.....	2
2 Visual survey.....	5
2.1 Visual survey.....	5
2.2 Case study: Neuchatel and Yverdon-Les-Bains.....	6
2.3 Building attributes.....	8
2.4 Results of visual survey.....	10
2.4.1 Distribution of building types.....	12
2.4.2 Distribution of period of construction.....	14
2.4.3 Mapping schemes.....	16
2.5 Investigation on the other building features.....	17
2.5.1 Numerical features.....	18
3 The Random Forest Classifier Method (RF).....	23
3.1 Description of the method.....	23
3.2 Application of the RF method on Neuchatel and Yverdon-Les-Bains.....	25
3.2.1 Results.....	25
3.3 Accuracy of RF.....	28
3.4 Parameters affecting accuracy of RF.....	30
4 Application of the RF model to other cities.....	33
4.1 Survey.....	34
4.2 The RF model based on Neuchatel and Yverdon-Les-Bains datasets.....	35
4.3 Application of RF model to Solothurn and Visp.....	36

5	Damage models.....	39
5.1	Hazard models.....	39
5.2	Exposure models	40
5.3	Fragility curves.....	40
5.4	Damage assessment of Neuchatel and Yverdon	41
5.5	Damage assessment of Solothurn: real and predicted exposure models.....	44
5.6	Damage assessment of Visp: real and predicted exposure models	47
5.7	Damage assessment in Visp: comparison between hazard models of magnitude 5, 6 and 7	52
5.8	Damage assessment in Visp: probability of damage grade for different classes of damage.....	55
6	Conclusion	59
7	Future works	63
8	Bibliography	65
9	Acknowledgments.....	69
10	Appendix.....	71

LIST OF FIGURES

Figure 1 - Building types: (a) M3; (b) M4; (c) M6; (d) RCW.....	6
Figure 2 - Neuchatel and Yverdon-Les-Bains	7
Figure 3 – Flowchart of the visual survey	8
Figure 4 –Building type distribution in Neuchatel	11
Figure 5 - Building type distribution in Yverdon-Les-Bains.....	11
Figure 6 - Distribution of building types: (a) Neuchatel; (b) Yverdon-Les-Bains.	12
Figure 7 – Distribution of building types per height class in Neuchatel: (a) Low-rise; (b) Mid-rise; (c) High-rise.	13
Figure 8 - Distribution of building types per height class in Yverdon-Les-Bains: (a) Low-rise; (b) Mid-rise; (c) High-rise.....	14
Figure 9 - Distribution of building per period of construction: (a) Neuchatel; (b) Yverdon-Les-Bains.	15
Figure 10 - Heat-map of the pair-wise correlations between the building features: (a) Neuchatel; (b) Yverdon-Les-Bains.	19
Figure 11 - Distribution of number of housing units for (a) Low-rise Neuchatel; (b) Mid-rise Neuchatel; (c) Low-rise Yverdon-Les-Bains; (d) Mid-rise Yverdon-Les-Bains.....	21
Figure 12 - RF diagram (Viet-Hung Dang, 2020)	24
Figure 13 - Real and predicted distributions of building types in (a-b) Neuchatel and (c-d) Yverdon-Les-Bains.	26
Figure 14 - Real and predicted distributions of building types with different height classes in Neuchatel	26
Figure 15 - Real and predicted distributions of building types with different height classes in Yverdon-Les-Bains.....	27
Figure 16 - Confusion matrices for (a) Neuchatel; (b) Yverdon-Les-Bains.....	28
Figure 17 - Accuracy variation according to the training set size	30
Figure 18 - Accuracy variation by changing number of trees in RF model	31
Figure 19 - Neuchatel, Yverdon-Les-Bains, Solothurn and Visp. The Swiss border is highlighted in black lines.	33
Figure 20 - Distribution of building types: (a) Solothurn; (b) Visp.....	34

Figure 21 - Distribution of building types in Neuchatel, Yverdon-Les-Bains, Visp and Solothurn.....	35
Figure 22 - Confusion matrix of RF model, trained and tested on the concatenated dataset of Neuchatel and Yverdon-Les-Bains.....	36
Figure 23 - Application on Solothurn of the RF model trained on Neuchatel and Yverdon-Les-Bains: (a) Solothurn – real distribution; (b) Visp – real distribution; (c) Solothurn – predicted distribution; (d) Visp – predicted distribution.....	37
Figure 24 –Confusion matrix of the RF model, trained on the Neuchatel and Yverdon-Les-Bains dataset and applied on (a) Solothurn; (b) Visp.	38
Figure 25 - Sierre earthquake (1946).....	40
Figure 26 - Distribution of damage grade: (a) Neuchatel; (b) Yverdon-Les-Bains.....	42
Figure 27 - Distribution of damage grade: comparison between Neuchatel and Yverdon-Les-Bains	44
Figure 28 - Distribution of damage grade of Solothurn based on the real and predicted exposure models.....	45
Figure 29 - Distribution of damage grade for different building types in Solothurn: (a) reality; (b) prediction.	46
Figure 30 - Differences of the expected damage evaluated from real and predicted exposure model in Solothurn	46
Figure 31 - Distribution of damage grade: comparison between Solothurn reality and Solothurn prediction.....	47
Figure 32 - Distribution of damage grade of Visp based on the real and predicted exposure models.....	48
Figure 33 - Distribution of damage grade in Visp resulted from: (a) real exposure model; (b) predicted exposure model.....	49
Figure 34 - Differences of damage grade evaluated from real and predicted exposure model for Visp	49
Figure 35 - Distribution of damage grade: comparison between Visp reality and Visp prediction	50
Figure 36 – Difference in probability between reality and prediction – DG0.....	51
Figure 37 – Difference in probability between reality and prediction – DG1	51
Figure 38 – Distribution of grade of damage: (a) Magnitude 5; (b) Magnitude 6; (c) Magnitude 7.	53

Figure 39 – Distribution of damage grade: (a) Magnitude 5 – reality; (b) Magnitude 5 – prediction; (c) Magnitude 6 – reality; (d) Magnitude 6 – prediction; (e) Magnitude 7 – reality; (f) Magnitude 7 – prediction.....	54
Figure 40 – Difference of damage grades in Visp: (a) Magnitude 5; (b) Magnitude 6; (c) Magnitude 7.....	55
Figure 41 - Visp - Distribution of building types	56
Figure 42 - Visp - DG0	57
Figure 43 - Visp - DG1	57
Figure 44 - Flowchart of visual survey based on experience.....	71
Figure 45 - Distribution of damage grade according to taxonomy of buildings: (a) Neuchatel; (b) Yverdon-Les-Bains.....	72
Figure 46 – Difference in probability between reality and prediction – DG2.....	74
Figure 47 - Visp - DG2	74

LIST OF TABLES

Table 1 - Building types.....	5
Table 2 - Building features reported in the dataset of Federal Office of Buildings and Logistic	9
Table 3 - Classification of the construction periods.....	10
Table 4 - Classification of the number of stories.....	10
Table 5 - Mapping scheme of Neuchatel based on the two features of the construction period and height class.....	16
Table 6 - Mapping scheme of Yverdon-Les-Bains, based on the two features of construction period and height class.....	17
Table 7 – Building features used for the building type classification.....	18
Table 8 – Real and predicted distributions of building types and its difference (Neuchatel)	29
Table 9 - Real and predictd distributions of building types and its difference (Yverdon-Les-Bains).....	29
Table 10 - Accuracy measurements for the Neuchatel and Yverdon-Les-Bains RF model	30
Table 11 - Accuracy measurements of the RF model, trained and tested on the Neuchatel and Yverdon-Les-Bains dataset	36
Table 12 - Accuracy measurements of the RF model, trained on the Neuchatel and Yverdon-Les-Bains dataset and applied on Solothurn and Visp	38
Table 13 - Building types used for damage assessment in Neuchatel and Yverdon-Les-Bains	42
Table 14 - Distribution of damage grade in Neuchatel.....	43
Table 15 – Distribution of damage grade in Yverdon-Les-Bains.....	43
Table 16 – Building types in Solothurn: reality and prediction.....	44
Table 17 - Building types in Visp: reality and prediction.....	47
Table 18 - Solothurn - reality: distribution of damage grade.....	72
Table 19 - Solothurn - prediction: distribution of damage grade.....	72
Table 20 - Visp - reality: distribution of damage grade.....	73
Table 21 - Visp - prediction: distribution of damage grade.....	73

1 INTRODUCTION

1.1 NATURAL DISASTERS: SEVERITY AND CONSEQUENCES

Natural disasters always represented a menace for human life for both the economical and physical aspects; among them, earthquakes represent a serious hazard worldwide (Wei et al., 2013). Indeed, approximately 1,600,000 deaths have been caused by earthquakes worldwide (Wu et al., 2014). The considerable damage observed after the latest moderate-to-strong earthquakes has increased awareness regarding natural disasters in the last decades (Jackson & Conway, 2006). A recent example in Europe is the severe L'Aquila earthquake with magnitude 5.9 M_L occurred in Italy on April 6th, 2009 and caused more than 300 victims, 1,600 injured people and financial losses of ~ 10 billion Euros (Greco et al., 2018). The most recent large-magnitude earthquake in Europe is the M_w 5.3 Zagreb earthquake, occurred on March 22nd, 2020. The epicentre was 7 kilometres (4.3 mi) north of the city centre of Zagreb (Croatia). The maximum intensity felt was VII-VIII (very strong to damaging) on the Medvedev-Sponheuer-Karnik scale. It was the strongest earthquake in Zagreb since the 1880 earthquake and caused substantial damage to the city's historic centre. The direct damage inflicted by the earthquake on Zagreb and Krapina-Zagorje County was estimated 86 billion Croatian kuna (11.5 billion euros) (Josip Atalić, Uroš et al., 2021). Moreover, the impact of natural disasters has raised in the last decades. Global urbanization processes and increasing spatial concentration of exposed elements (e.g., people, buildings, infrastructure, and financial assets) in earthquake-prone area led to an increased seismic risk (Geiß et al., 2016).

1.2 RISK ASSESSMENT: DIFFERENT TECHNIQUES

In general, risk of earthquake is defined as the probability of earthquake consequences in terms of environmental, economical and social aspects and quantified by using loss modelling procedures, which includes earthquake hazard and fragility modelling (Erdik, 2017). Earthquake hazard are focused in estimation of strong ground motion parameters by means of probabilistic or deterministic models. For each strong ground motion, the structural damages are determined by using fragility models, linking the probability of damage to the level of intensity. Consequence model quantifies the loss (costs for retrofitting) or loss ratio (for example the repair cost divided by replacement cost) as a function of structural damages. Thus, some uncertainties should be considered during the earthquake risk estimations (Erdik, 2017).

Every year, hundreds of risk assessment studies are carried out. Some of them follow more traditional methods, others are more innovative (e.g. use of remote sensing, machine

learning methods). Concerning the traditional methods, for instance, in (Zuccaro & Cacace, 2015), a quick method for seismic vulnerability assessment is proposed, according to the EMS 98 classification. The criteria to assign the typological classes are mainly based on the vertical structure type. The purpose of this methodology is to reduce the uncertainty in the class assessment using a set of parameters, related to typological features, that are identified as modifiers of the vulnerability level. A synthetic damage parameter is defined for comparing the seismic response of different sets of buildings under the same seismic intensity and, in the end, the vulnerability assessment is obtained. In (Brando et al., 2017), a predictive model, Damage Probability Matrices (DPMs), for assessing the seismic vulnerability of small historic centres in Abruzzi region, in Italy, is presented. This model needs a certain number of parameters and it is based on information collected in the consequences of the 2009 L'Aquila earthquake. Concerning the more innovative methods, we can mention data mining methods (e.g. ARL and SVM), the use of machine learning and remote sensing techniques, which are aimed at creating an exposure model to perform the seismic damage assessment.

1.3 EXPOSURE MODELS: TECHNIQUES USED IN THE PAST AND OUR METHOD

Assessing the seismic vulnerability is a complex and expensive process, especially when entire urban area is concerned and a huge number of buildings are involved, as it includes the evaluation of existing buildings and their expected response in case of earthquake. With that goal, a large amount of data is required, while they are almost always incomplete. Over the last years, several methodologies have been implemented to overcome the incompleteness of data. For instance, Riedel et al. (2014) proposed to assess vulnerability exploiting the data already available from the country or the region. For implementing this concept, data mining methods were proposed. Indeed, Data mining are computational processes used to discover patterns within large datasets through a combination of machine learning, statistics, and dataset systems (Linoff & Berry, 1997). One of these methods is the Association rule learning (ARL), a rule-based data mining method used to reveal interesting relationships among variables within a large dataset (Frawley et al., 1992) using some measures of interestingness (Agrawal et al., 1993). In the last years, the ARL method is used to establish relationships between different building features (e.g., shape of roof, period of construction) and building types in different studies (Riedel et al., 2014), (Liu et al., 2019), (Diana et al., 2019).

Another method is the support vector machine (SVM) (C. Cortes, 1995), that is a classification algorithm (Noble, 2006). In essence, it is a mathematical entity that maximizes particular mathematical function with respect to a given high-dimension collection of data. It was employed by Torres et al. (2019) for vulnerability estimation in Lorca, Spain and by Riedel et al. (2015) for the seismic vulnerability assessment of urban environments. This method has also been applied by Harirchian et al. (2020), for investigating the efficacy of the Machine Learning (ML) application in damage prediction. Indeed, in that study, a fast, reliable and rapid method, known as Rapid Visual Screening (RVS) has been applied as a preliminary screening platform to identify vulnerable buildings by assigning the damage index for different types of structures. Moreover, SVM has been used in (Han & Kim, 2019) for seismic

vulnerability assessment and for mapping the M_L5.8 Gyeongju Earthquake in Gyeongju, South Korea, as a case study. They applied logistic regression (LR) and four kernel models (linear, polynomial, radial basis function, and sigmoid kernels), based on the support vector machine (SVM) learning method to derive suitable models for assessing seismic vulnerabilities.

Another field that has been exploited for the vulnerability assessment is the remote sensing, that is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance (typically from satellite or aircraft) (usgs.gov). For instance, in (Geiß et al., 2016), the use of multi-sensor remote sensing has been proposed. It has proven a great potential to extract relevant features for prevent vulnerability analysis of buildings. The intrinsic advantage of remote sensing is the ability to offer an overview of building stocks and serve as a screening method for derivation of building vulnerability related features, such as shape characteristics, height, roof material, period of construction, structure type, and spatial context (Geiß & Taubenböck, 2013).

In the field of new methods for assessing seismic vulnerability, Lee et al. (2019) implemented a decision-maker for identifying areas with potential vulnerability to reduce seismic damage. In particular, a GIS-based opensource software entitled Seismic-Related Vulnerability Calculation Software (SEVUCAS), based on the Step-wise Weight Assessment Ratio Analysis (SWARA) method and geographic information system, has been developed to assess seismic vulnerability by considering four groups of criteria (i.e., geotechnical, structural, socio-economic, and physical distance to needed facilities and away from dangerous facilities). After weighing the criteria (indicators) and alternatives (sub-indicators), the weighted overlay analysis was used to determine the final vulnerability map in the form of contours and statistical data.

Furthermore, Torres et al. (2019) proposed to characterize the built stock of the Spanish city of Lorca by integrating airborne LiDAR points, orthophotos and satellite images to create an exposure and earthquake vulnerability dataset. The procedure for data integration intended to be fast and replicable in the rest of the cities listed by the GDCPE of Spain (Torres et al., 2019). Despite to the possibility of employing all these new methodologies, it is important to remember that a ground truth dataset (e.g., dataset from visual survey) is necessary for training and validating machine learning models.

In the present paper, different cities have been surveyed to understand building taxonomy and building types in Switzerland. As a first step, 3537 buildings of Neuchatel and 2808 buildings of Yverdon-Les-Bains are analysed through a visual survey. The building taxonomy proposed by Lagomarsino et al. (2006) are considered and building types are detected by taking into the consideration of building characteristics (e.g., the roof shape, the façade aspect, the presence of balconies). After that, the obtained dataset are enriched by the building database from BAFU, with more than 18 building features (e.g., building height, period of construction, number of building units). As a first outcome of survey, the mapping schemes of the cities are obtained based on periods of construction and height classes. In the second part of the paper, the obtained results of surveys are used to develop a deep learning model for building type prediction based on building features. The Random Forest (RF) method, a supervised learning algorithm, is applied. This algorithm is an ensemble of random decision Tree classifiers, that makes predictions by combining the predictions of the individual

trees (Vens, 2013). The model is trained and tested on the cities of Neuchatel and Yverdon-Les-Bains, separately. The performance of the models are evaluated with three accuracy measures: AM1, based on the confusion matrix (also known as error matrix), AM2, obtained as the weighted average of the precision for each building type, and AM3, accuracy based on the distribution of building types. In the third section of the paper, a new RF model is trained and tested on the concatenated datasets of Neuchatel and Yverdon-Les-Bains. Next, in order to see the performance of the method for other cities, the model is applied on two different cities, Solothurn (3238 buildings) and Visp (307 buildings), visually/physically (e.g. Google street) surveyed. Accuracy of application of the new model to the different cities is comparable to the two models, which are separately trained and tested on the building datasets of Neuchatel and Yverdon-Les-Bains, showing that the methodology is also successful in predicting building types for cities with different distributions. In the last part of this paper, seismic damage assessment is carried out on the obtained exposure models. First, the damage assessment of Neuchatel and Yverdon-Les-Bains is performed. Next, the damage assessment of Solothurn and Visp for the two most-destructive historical events in Switzerland is conducted, considering both the real and predicted building exposure models, to study the performance of the proposed building type detection method in final output of seismic risk assessment. A negligible discrepancy between the estimated damages based on the real and predicted exposure models demonstrate the successfulness of the method in risk assessment with high accuracy.

2 VISUAL SURVEY

In order to assess the seismic vulnerability of existing buildings at urban scale, it is necessary to detect building types and categorize the buildings in some limited building types to minimize the complexity of the computation. Performing a visual survey at urban scale is facing considerable challenges; it requires a big amount of time and money. Although the most innovative methods (i.e., use of remote sensing and deep learning) can simplify the process of building type detection classification; it is, however, important to mention that a visual survey remains necessary for testing the methods and verifying the results. In this chapter, the visual survey and the analysis of the results are presented.

2.1 VISUAL SURVEY

In exposure modeling, we need to define building taxonomy (Porter et al., 2001), a list of building types, classified based on method and/or materials used for construction. Here, the building types proposed by Lagomarsino et al. (2006) are used. Building types and their definitions are presented in Table 1 and images of the building types are shown in Figure 1.

Table 1 - Building types

BUILDING TYPE	DESCRIPTION
M3	Masonry buildings with simple stone
M4	Masonry buildings with massive stone
M5	Unreinforced masonry (bricks) with flexible floors
M6	Unreinforced masonry—RC floors (rigid floors)
RCW	Reinforced concrete buildings with shear walls
RCF	Reinforced concrete buildings
W	Wood structures
S	Steel structures



Figure 1 - Building types: (a) M3; (b) M4; (c) M6; (d) RCW.

In order to better understand the visual survey, it is appropriate to give a brief description of each category. (1) M3 are masonry buildings built with simple stones. The roof shape is traditionally sloped and small balconies are observable if there is balcony. (2) M4 are masonry buildings with massive stone. They are identifiable by massive dressed stone and are usually used for important administrative purposes. (3) M5 are unreinforced masonry (bricks) with flexible floors. (4) M6 are unreinforced masonry buildings with RC floors (rigid floors). It is usually possible to identify some concrete elements on the façade of these structures. Moreover, they are characterized by deep balconies made of concrete. All M3, M5, M6 are mostly used for residential purposes. (5) RCW buildings are made of reinforced concrete walls, which are more common in comparison to reinforced concrete frame (RCF) buildings in Switzerland. RC buildings are usually recently built and characterized by a flat roof and big openings.

2.2 CASE STUDY: NEUCHATEL AND YVERDON-LES-BAINS

The case study presented in this chapter are the cities of Neuchatel and Yverdon-Les-Bains. Neuchatel is located in the northwestern part of Switzerland and overlooking on the homonymous lake. Yverdon-Les-Bains is a swiss city, located in the district of Jura-Nord vaudois of the canton of Vaud in Switzerland and it also overlooks Lake Neuchatel. In Figure 2, a map of Switzerland that includes Neuchatel and Yverdon-Les-Bains is presented.

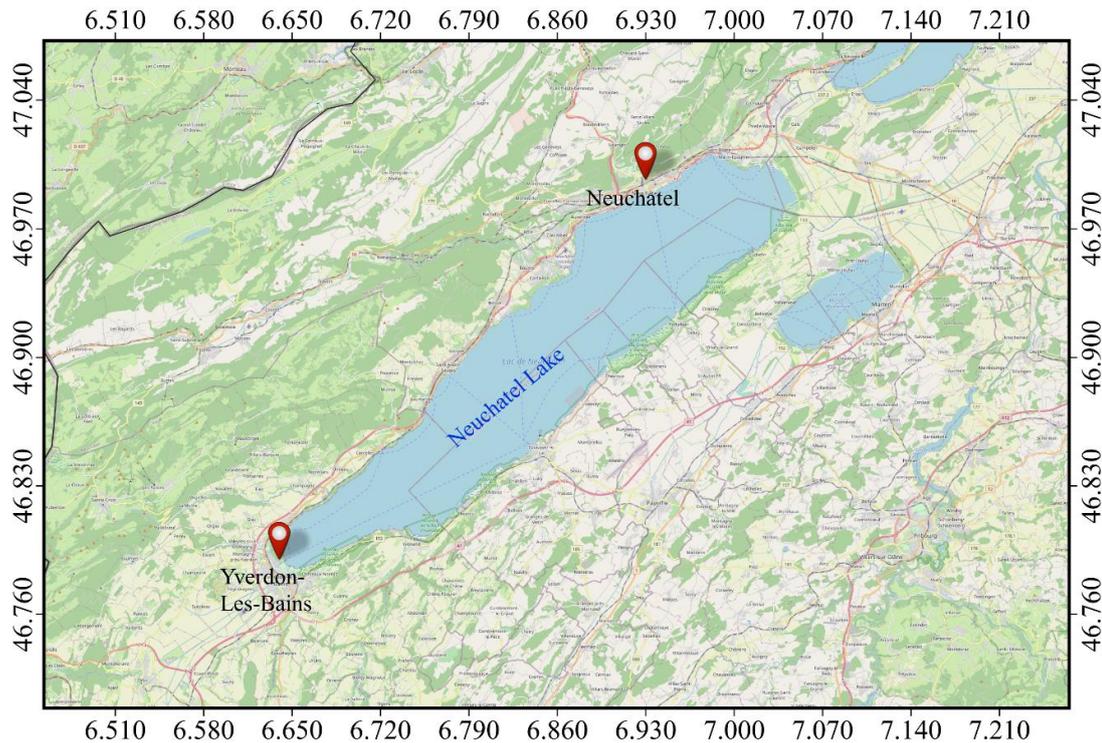


Figure 2 - Neuchâtel and Yverdon-Les-Bains

In total, 3537 buildings in the city of Neuchâtel and 2808 buildings in the city of Yverdon-Les-Bains are surveyed and type of any single building has been determined based on taxonomy presented in Table 1. In Figure 3, a flowchart that can be used as a hint during the visual survey is presented. Some main features are considered in order to assign the most appropriate building type to each building. The most relevant features that have been taken into account are the roof shape, the façade aspect, the presence of balconies. The first distinction is made with the roof shape, that can be sloped or flat. Generally, buildings with a sloped roof are the masonry ones, whereas buildings with a flat roof are concrete structures. Considering buildings with the flat roof, a distinction is made by exploring the structure elements; if the structure presents shear walls, it is probably a RCW structures, otherwise, if it is made with frames, it is a RCF structure. On the other hand, concerning the sloped roof structures, the façade aspect is considered. The buildings with stone masonry are identified as M3 or M4, according to the presence of massive stone on the façade. The structures made of bricks and concrete blocks masonry are identified as M5 and M6. In this field, the presence of balconies and their shape can be helpful for distinguish these two building types. Infact, M5 buildings tends to have no balconies or small ones whereas M6 structures use to have significant balconies. It is important to mention that the flowchart of Figure 3 can be a good guideline for performing a visual survey; however, it has been built based on the experience and it has its own limitation.

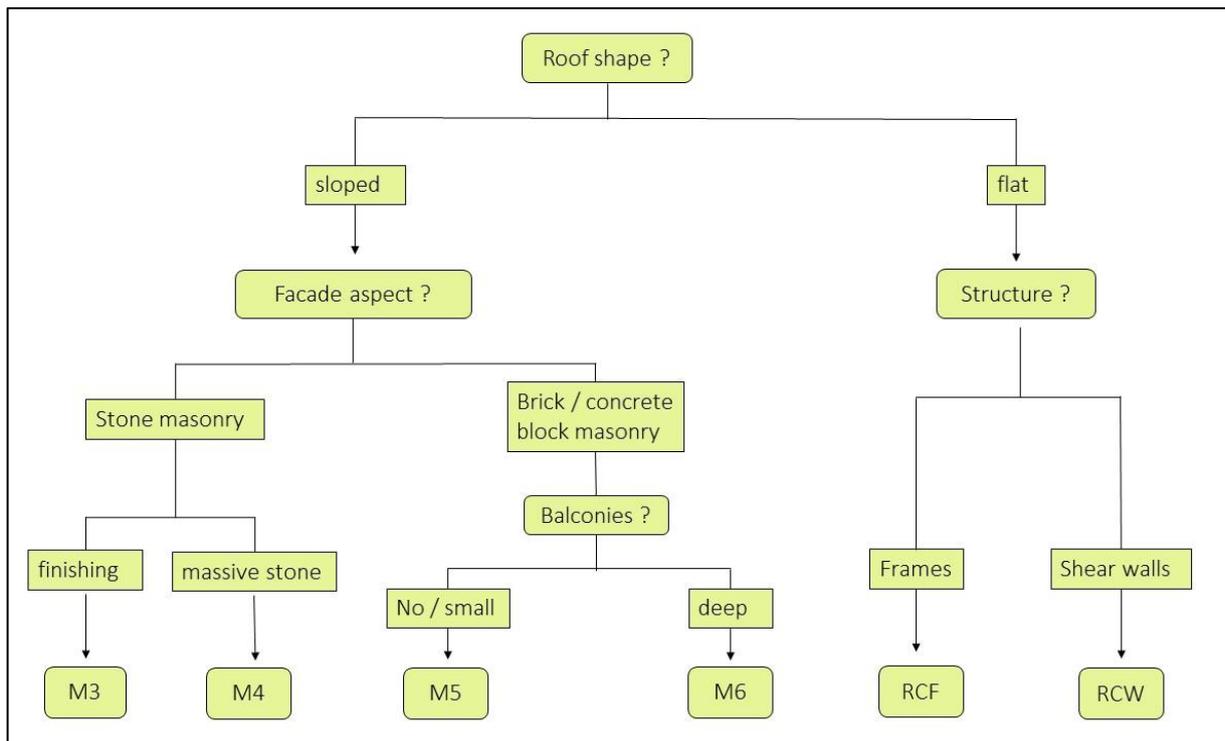


Figure 3 – Flowchart of the visual survey

2.3 BUILDING ATTRIBUTES

Fortunately, there is a dataset of buildings located in Switzerland from Federal Offices. It contains several data about building features: EGID - Federal building identification number, building location (canton's name, ZIP-code, coordinates of building location), structural characteristics (period of construction, footprint, number of stories, roof type), characteristics of housing units (number of housing units, cumulative number of rooms in housing units), usage (number of inhabitants or equivalent full time employees) and financial value (Replacement value in CHF, Value of mobile goods in CHF). The list of building features is presented in Table 2.

Table 2 - Building features reported in the dataset of Federal Office of Buildings and Logistic

ABBREVIATION	DESCRIPTION
EGID	Federal building identification number
ANZEDID	Number of federal entry identification numbers
GDEKT	Abbreviation of cantons's name
GDENRG	BFS commune number from GIS analysis
PLZ4G	ZIP-code from GIS analysis
PLZZG	ZIP-code additional number from GIS analysis
AREBAUZ	Building zone category (ARE)
GDETYP	Commune category (Infoflan ARE)
GKODE	East coordinate
GKODN	North coordinate
GKAT	Category of the building
GKLAS	Classification of the building
GBAUJ	Year of construction
GBAUP	Period of construction
FOOTPRI	Area of the building footprint in m2
GEBHOHE	Average height above ground in m
GEBVOL	Building volume in m3
ROOFTYPE	Roof inclination
ANZWHG02	Number of housing units
KUMWAREA	Cumulative area of housing units in m2
KUMWAZIM	Cumulative number of rooms in housing units
EINWMOD	Number of inhabitants
ANGMOD	Equivalent full time employees
PUBLMOD	Customers capacity for selected
GEBWERT	Replacement value in CHF
INHWERT	Value of mobile goods in CHF
UMSATZ	Production account as a proxy for the yearly sales revenue in CHF/year
EIGKAT	Owner's category
WITAKAT	Type of economic activity
RELPOS	Relative position of a building in an aggregate
GEOMETRY	Latitude and longitude of the building
CONSTRPRD	Construction period
NOSTORIES	Number of stories (only ground floor and upper floors)

In order to simplify the visual survey, a second dataset that contains street and building number of each building is used. The first and the second datasets are merged with respect on the EGID (Federal building identification number) and a new dataset is obtained. It contains the list of buildings of the city with the related features that are listed in Table 2, and the addresses of the buildings (street and building number). This dataset is used for performing the visual survey and for the analysis of the results.

For sake of simplicity in the interpretation of results, some features of the dataset have been categorized. The period of construction has been categorized in six categories, corresponding to ranges of period of construction (Table 3). The number of stories has been categorized in three categories (i.e., low-rise, mid-rise, high-rise) (please see Table 4).

Table 3 - Classification of the construction periods

CONSTRUCTION PERIOD	CODE
< 1919	1
1919 - 1945	2
1946 - 1970	3
1971 - 1990	4
1991 - 2003	5
> 2003	6

Table 4 - Classification of the number of stories

NUMBER OF STORIES	HEIGHT CLASS
1-3	L
4-6	M
>7	H

2.4 RESULTS OF VISUAL SURVEY

The maps of building types distribution in Neuchatel and in Yverdon-Les-Bains are presented in Figure 4 and in Figure 5, respectively. It is seen that the masonry buildings with stone (M3 and M4) are highly concentrated in the city centre and in the most ancient areas of the cities. On the other hand, the unreinforced masonry buildings (M5 and M6) and the reinforced concrete buildings (RCW and RCF) are mainly located in the peripheral zones of the cities.

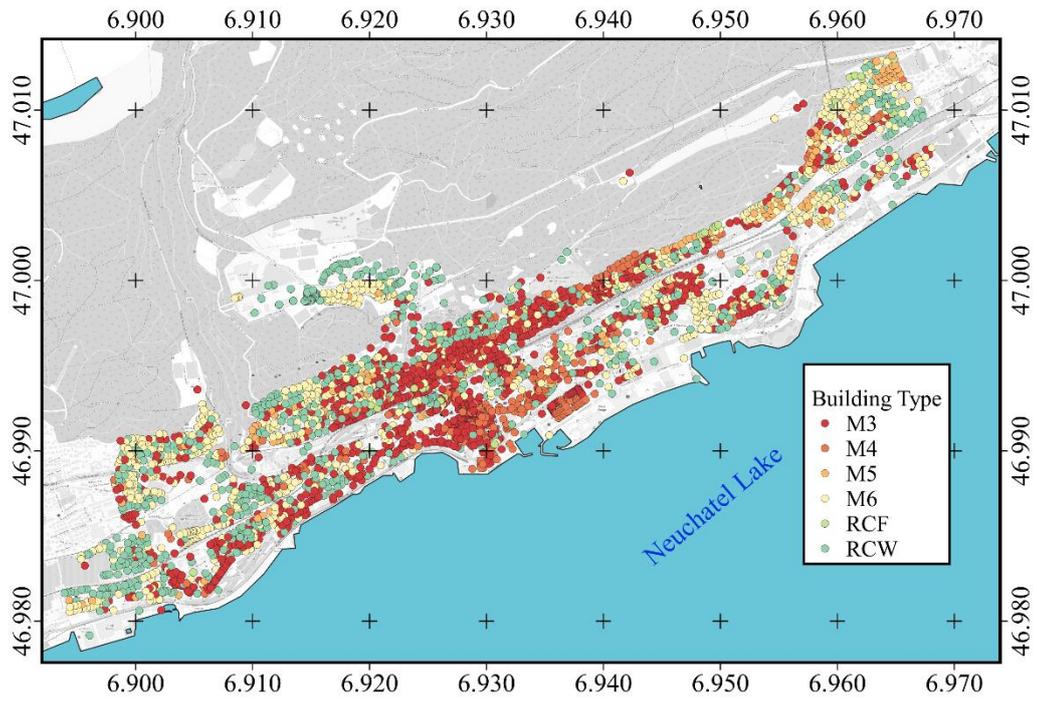


Figure 4 –Building type distribution in Neuchatel

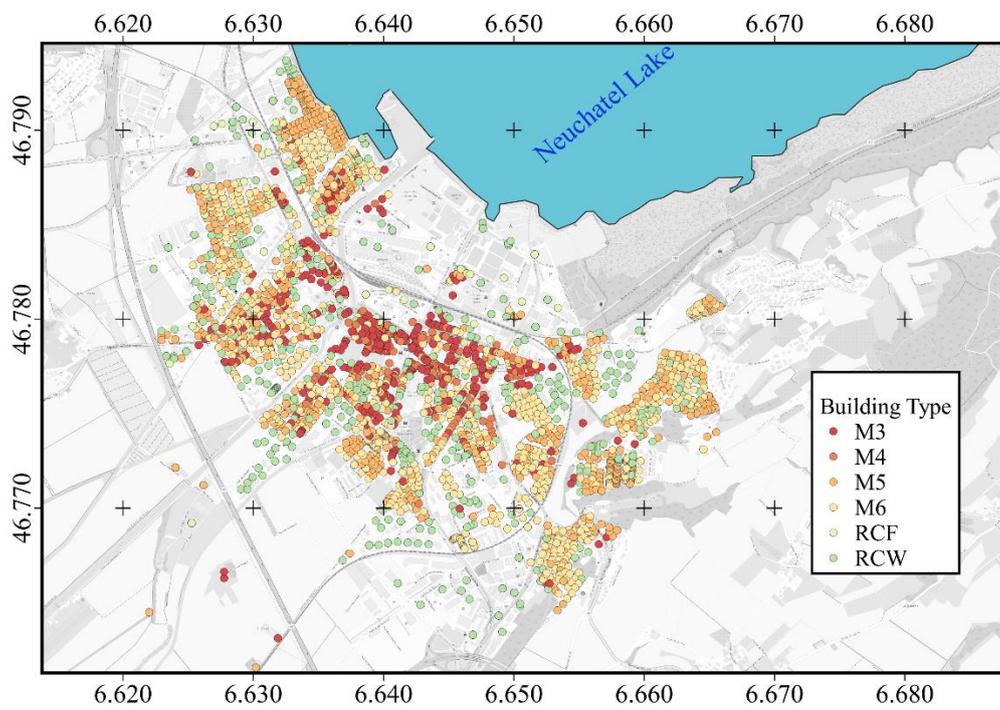


Figure 5 - Building type distribution in Yverdon-Les-Bains

2.4.1 Distribution of building types

Figure 6 shows the distribution of building types in percentages. Concerning Neuchatel, the majority of buildings are M3 (34.67%). The second and third majorities are M6 and RCW buildings, with 26.38% and 24.46% of the total number of buildings, respectively. However, the majority of buildings in Yverdon-Les-Bains are M6, with 30.63% of the total, whereas a contribution of M5 buildings and RCW buildings is considerable with 25.51% and 20.01% of the total, respectively. Comparing two parts of Figure 6 shows the biggest difference between the two cities lies on the M3 and M5 building types; while Neuchatel has a majority of M3 buildings, a large portion of M5 buildings can be found in Yverdon-Les-Bains. That fact can be explained by presence of some residential areas, where the same building types (mostly M5 and M6) are built in Yverdon-Les-Bains (please see Figure 5).

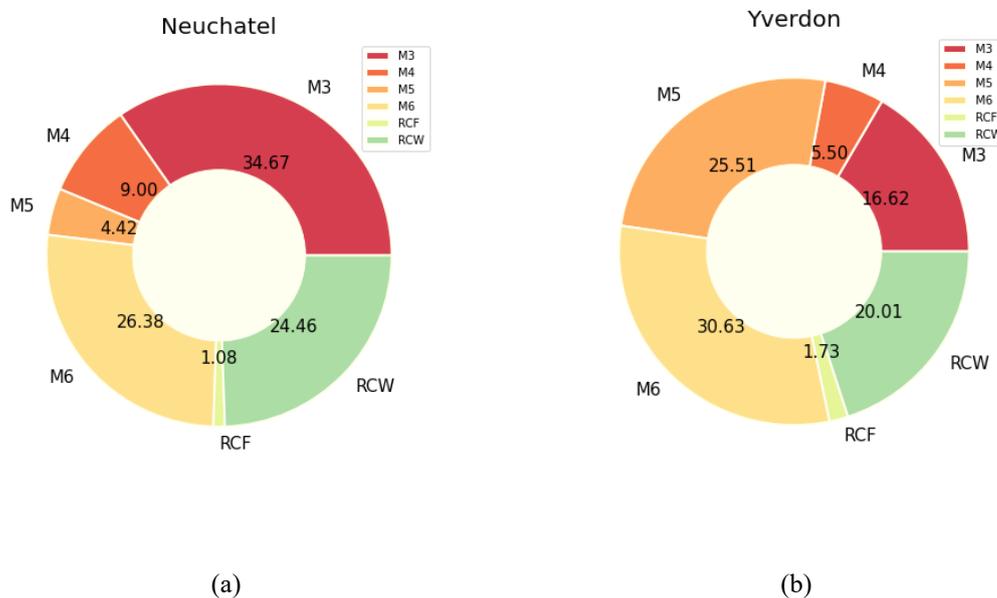


Figure 6 - Distribution of building types: (a) Neuchatel; (b) Yverdon-Les-Bains.

Focusing on height classes, three different distributions of building types are obtained, as shown in Figure 7 and in Figure 8. Regarding the low-rise building in Neuchatel, the majority of buildings constitutes the M3 building type, as it represents the 37.33% of the total. The second majority is represented by M6 buildings, with 32.78%, and the third one is constituted by RCW buildings, with 17.96%. The remaining buildings belong to M4, M5 and RCF building types, as a minority. Considering the mid-rise class of building, the distribution is similar to the one of low-rise buildings, except for a relevant decrease of M6 buildings, that represent 22.76% of the total, and a complementar increase of RCW, that represent 26.43% of the total. On the other hand, M3 still represents the majority of buildings, with 34.27% of the total. Moving to the last height-class, we can notice a substantial difference with respect on the other two. Indeed, the RCW buildings are more than a half of the total number of buildings,

with 57.14%. After that, the second majority is represented by M3 buildings, with 17.73% and the remaining buildings belong to M4, M5, M6 and RCF building types.

The first two majorities of low-rise buildings in Yverdon-Les-Bains are represented by M5 and M6 buildings, with 31.52% and 30.74% of the total, respectively. After that, the third and fourth biggest portions belong to M3 and RCW building types, with 17.18% and 14.80% of the total. The remaining buildings belong to M4 and RCF building types, as a minority. Considering the mid-rise building, the first two majorities of buildings are represented by M6 and RCW buildings, with 32.69% and 31.47% of the total, respectively. After that, 16.61% of buildings belong to M3 category, and M4, M5 and RCF represent a minority. Moving to the last height-class(i.e., high-rise buildings), we can notice a substantial difference with respect to the other two. Indeed, the RCW buildings have a great contribution of 86.30%.

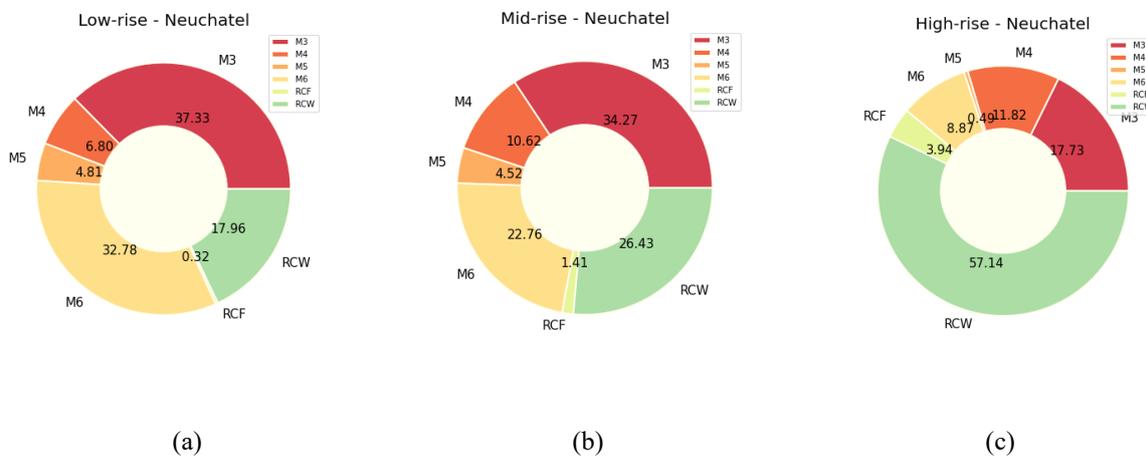


Figure 7 – Distribution of building types per height class in Neuchatel: (a) Low-rise; (b) Mid-rise; (c) High-rise.

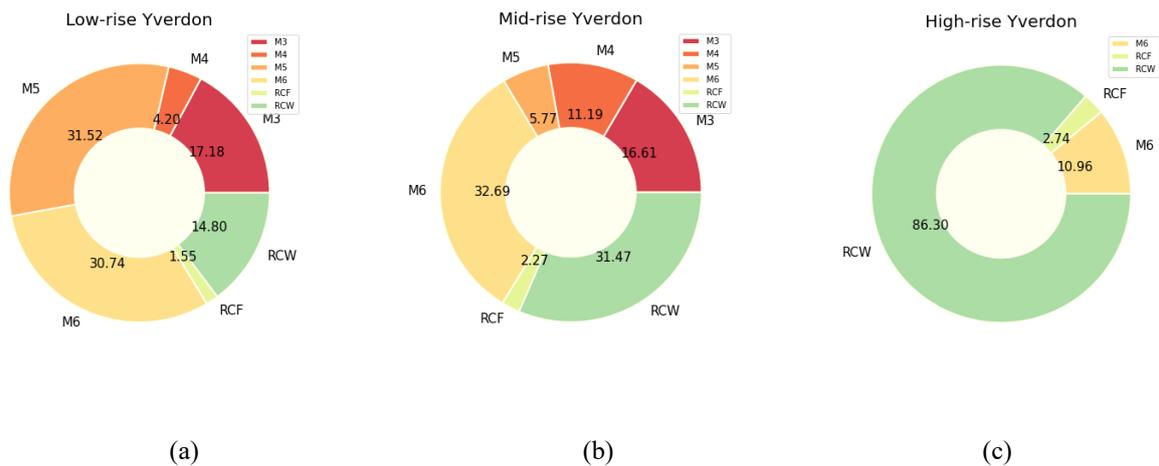
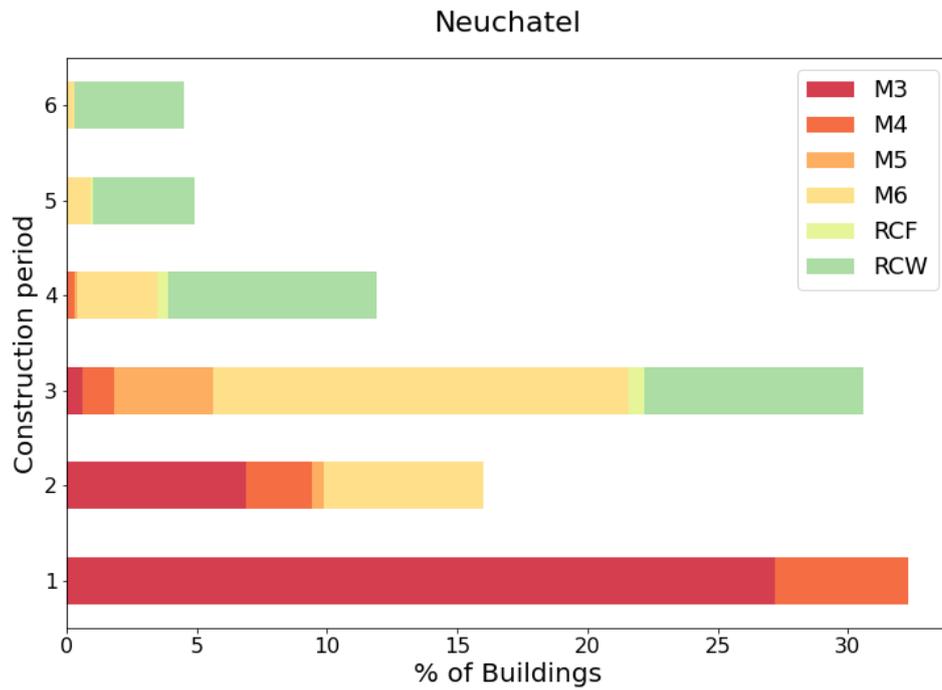


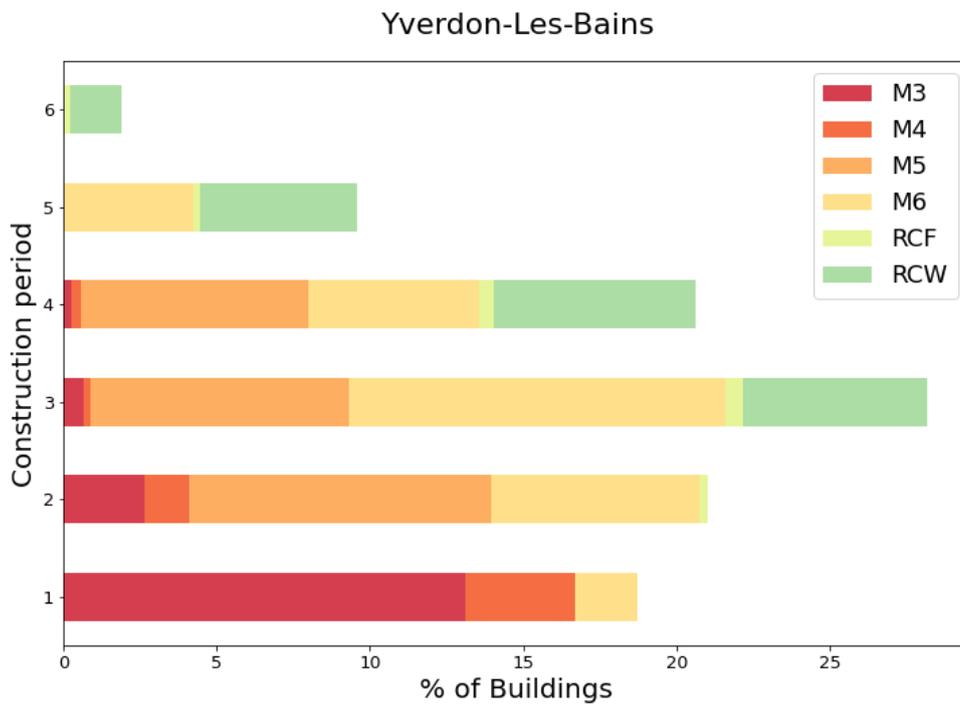
Figure 8 - Distribution of building types per height class in Yverdon-Les-Bains: (a) Low-rise; (b) Mid-rise; (c) High-rise.

2.4.2 Distribution of period of construction

In this section, the important feature of the period of construction is investigated. It is one of the fundamental characteristic to be considered for the building type detection. Indeed, considering materials that were available in different period of time, it could be a very reliable characteristic to find the type of building. Figure 9 shows the distribution of period of construction, categorised and presented in Table 3. Concerning Neuchatel, as shown in Figure 9(a), it is clear that the masonry buildings (M3, M4) were mainly constructed in the periods 1 and 2. Furthermore, the M5 and M6 buildings were constructed in the periods 2, 3 and 4. Concerning Yverdon-Les-Bains, as shown in Figure 9(b), most of masonry buildings (M3 and M4) were constructed in period 1. After that, the majority of M5 and M6 buildings were constructed in the range of 2-4. Finally, the RC structures are constructed from period 3 to 6.



(a)



(b)

Figure 9 - Distribution of building per period of construction: (a) Neuchatel; (b) Yverdon-Les-Bains.

2.4.3 Mapping schemes

Table 5 and Table 6 present the mapping schemes of Neuchatel and Yverdon-Les-Bains, respectively. They are obtained by considering two fundamental building features (i.e., period of construction and height class), categorized and presented in Table 3 and Table 4. The data are written in percentage so that 100% is obtained for each line (Construction period *i*, height class *j*). It is seen that the tables are almost diagonal, meaning by moving from the construction period of 1 to 6, the building type varies from M3 to RCW. Making a comparison between the two cities, it is possible to observe that in Neuchatel there are three height classes for each period of construction (L, M, H), whereas in Yverdon-Les-Bains the height class H is missing for the periods of construction 1 and 2. This means that no high-rise building, constructed before 1945 in Yverdon-Les-Bains are available in the dataset. Looking at the distribution of buildings for each category (Construction period *i*, height class *j*), we can say that the period of construction 3 is the most widespread as different building types were built during that period; nevertheless, M6 and RCW are the most common building types. From period of construction 4, concrete buildings constitute the majority of buildings.

Table 5 - Mapping scheme of Neuchatel based on the two features of the construction period and height class

		BUILDING TYPE					
CONSTRUCTION PERIOD	HEIGHT CLASS	M3	M4	M5	M6	RCF	RCW
1	H	70.0	30.0	0.0	0.0	0.0	0.0
	L	88.5	11.5	0.0	0.0	0.0	0.0
	M	82.1	17.9	0.0	0.0	0.0	0.0
2	H	50.0	20.0	0.0	30.0	0.0	0.0
	L	50.9	13.0	3.7	32.4	0.0	0.0
	M	31.6	19.0	2.6	46.8	0.0	0.0
3	H	5.3	17.5	1.8	15.8	8.8	50.9
	L	2.2	2.0	12.4	64.2	0.2	18.9
	M	1.3	3.9	13.9	45.1	2.8	32.9
4	H	0.0	0.0	0.0	7.5	3.8	88.7
	L	0.0	0.5	1.1	41.9	2.2	54.3
	M	0.0	4.5	0.0	16.4	4.0	75.1
5	H	0.0	0.0	0.0	7.1	3.6	89.3
	L	0.0	0.0	0.0	17.9	0.0	82.1
	M	0.0	0.0	0.0	21.5	2.8	75.7
6	H	0.0	0.0	0.0	0.0	0.0	100.0
	L	0.0	0.0	0.0	10.0	0.0	90.0
	M	0.0	0.0	0.0	3.7	0.0	96.3

Key: please see Table 3 and Table 4 for the categorization of construction period and height class.

Table 6 - Mapping scheme of Yverdon-Les-Bains, based on the two features of construction period and height class

CONSTRUCTION PERIOD	HEIGHT CLASS	BUILDING TYPE					
		M3	M4	M5	M6	RCF	RCW
1	L	75.5	12.5	0.0	12.0	0.0	0.0
	M	56.0	36.0	0.0	8.0	0.0	0.0
2	L	13.0	6.6	48.4	30.6	1.3	0.0
	M	8.9	8.9	28.9	53.3	0.0	0.0
3	H	0.0	0.0	0.0	10.5	2.6	86.8
	L	2.8	0.7	38.4	41.1	1.7	15.3
	M	1.2	1.7	8.1	59.3	2.9	26.7
4	H	0.0	0.0	0.0	12.5	0.0	87.5
	L	0.7	1.1	45.6	29.0	2.5	21.1
	M	4.9	2.9	5.9	22.5	2.9	60.8
5	H	0.0	0.0	0.0	0.0	0.0	100.0
	L	0.0	0.0	0.0	49.5	1.1	49.5
	M	0.0	0.0	0.0	32.9	5.1	62.0
6	H	0.0	0.0	0.0	0.0	100.0	0.0
	L	0.0	0.0	0.0	0.0	12.1	87.9
	M	0.0	0.0	0.0	0.0	5.3	94.7

Key: please see Table 3 and Table 4 for the categorization of construction period and height class.

2.5 INVESTIGATION ON THE OTHER BUILDING FEATURES

In the subchapter 2.4, distribution of period of construction and height classes for different building types are shown. In this section, we focus on the other features that are listed in Table 2 for understanding their importance in the process of classification of buildings. First of all, only the features that are significant for the classification of buildings are selected. As it is listed in Table 7, 17 features have been selected, as they have been considered good indicators of the building type. Among these, 13 are numerical features and 4 are categorical features. This means that in the first group, the values are numerically meaningful, whereas in the second group, the features' value indicates a category of that feature. The two groups will be treated in different ways in the further analyses.

Table 7 – Building features used for the building type classification

ABBREVIATION	DESCRIPTION	TYPE
FOOTPRI	Area of the building footprint in m ²	Numerical
NOSTORIES	Number of stories (only ground floor and upper floors)	Numerical
GEBHOHE	Average height above ground in m	Numerical
GEBVOL	Building volume in m ³	Numerical
ANZWHG02	Number of housing units	Numerical
KUMWAZIM	Cumulative number of rooms in housing units	Numerical
KUMWAREA	Cumulative area of housing units in m ²	Numerical
EINWMOD	Number of inhabitants	Numerical
ANGMOD	Equivalent full time employees	Numerical
PUBLMOD	customers capacity for selected	Numerical
GEBWERT	Replacement value in CHF	Numerical
INHWERT	Value of mobile goods in CHF	Numerical
UMSATZ	Production account as a proxy for the yearly sales revenue in CHF/year	Numerical
GKAT	Category of the building	Categorical
GKLAS	Classification of the building	Categorical
GBAUP	Period of construction	Categorical
ROOFTYPE	Roof inclination	Categorical

2.5.1 Numerical features

The first group of features includes numerical features. First of all, the pair-wise correlations of the selected numerical features are calculated to detect any potentially multicollinear variables. In Figure 10, a heat-map of pair-wise correlations for Neuchatel and Yverdon-Les-Bains are shown; the darkest regions correspond to a low level of correlation while the lightest regions correspond to a high grade of correlation. In both the graphs, it is possible to identify some features as highly correlated. The first block of highly correlated features are “Number of inhabitants”, “Cumulative area of housing units in m²”, “Cumulative number of rooms in housing units”, “Number of housing units”. The second one is constituted by “Area of the building footprint in m²”, “Building volume in m³”, “Replacement value in CHF”, “Value of mobile goods in CHF”, “Production account as a proxy for the yearly sales revenue in CHF/year” and “Equivalent fulltime employees”. It is possible to say that the analysis of the pair-wise correlations of the selected numerical features in the two cities lead to the same result. If we had a limited computation power, it would have been reasonable to drop the highly correlated features for modelling; considering only one or two features per each block could be an option, as suggested by (Mumtaz, 2020).

Another investigation is carried out to measure the importance of each feature on the identification of building types. By looking at the feature importance, it is possible to decide which features have less contribution (or even no contribution) in the prediction process (Donges Niklas, 2021). The features in Figure 10 are presented in order of importance by applying the ANOVA test. Six numerical features with the biggest contribution are: (1) Number of stories; (2) Number of housing units; (3) Cumulative area of housing units in m²; (4) Cumulative number of rooms in housing units; (5) Number of inhabitants; (6) Average height above ground in m. However, if we wanted to create mapping schemes of a city by considering only two features, we would have had to use "Number of inhabitants" and "Cumulative area of housing units in m²" for Neuchatel; "Average height above ground in m" and "Cumulative area of housing units in m²" for Yverdon-Les-Bains. This means that the choice of classifying buildings in an aggregated way with mapping schemes may exclude a lot of information. For this reason, a more sophisticated classification method will be presented and used later, using the principles and tools of deep learning. We will consider all the selected features, presented in Table 7, for having the best model in estimation of building type.

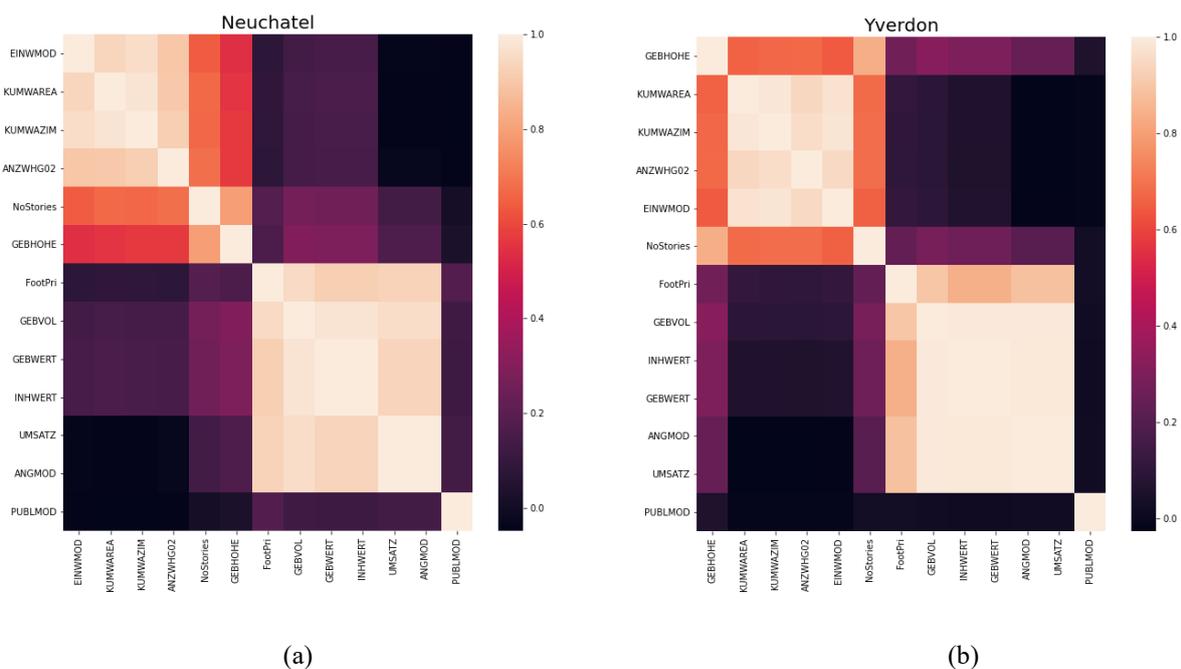
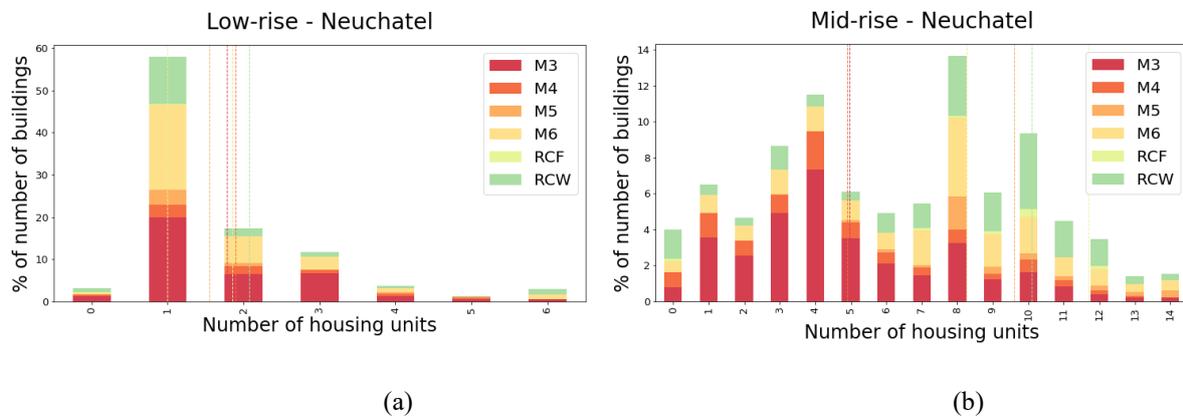


Figure 10 - Heat-map of the pair-wise correlations between the building features: (a) Neuchatel; (b) Yverdon-Les-Bains.

In the next subchapters, some of these features are explored in more details to understand how they contribute in the process of identification of building type.

2.5.1.1 Number of housing units per each building

The number of housing units per each building is a good example to show how features can be used for building typeification. Figure 11 shows distribution of this feature low-rise and mid-rise buildings. Figure 11(a) and (b) refers to Neuchatel dataset, Figure 11(c) and (d) refers to Yverdon-Les-Bains dataset. Concerning the low rise buildings of Neuchatel, the mode is 1 housing unit per building. At the same time, for all different building types, the average is between 1 and 2 housing units per building. Considering the mid-rise buildings of Neuchatel, the mode is 8 housing units per building. The average lines are quite spread and they are different for the different building types; the average is 5 for M3 and M4 buildings while the average for M6 and RCW buildings are 8 and 10 housing units per building, respectively. Regarding Yverdon-Les-Bains, the mode is 1 housing units per building for the low-rise buildings and the averages for different building types are between 1 and 3 housing units per building while the averages are quite spread and different for the mid-rise buildings; the average for M3 and M4 buildings is between 4 and 5 while the averages for M6, RCF and RCW are between 13 and 19 housing units per building. In conclusion, the feature of "Number of housing units per building" is useful to distinguish structural types of mid-rise buildings, for which there is a clear distinction; however, that feature could not do that for Low-rise buildings as the number of housing units for low-rise building is in the same range for all building types.



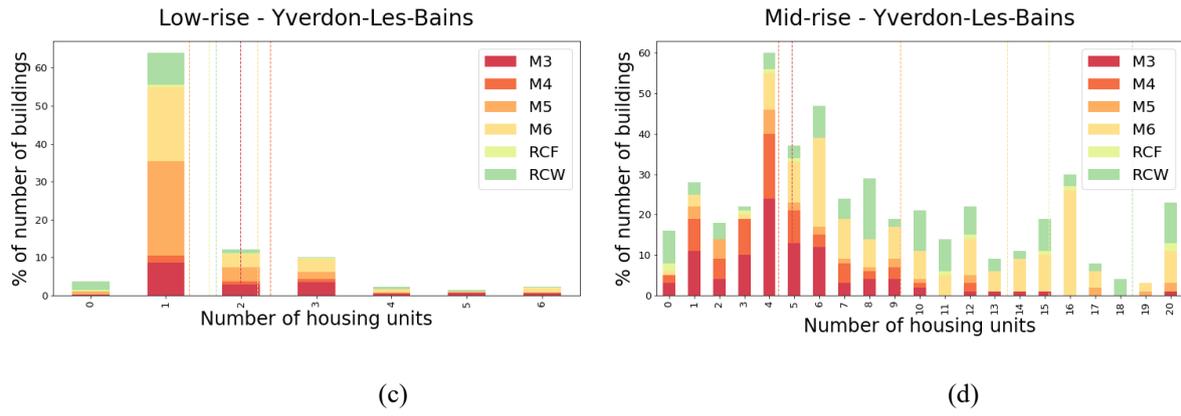


Figure 11 - Distribution of number of housing units for (a) Low-rise Neuchatel; (b) Mid-rise Neuchatel; (c) Low-rise Yverdon-Les-Bains; (d) Mid-rise Yverdon-Les-Bains.



3 THE RANDOM FOREST CLASSIFIER

METHOD (RF)

In subchapter 2.5, a general description of building features, that are listed in Table 7, is presented. In Chapter 3, a deeper analysis of these features is performed. In this chapter, a supervised learning algorithm is applied to predict building types based on the building features. The method is called Random Forest (RF), it is a supervised learning algorithm and it can be used both for classification and regression. It is also a flexible and easy-to-use algorithm.

RF has been widely used in areas of geography, economics, medicine, and engineering. It can be used to classify loyal loan applicants, identify fraudulent activity and predict diseases (Navlani, 2018). Fan et al. (2013) extracted building geometrical features from LiDAR point clouds using the RF method. Bosch et al. (2007) explored the problem of classifying images by the object categories by combining RF classification and multi-way SVM. For its applications in remote sensing, Ham et al. (2005) applied two methods inside a binary hierarchical multi-classifier system, generalizing the RF classifiers in an analysis of hyperspectral data, when the number of training samples is small. In our research, we implement RF method to detect building type from the building attributes.

3.1 DESCRIPTION OF THE METHOD

RF method is an ensemble of random decision tree classifiers, which discriminate between different classes based on features (Ho, 1995). The final prediction is made by combining the predictions of individual trees, that form the decision forests. In other word, a decision forest includes a set of expert tree classifiers and all of these classifiers would entirely vote for the most probable class of an input vector of features (Ho, 1998). There are different approaches to introduce randomness in the decision tree construction method. For example, we randomly select subsamples from the dataset (Pal, 2005). Generally, the more trees it has, the more robust a forest is. Decision trees are then built based on randomly selected subsamples. An attractive aspect of the RF classifier is that it can avoid overfitting and can rapidly adjust to the training data. Furthermore, even if some information is missing, the RF classifier can still accurately estimate the missing samples and maintain a stable classification performance (Breiman, 2001).

The algorithm works in four steps: (1) Elect training samples from a given dataset. The samples should include all of the building types and contain information on building type. (2) Construct a decision tree for each subsample. Then, every tree would modify itself and select the optimal type and branches to have the best performance in prediction. (3) Assign a vote

for each predicted result. (4) Select the prediction with the most votes as a final prediction (Huang, 2017). The procedure is schematically depicted in Figure 12.

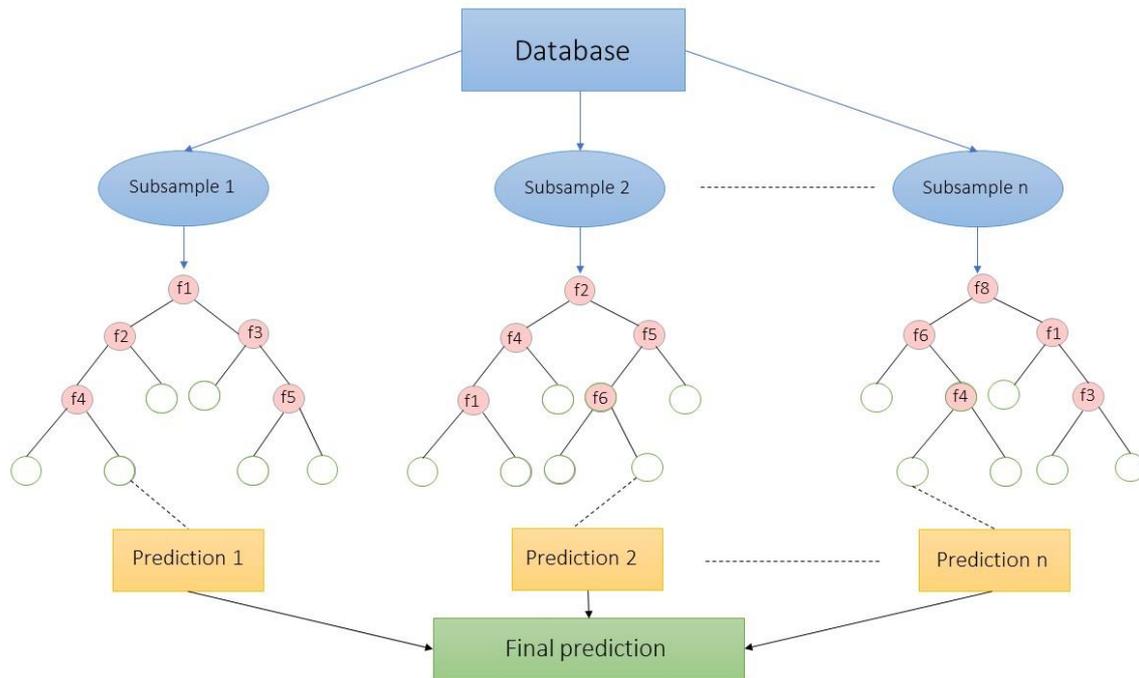


Figure 12 - RF diagram (Viet-Hung Dang, 2020)

Summing up, RF is one of the best performing predictive models (Vens, 2013). It has several advantages: (1) It is considered as a accurate and robust method because of the number of decision trees participating in the process. (2) It does not suffer from the overfitting problem. The main reason is that it takes the average of all the predictions, which cancels out the biases. (3) The algorithm can be used in both classification and regression problems. (4) RF can also handle missing values. There are two ways to handle these: using median values to replace continuous variables and computing the proximity-weighted average of missing values. (5) It is possible to get the relative feature importance, which helps in selecting the most contributing features for the classifier (Navlani, 2018). On the other hand, it also has some disadvantages: (1) It might be slow in generating predictions because it has multiple decision trees. Whenever it makes a prediction, all the trees in the forest have to make a prediction for a given input and then perform voting on it. This whole process is time-consuming. (2) The model is difficult to interpret compared to a decision tree, where it is possible to easily make a decision by following the path in the tree (Navlani, 2018).

The Scikit-learn, an open-source Python module for machine learning, is used to implement the RF classifier (Pedregosa et al., 2012). The dataset of labelled samples is randomly divided into two sets: a training sample set for model calibration and a test set for

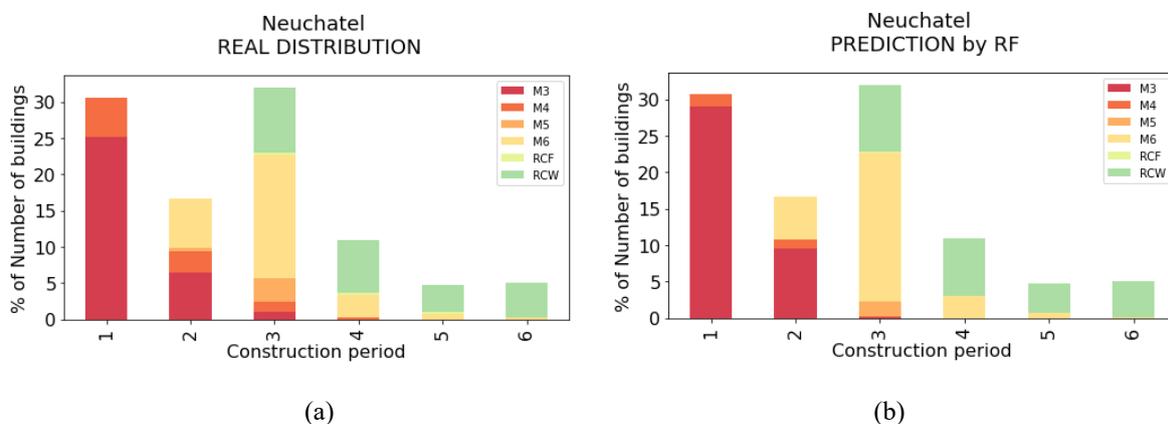
validation. In the present study, 80% of the dataset is used for the training sample, 20% of the dataset is used for as the test set. However, some simulations have been carried out in order to find the perfect balance leading to the best accuracy.

3.2 APPLICATION OF THE RF METHOD ON NEUCHATEL AND YVERDON-LES-BAINS

In the following section, we describe two RF models developed based on the building datasets of Neuchatel and Yverdon-Les-Bains, seperately. The dataset encompasses records of 3533 and 2808 buildings in Neuchatel and Yverdon-Les-Bains, respectively. All features listed in Table 7 are employed for development of RF model. Numerical and categorical features are important to be used in the modelling. In categorical variables, the different values do not have significant numerical relation with each other. Dummy indexes are, therefore, created for the categorical features. The purpose of the dummy variables is to convert categorical features into numerical values, which allows a clear interpretation of the categorical data by the RF model. That procedure can be explained with an example from (Mumtaz, 2020); Considering a categorical feature called grade with the following unique values in the dataset: A, B, C, and D. It does not make sense to assign numbers of 1 to 4. Therefore, the column “grade” will be split in four columns (i.e., A, B, C, D) with only value of 0 or 1. It is reasonable to create a variable called “A” and interpret it as meaning that someone assigned a 1 on this variable is A and someone with an 0 is not.

3.2.1 Results

Considering the test datasets, a comparison between the real and predicted distribution of building types for the two cities of Neuchatel and Yverdon-Les-Bains is shown Figure 13(a-b) and Figure 13(c-d), respectively. The models provide a good prediction of the building types in general. The biggest confusion is found between M3 and M4 in period of construction of 1 and 2. That could be explained by the low population of M4 building type.



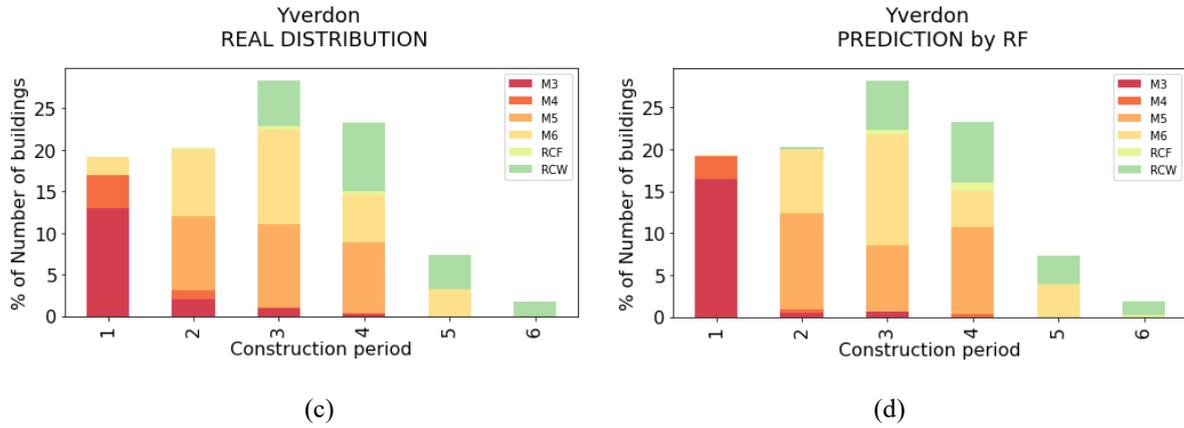


Figure 13 - Real and predicted distributions of building types in (a-b) Neuchatel and (c-d) Yverdon-Les-Bains.

Focusing on height classes:

Neuchatel

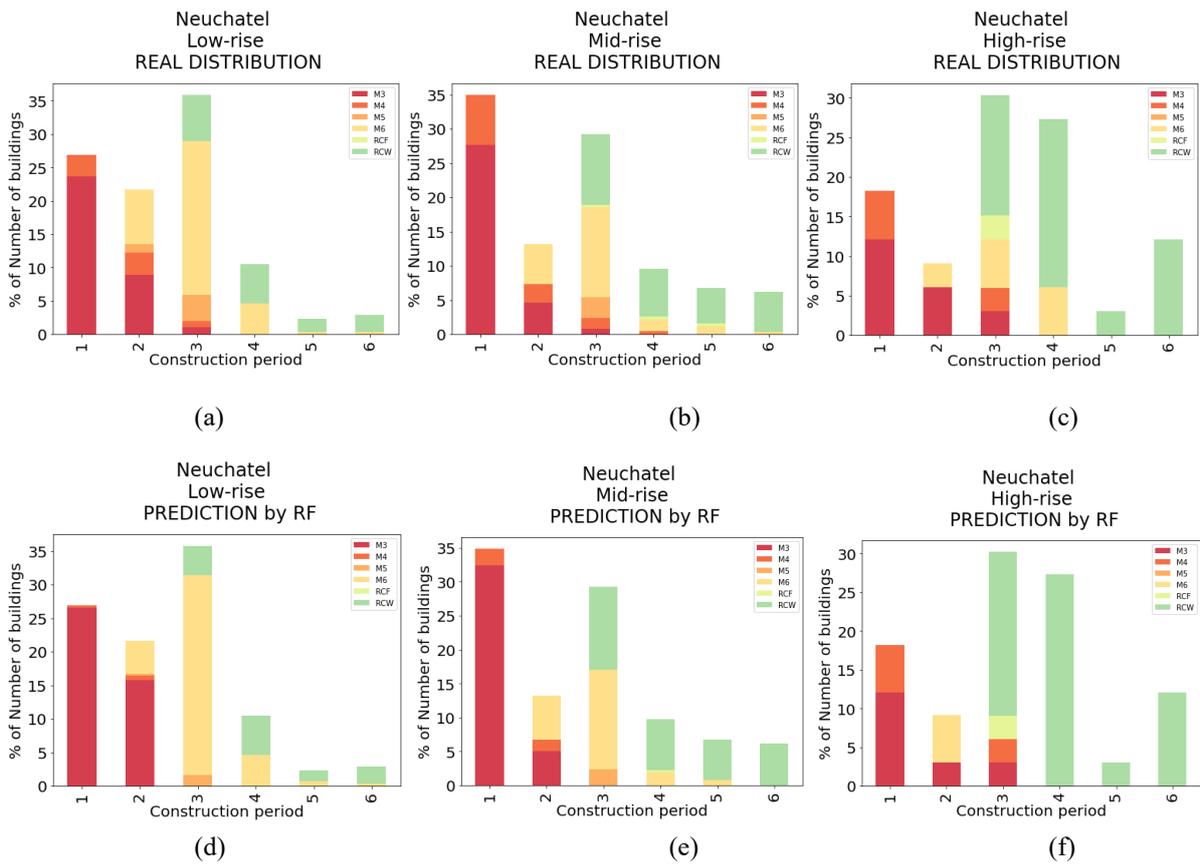


Figure 14 - Real and predicted distributions of building types with different height classes in Neuchatel

Yverdon-Les-Bains

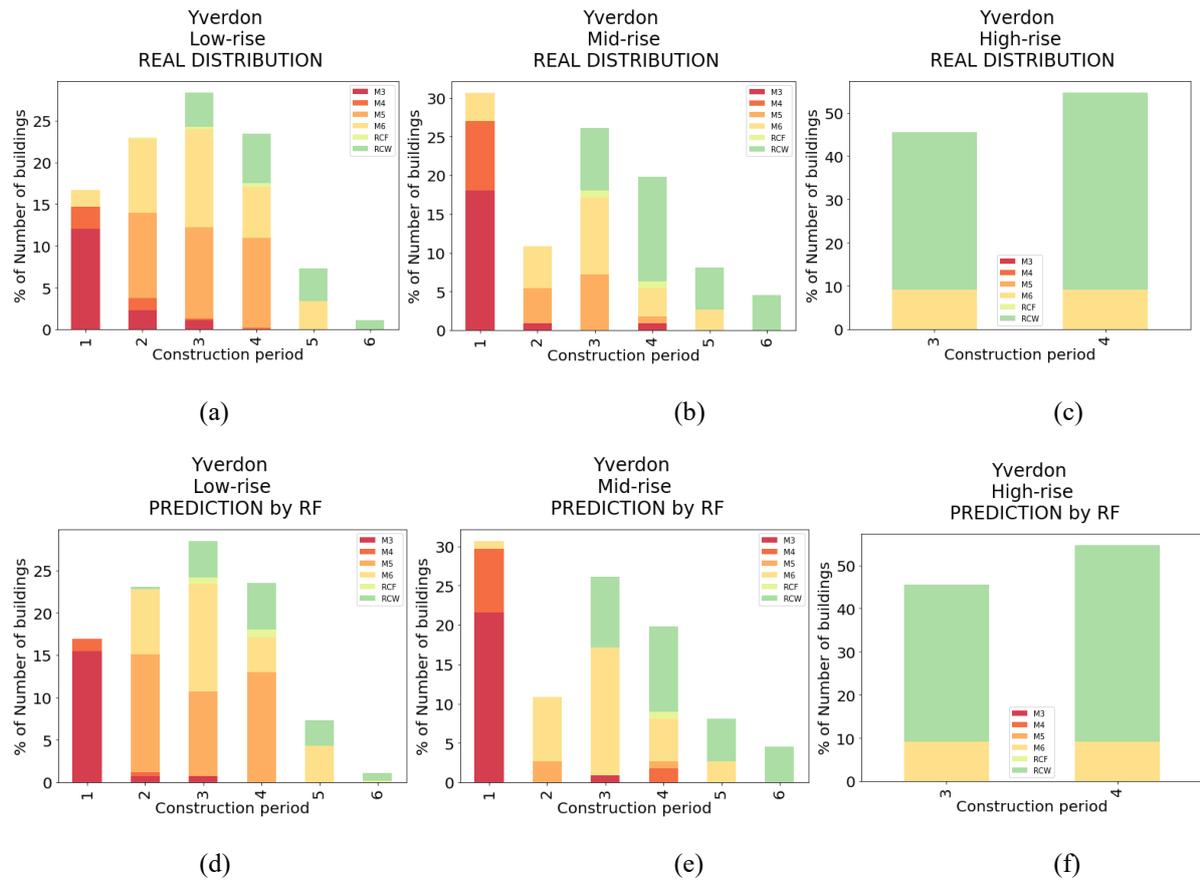


Figure 15 - Real and predicted distributions of building types with different height classes in Yverdon-Les-Bains

3.3 ACCURACY OF RF

To evaluate the performance of the RF, the Confusion Matrix (CM), also known as error matrix, is obtained (Figure 16). The CM is a $N \times N$ matrix; where N is the number of building types. A CM cell indicates the number of test samples for each combination of ground-truth building types (N) and assigned building types (N). The diagonal elements show the number of buildings that have been correctly predicted by the RF model. On the off-diagonal elements show the number of buildings that have been incorrectly predicted by the RF model. It is important to mention that the fewer building number for each building type, the harder it will be for the model to predict that building type. For this reason, the biggest error is expected for the building types M4 and RCF, which have a lowest contribution on the portfolio.

Concerning Neuchatel, the major errors in the prediction are identified between M3 and M4, rather than error is also identified between M6 and RCW. Concerning Yverdon-Les-Bains, the biggest mismatching is identified between M5 and M6. Considering the two CMs, it should be mentioned that the Yverdon-Les-Bains RF model was able to correctly classify M3 buildings for 77% of the cases while the Neuchatel RF model reached 88% of the cases. Concerning M6, 78% of buildings is correctly identified in the case of Neuchatel whereas that is 56% of buildings in case of Yverdon-Les-Bains. On the contrary, M5 buildings are correctly classified in 65% of the case Yverdon-Les-Bains model as opposed to 23% for RF Neuchatel model. This might be related to the fact that the number of M5 buildings in Yverdon-Les-Bains is higher than those in Neuchatel (154 vs 35). Finally, RCW are equally identified with success rates of 86% and 85%, in the two cases of Neuchatel and Yverdon, respectively. M4 and RCF have a small contribution on the training and test sets and consequently, the highest level of error is expected for those building types.

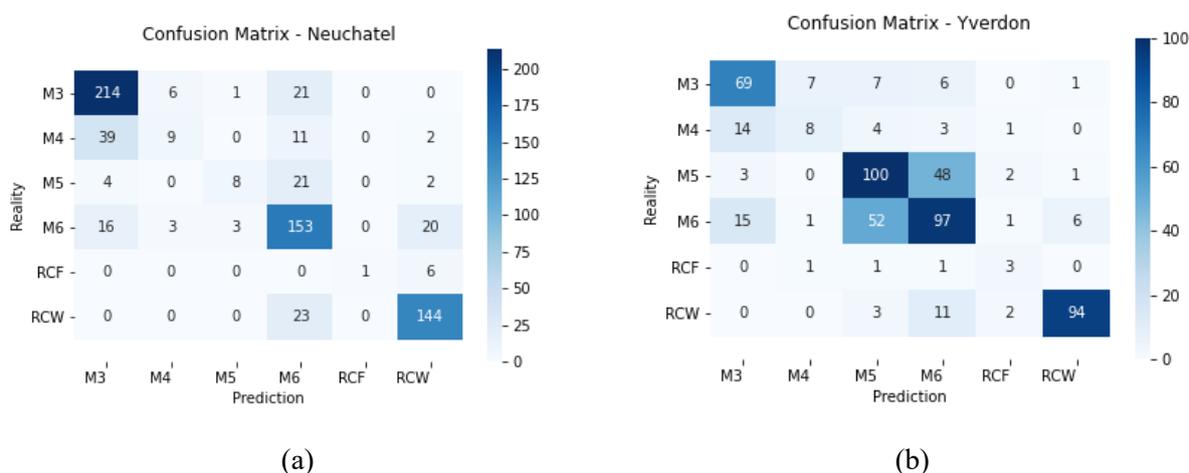


Figure 16 - Confusion matrices for (a) Neuchatel; (b) Yverdon-Les-Bains

The results of the RF models are evaluated with three accuracy measures: the first one, AM1, is overall accuracy of building types, which is based on the confusion matrix; the second one, AM2, is obtained as the weighted average of the precisions for each building type; the third one, AM3, is based on the distribution of building types. AM1 is calculated as the ratio between the number of correctly detected buildings to the total number of buildings, as given by:

$$AM1 = \frac{\sum_i A_{ii}}{\sum_i \sum_j A_{ij}} \quad (1)$$

Moreover, one parameter is directly presented from the algorithm for the evaluation of precision of the RF method: the weighted average of the precision (AM2). It is calculated by weighted averaging the precisions for each type with respect to the number of buildings on each type. AM3 is focused on the general distribution of building types; AM3 is calculated based on the difference (in number of buildings) between the “real” and the “predicted” distributions, as presented in equation 2. The real distribution of building types is the outcome of visual surveying (the test sample set) and the predicted distribution is provided by the RF model. These distributions and their differences for Neuchatel and Yverdon-Les-Bains are presented in Table 8 and Table 9.

$$AM3 = 1 - \frac{\sum_k |N_k^{real} - N_k^{pre}|}{\sum_k N_k^{real}} \quad (2)$$

k: building types

Table 8 – Real and predicted distributions of building types and its difference (Neuchatel)

	M3	M4	M5	M6	RCF	RCW
REAL DISTR. (TEST SAMPLE SET)	242	61	27	198	6	175
RF PREDICTED DISTR.	273	18	12	229	1	174
Δ (IN ABSOLUTE VALUE)	43	53	15	31	5	1

Table 9 - Real and predictd distributions of building types and its difference (Yverdon-Les-Bains)

	M3	M4	M5	M6	RCF	RCW
REAL DISTR. (TEST SAMPLE SET)	90	30	154	172	6	110
RF PREDICTED DISTR.	101	17	167	166	9	102
Δ (IN ABSOLUTE VALUE)	11	13	13	6	3	8

The three accuracies for Neuchatel and Yverdon-Les-Bains are reported in Table 10. AM1 and AM2 are quite similar; the main difference between the AM1 and AM3 is the fact that AM1 is evaluated at building-by-building level and based on the confusion matrix and all mispredicted buildings classification are considered as errors, although AM3 is based on the general distribution of building types. Thus, errors (whether false positives and false negatives) can compensate each other without impacting the final result. The lower value of AM1 in comparison to AM3 are therefore expected. It should be mentioned that seismic risk assessment is usually carried out for a big region; an estimate of buildings at city/district level, rather than building level, is considered. AM3 could be effective to show the performance of RF prediction with that generic purpose of risk calculations.

Table 10 - Accuracy measurements for the Neuchatel and Yverdon-Les-Bains RF model

	NEUCHATEL	YVERDON-LES-BAINS
AM1	0.73	0.65
AM2	0.74	0.66
AM3	0.81	0.91

3.4 PARAMETERS AFFECTING ACCURACY OF RF

In the present study, 80% of the dataset is used for the training purpose, 20% of the dataset is used as a test set. However, some simulations have been carried out for finding the perfect balance leading to the best accuracy. Figure 17 shows that, as the fraction of train set to test set changes, the accuracy measures vary. The ratio of 0.8 should be used to achieve the best accuracy. In this simulation, the dataset of Neuchatel was used.

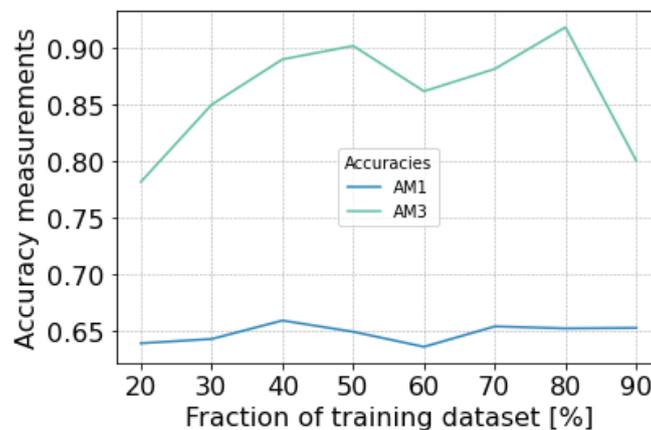


Figure 17 - Accuracy variation according to the training set size

Another parameter that may influence the accuracy of the RF model is the number of decision trees. Several simulations are performed and starting from about 450 trees, the accuracy becomes stable, as shown in Figure 18. In the present study, 600 trees are used for in the training phase of the RF models.

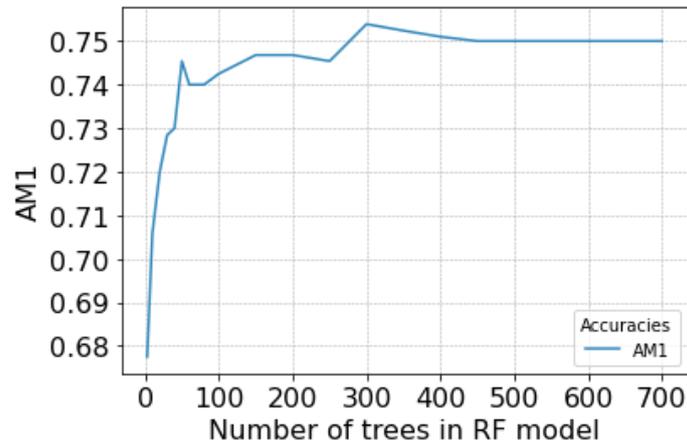


Figure 18 - Accuracy variation by changing number of trees in RF model



4 APPLICATION OF THE RF MODEL TO OTHER CITIES

In the previous chapter, the RF method has been applied on the two building datasets of Neuchatel and Yverdon-Les-Bains, separately. In the following chapter, a RF model, trained based on the concatenated datasets of Neuchatel and Yverdon-Les-Bains, is applied to two other Swiss cities (i.e., Solothurn and Visp) in order to validate the proposed method's applicability in other areas.

Case study: Solothurn and Visp

Solothurn is located in the north-west of Switzerland on the banks of the Aare and on the foot of the Weissenstein Jura mountains. Visp is located in the Canton Valais, it lies in the Rhône valley and 9 km west of Brig-Glis. The locations of Neuchatel, Yverdon-Les-Bains, Solothurn and Visp are shown in Figure 19.

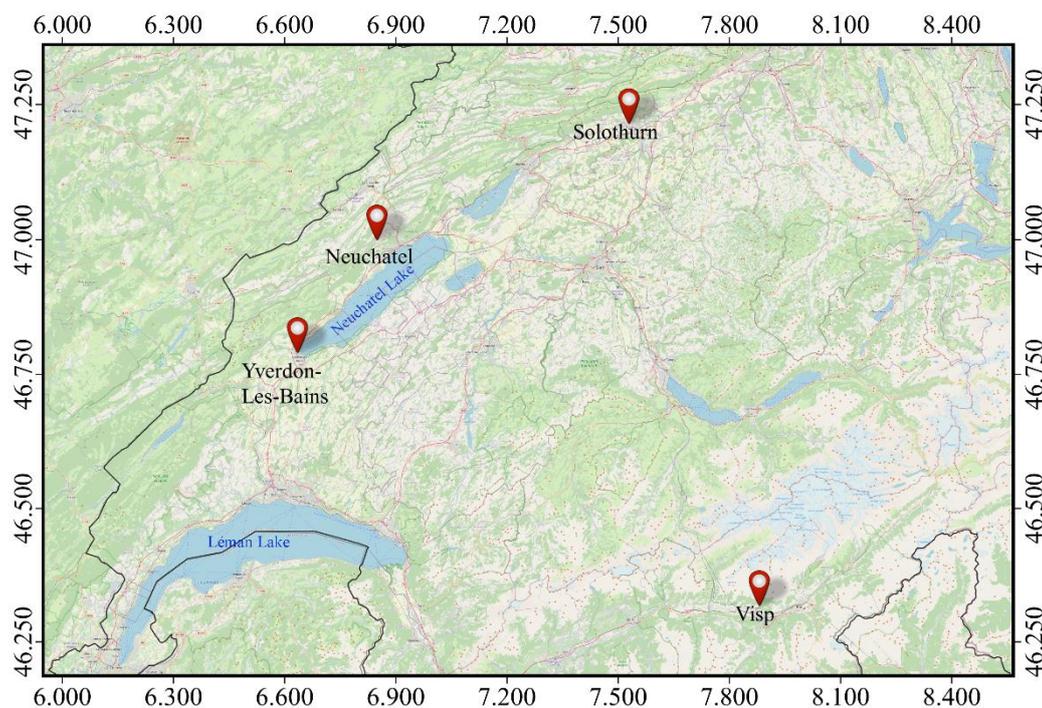


Figure 19 - Neuchatel, Yverdon-Les-Bains, Solothurn and Visp. The Swiss border is highlighted in black lines.

4.1 SURVEY

Before applying the RF model to the cities of Solothurn (3238 buildings) and Visp (307 buildings), surveys have been carried out for providing a ground-truth datasets. The survey of Visp has been performed as a physical visual survey. The survey of Solothurn has been made by using two different techniques. First of all, a general classification has been done using Google satellite and street views. This classification has been carried out by categorizing the buildings according to other features (Footprint and Roof shape). After that, a check on the uncertain data has been performed with a physical visual survey. This technique was only possible thanks to the experience gained with the physical visual surveys. The distributions of building types are shown in Figure 20. Concerning Solothurn, the three main building types are M3, M6, and RCW, with 33.17%, 32.44% and 24.89% of buildings, respectively. A minority of buildings, 9.47%, is constituted by M5. It is possible to notice that only few buildings are M4. Concerning Visp, it has a more diverse distribution. The main building type is M6, with 31.56% of buildings. After that, 35.47% of buildings are RCW, 14.25% are M3. A minority of buildings is M1 (rural buildings) and RCF.

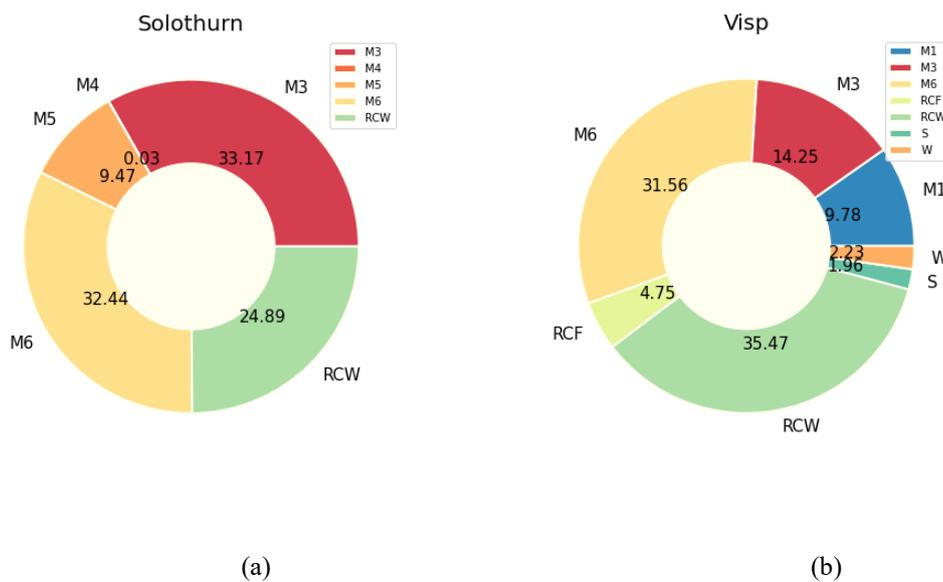


Figure 20 - Distribution of building types: (a) Solothurn; (b) Visp.

It is worth mentioning that the distributions of building types in the two cities are quite different, making them as good candidates for evaluating performance of the RF model on other cities. Indeed, Figure 21 shows a comparison between the distribution of building types of Neuchatel, Yverdon-Les-Bains, Visp and Solothurn and it is clear that the four cities have different distributions of building types. In Valais, because the territory is more prone to earthquake, we have a greater preference to more stronger buildings. In fact, RC buildings in Visp are in majority while the distribution of buildings in Neuchatel, Yverdon and Solothurn are similar. Nevertheless, we will see that RF works well for all these building type distributions.

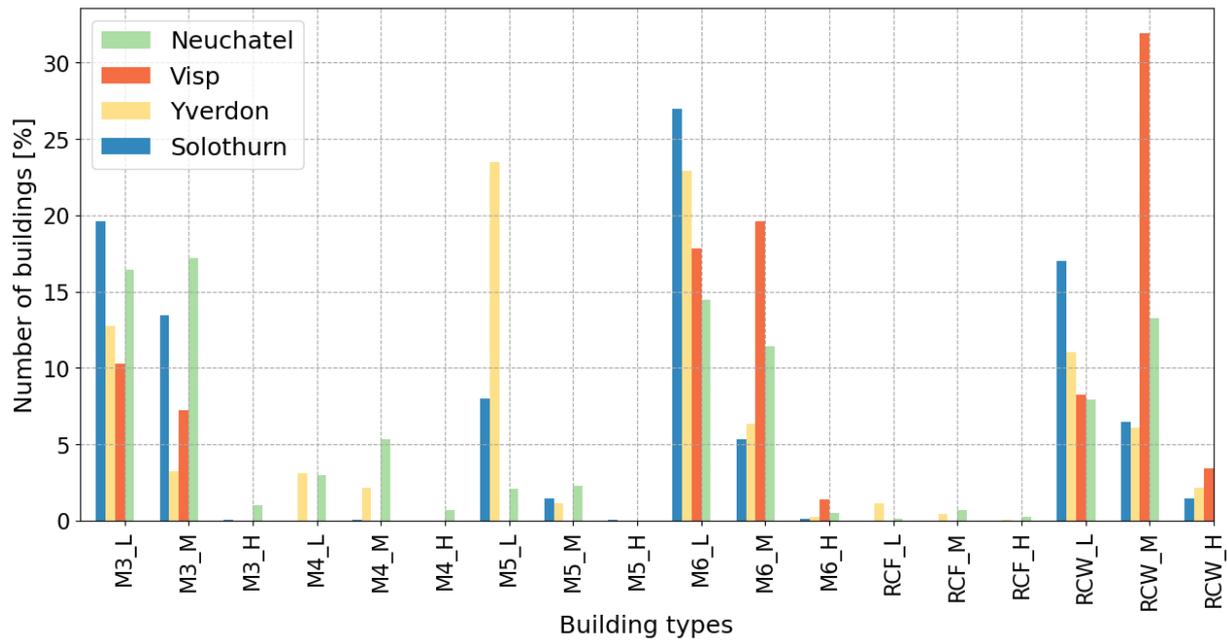


Figure 21 - Distribution of building types in Neuchatel, Yverdon-Les-Bains, Visp and Solothurn.

4.2 THE RF MODEL BASED ON NEUCHATEL AND YVERDON-LES-BAINS DATASETS

In this section, a RF model, based on the joined datasets of Neuchatel and Yverdon-Les-Bains (RF N&Y), is developed. The confusion matrix for the prediction in the test dataset is depicted in Figure 22 and a comparison is carried out between the current model and the previous ones, which are separately trained based on the datasets of Neuchatel and Yverdon-Les-Bains.

A comparison is established between the real building types and the correctly predicted building types by the three models: Neuchatel, Yverdon and N&Y, respectively. For example, concerning M3, the RF N&Y model was able to correctly identify in 86% of the cases while RF Neuchatel and RF Yverdon-Les-Bains reached respectively 88% and 77%. M5 buildings are correctly classified in 48% of the cases by the RF N&Y model whereas a value between 30% and 65% from the RF Neuchatel and the RF Yverdon-Les-Bains models are achieved. Concerning M6, 71% of buildings is correctly identified in the case of RF N&Y, whereas 77% of buildings in case of RF Neuchatel and 56% in case of RF Yverdon-Les-Bains. In overall, we can say that the RF N&Y model performs with an accuracy slightly higher than the average of the RF Neuchatel and the RF Yverdon-Les-Bains. The three accuracy measurements are also calculated for this model and presented in Table 11. The accuracies of the model trained on the joint dataset are really similar to the accuracies of models based on single datasets, reported in Table 10.

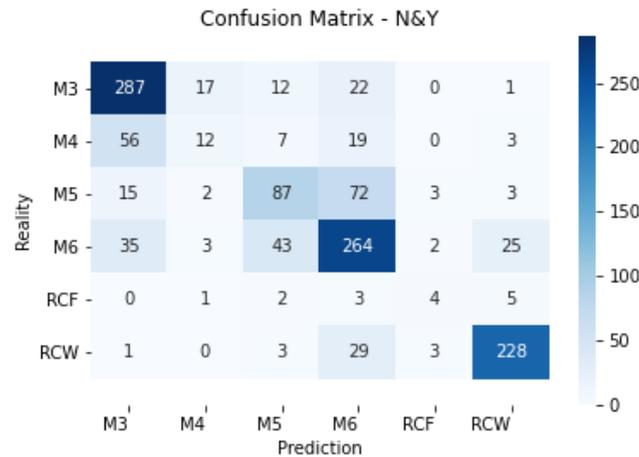


Figure 22 - Confusion matrix of RF model, trained and tested on the concatenated dataset of Neuchatel and Yverdon-Les-Bains

Table 11 - Accuracy measurements of the RF model, trained and tested on the Neuchatel and Yverdon-Les-Bains dataset

	N&Y
AM1	0.70
AM2	0.67
AM3	0.85

4.3 APPLICATION OF RF MODEL TO SOLOTHURN AND VISP

By having the RF model, trained and tested on the concatenated datasets of Neuchatel and Yverdon-Les-Bains, we apply it on the two cities of Solothurn and Visp. For the application of the model, some adjustments have been made on the dataset. First of all, as described in Chapter 3 for Neuchatel and Yverdon-Les-Bains, only the features listed in Table 7 have been considered. Concerning the categorical features, dummies have been created, as described in Chapter 3. In this compound, it is important to underline that the dataset on which the model has been trained and the dataset on which the model will be applied must be exactly harmonized. It means that both of them must have the same features. Due to that, it is necessary to harmonize for full compatibility. For example, some columns concerning specific Building category (GKLAS) have been added, since some of them were not present in both the datasets. Moreover, the model is able to predict the building types that are present in the training datasets (i.e., building datasets of Neuchatel and Yverdon-Les-Bains). With that fact, the M1 building type (rural buildings), that is present in the Visp dataset, is excluded from the analyses. This means that 12.08% of buildings in Visp are not considered for the application of the model.

In Figure 23, a comparison between the real and predicted distribution of building types is presented. Generally, the graphs show that the model well predicts the building types for the two cities. Considering Solothurn, the biggest difference between the real distribution and the prediction is seen in the detection of M4 and in the detection of M5 during the construction periods of 2, 3 and 4. For the city of Visp, the biggest difference between the real distribution and the predicted distribution regards the absence of M3 in the periods of construction 3, 4 and 5 and the absence of RCF in the predicted distribution. Moreover, no M6 are predicted for periods of construction 5 and 6.

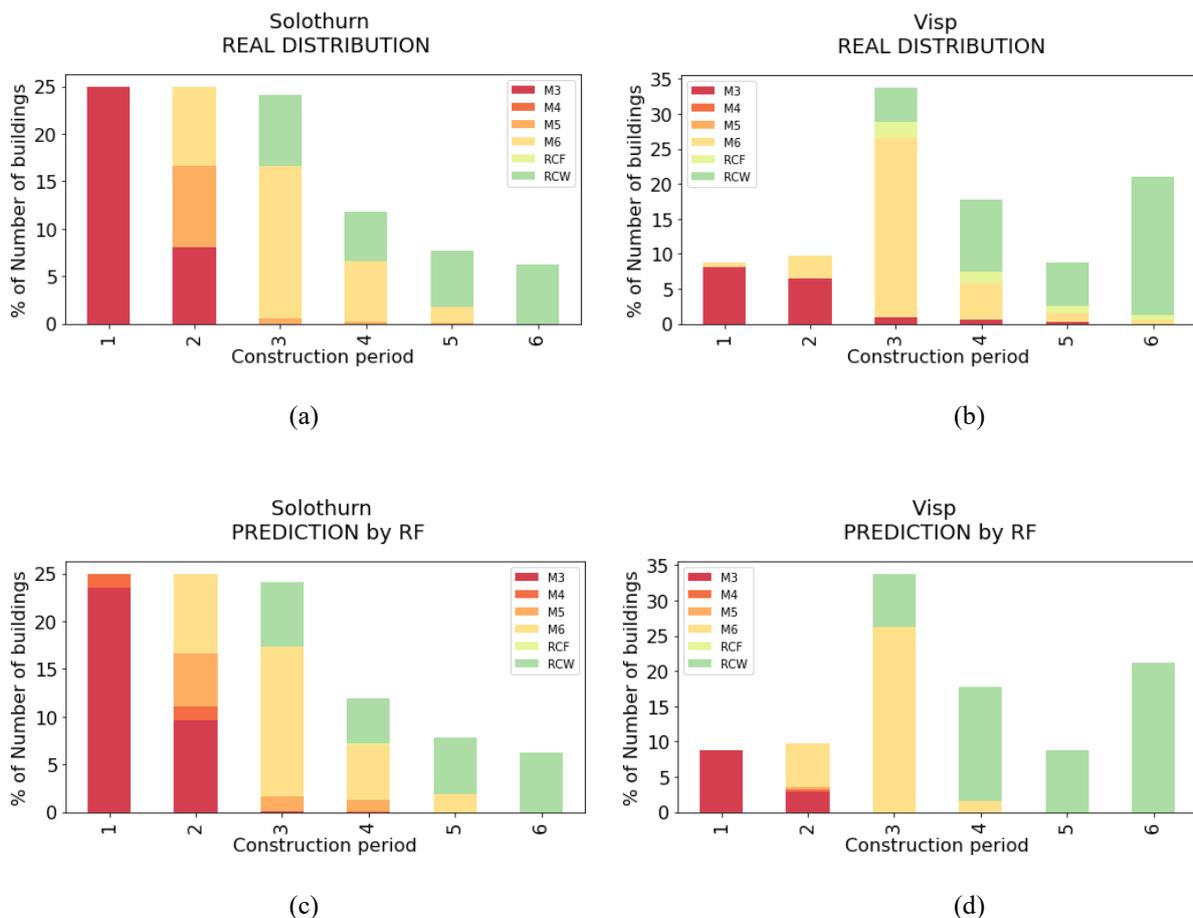


Figure 23 - Application on Solothurn of the RF model trained on Neuchatel and Yverdon-Les-Bains: (a) Solothurn – real distribution; (b) Visp – real distribution; (c) Solothurn – predicted distribution; (d) Visp – predicted distribution.

In Figure 24, the confusion matrix obtained from the application of the RF N&Y model on Solothurn and Visp dataset is presented. Concerning Solothurn, the model was able to correctly identify 84% of M3 buildings, 80% of M6 buildings and 87% of RCW buildings, showing a very good level of prediction. Indeed, there is a good correspondence between the reality and the prediction, as it is possible to see on the diagonal of the confusion matrix. On

the other hand, looking into the misclassification, the biggest errors are present in the detection of M3, M5 and M6, as it is possible to see from the off-diagonal elements in Figure 24. A confusion is also identified in the detection of M6 and RCW. Concerning Visp, the model was able to correctly identify 63% of M3 buildings, 70% of M6 buildings and 95% of RCW buildings. The major uncertainties in this case are identified between M6 and RCW. In Table 12, the accuracy measurements for the application of the model to the two cities are reported.

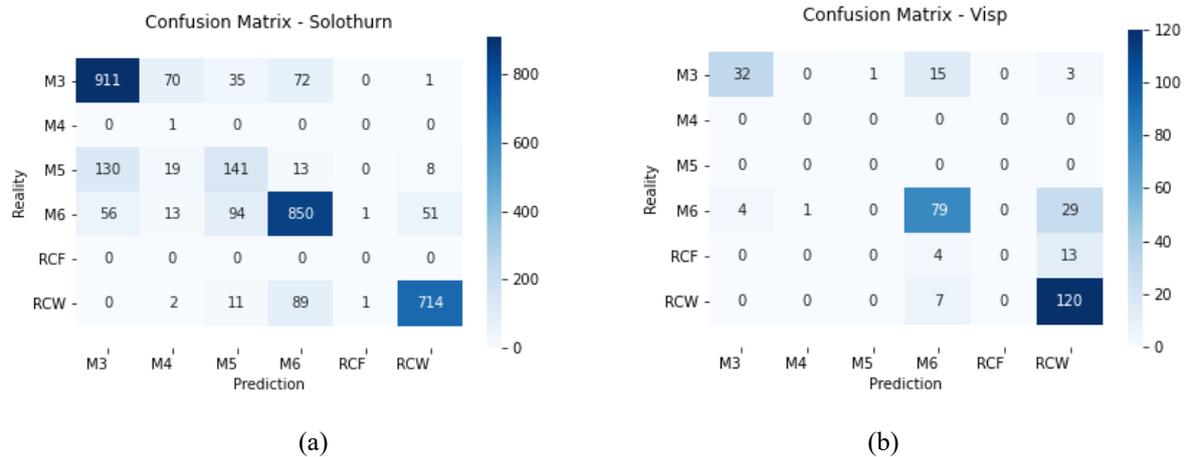


Figure 24 –Confusion matrix of the RF model, trained on the Neuchatel and Yverdon-Les-Bains dataset and applied on (a) Solothurn; (b) Visp.

Table 12 - Accuracy measurements of the RF model, trained on the Neuchatel and Yverdon-Les-Bains dataset and applied on Solothurn and Visp

	SOLOTHURN	VISP
AM1	0.80	0.75
AM2	0.82	0.71
AM3	0.93	0.81

According to Table 12, the RF N&Y applied on Visp is less accurate in comparison to its application to the city of Solothurn. This may be due to the difference in the size and distribution of building types in these two cities. The city of Solothurn is more comparable with the cities of Neuchatel and Yverdon-Les-Bains in terms of building types, as opposed to the city of Visp (please see Figure 21). Despite to this, the RF model, trained and tested on the joined datasets of Yverdon-Les-Bains and Neuchatel, applied on two cities with a different distribution of building types, provides a good prediction of building types.

5 DAMAGE MODELS

In the previous chapters, a classification of buildings has been carried out on four cities (i.e., Neuchatel, Yverdon, Solothurn and Visp), by means of visual/physical surveys and by the application of a deep learning model (i.e., RF). This allows us to have several exposure models, which will be subjected to damage assessment.

In order to perform the damage assessment, two elements are needed: the hazard model and the exposure model. The hazard model is the source of hazard (e.g. earthquakes), while the exposure model is the set of buildings exposed to the hazard. In our study, several damage assessments are carried out: first of all, the damage assessments of the real exposure models of Neuchatel and Yverdon-Les-Bains are realized; after that, the damage assessments of the real and predicted exposure models of Solothurn and Visp are performed. In the latter case, we analyse the comparison between the results of the damage assessment of the real exposure model and those of the predicted exposure model. In addition, a study is done on the Visp exposure model, subjecting it to a magnitude 5, 6 and 7 hazard model, and the differences between the different cases are presented.

5.1 HAZARD MODELS

In our study, for the damage assessment, two different hazard models were used. For the Visp damage assessment, Sierre earthquake of 1946 was considered (Figure 25). It happened in the central Valais region in the southwest of Switzerland and it was the strongest for the last 150 years. A moment magnitude of $M_w = 6.1$ was firstly assigned to the event from the Swiss Earthquake Catalogue (ECOS 2002). After that, a moment magnitude of 5.8 was detected from European stations. The latter value has been used in this study. The event resulted in 3500 damaged buildings and CHF 26 million in today's money in damages (Fritsche & Donat, 2009). For Neuchatel, Yverdon and Solothurn damage assessment, the Basel earthquake of 1356 with magnitude 6.6 was considered. It is known as one of the most damaging events in intra-plate Europe within historical times and it was one of several devastating catastrophes in the 14th century (Fritsche & Donat, 2009).

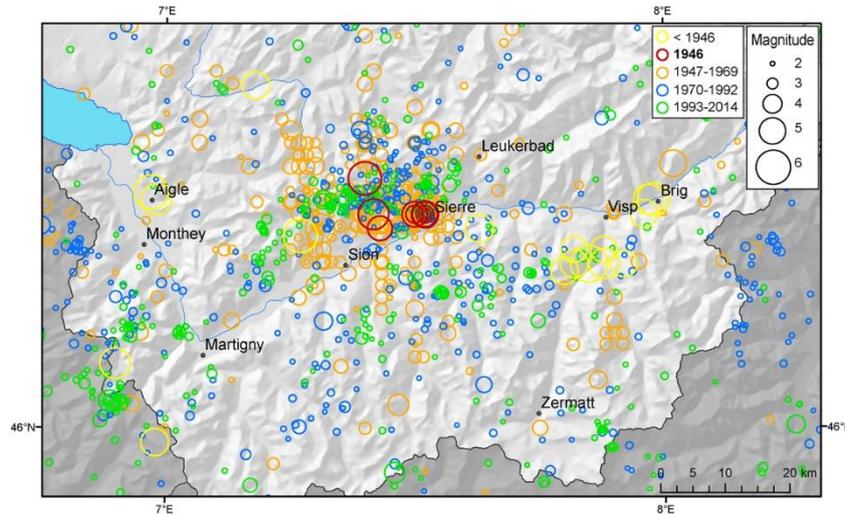


Figure 25 - Siere earthquake (1946)

5.2 EXPOSURE MODELS

For each damage assessment, the exposure model was derived from the dataset of buildings of the considered city. The exposure model quantifies the building stock considering the structural characteristics, the location and the occupancy (Pavić et al., 2020). In details, buildings are classified according to a taxonomy which uniformly covers all the structural typologies (as described in Chapter 2). In this study, the exposure model contains the following information about each building: (1) ID, which corresponds to the EGID and univocally indicates a building; (2) longitude and latitude of the building, which identify its geographical location; (3) taxonomy (e.g. M3_L, M3_M), which indicates the building type and height class, as reported in Table 1 and Table 4; (4) number, which indicates the quantity of buildings (i.e., 1). In our study, we considered a taxonomy that includes ten classes of buildings: five building types (i.e., M3, M4, M5, M6, RCW), subdivided in low-rise (L) and mid-rise (M). Therefore, rural buildings (M1) and reinforced concrete buildings with frame (RCF) were not considered in the damage assessment, since they are only present in a small percentage in Switzerland. High-rise buildings are also excluded from our assessment, since they constitute about 5% of the entire dataset of buildings.

5.3 FRAGILITY CURVES

In order to quantify earthquake loss estimation, it is fundamental to define the relationships between earthquake damage and ground motion. They are also called fragility curves and depict the probability of exceeding a certain level of damage as a function of ground-shaking intensity (Lallemant et al., 2015). Moreover, they can be classified as (1) empirical fragility curves, based on data of post-earthquake damage evaluation; (2) analytical fragility curves, based on response simulation and structural modelling; (3) heuristic fragility curves, based on

opinion of experts. The fragility curves can be represented as both discrete and continuous form (Lallemant et al., 2015). It is important to underline that fragility curves of this study are taken from Earthquake risk model. No coefficients for soil amplification factors are considered.

Although there are already several fragility models for European buildings, direct application of them in a large-scale seismic risk analysis can be challenging as the methodologies and damage criteria considered on available studies are often different. Moreover, structural characteristics of Swiss buildings are different in compared to buildings, common in other European countries; structural modelling of Swiss-specific buildings is the only way to have an estimate of possible damages. It is also worth noting that empirical fragility models developed based on macroseismic intensity are usually suffering from the lack of sufficient datasets of recorded damages due to the past events. Those models are, therefore, usually used for damage assessment of common scenarios whereas mechanical-based fragility functions can be a better option when it comes to calculating seismic risk in the probabilistic framework. The aforementioned factors demonstrate the necessity of a fragility model capable of overcoming these issues. For this study, we used the fragility curves, which are specifically derived for Swiss buildings for the big project of Earthquake risk model (ERM-CH).

5.4 DAMAGE ASSESSMENT OF NEUCHATEL AND YVERDON

In this section, the damage assessment of the real exposure model of Neuchatel and Yverdon-Les-Bains is carried out. First of all, the exposure model of the two cities has been created, with 3300 buildings for Neuchatel and 2401 buildings for Yverdon-Les-Bains. After that, the damage assessment has been realized. The considered hazard model is the Basel earthquake of 1356 with magnitude 6.6.

Before analysing the results of damage assessment, it is interesting to know how many buildings have been considered for each building type. Table 13 shows the distribution of buildings according to the used taxonomy. The majority of buildings considered in Neuchatel is composed by masonry buildings with simple stones – low-rise and mid-rise (i.e., M3_L and M3_M) and unreinforced buildings with rigid floors (i.e., M6_L and M6_M). Moreover, reinforced concrete buildings – mid-rise (i.e., RCW_M) are present with a significant percentage. Considering Yverdon-Les-Bains, three major classes of buildings are present: (1) unreinforced buildings with rigid floors – low-rise (i.e., M6_L), as 27.49% of total buildings; (2) unreinforced buildings with flexible floors – low-rise (i.e., M5_L), as 26.07% of total; (3) masonry buildings with simple stones (i.e., M3_L), as 13.66% of total.

Table 13 - Building types used for damage assessment in Neuchatel and Yverdon-Les-Bains

TAXONOMY	NEUCHATEL		YVERDON-LES-BAINS	
	Number of buildings	% of buildings	Number of buildings	% of buildings
M3_L	582	17.64	328	13.66
M3_M	607	18.39	82	3.42
M4_L	106	3.21	78	3.25
M4_M	188	5.70	57	2.37
M5_L	75	2.27	626	26.07
M5_M	80	2.42	31	1.29
M6_L	511	15.48	660	27.49
M6_M	403	12.21	180	7.50
RCW_L	280	8.48	210	8.75
RCW_M	468	14.18	149	6.21

Figure 26 shows graphically the distribution of the degree of damage, as an average of percentage of damage between the different buildings. The degree of damage is divided into 6 classes (e.g. DG0, DG1 etc.), where DG0 represents the no damage and DG5 represents the fully collapsed building.

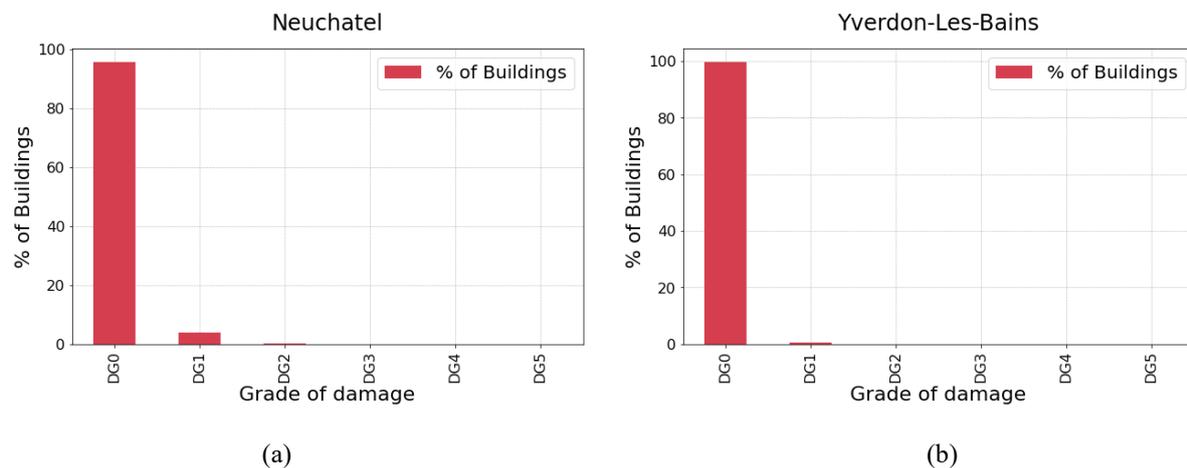


Figure 26 - Distribution of damage grade: (a) Neuchatel; (b) Yverdon-Les-Bains.

The distribution of damage degree for the two cities can be represented with the probability of damage degrees according to the taxonomy of buildings, as shown in Table 14 and in Table 15. The results are also graphically shown in Figure 45 of the Appendix.

Table 14 - Distribution of damage grade in Neuchatel

DAMAGE GRADES						
TAXONOMY	DG0	DG1	DG2	DG3	DG4	DG5
M3_L	98.86	1.06	0.08	0.00	0.00	0.00
M3_M	92.41	6.74	0.61	0.13	0.10	0.01
M4_L	99.29	0.66	0.05	0.00	0.00	0.00
M4_M	94.60	4.89	0.47	0.02	0.02	0.00
M5_L	98.25	1.59	0.16	0.00	0.00	0.00
M5_M	94.06	5.26	0.36	0.08	0.23	0.01
M6_L	98.11	1.82	0.07	0.00	0.00	0.00
M6_M	94.49	5.15	0.34	0.00	0.01	0.00
RCW_L	99.00	0.97	0.03	0.00	0.00	0.00
RCW_M	90.79	8.58	0.62	0.01	0.00	0.00

Table 15 – Distribution of damage grade in Yverdon-Les-Bains

DAMAGE GRADES						
TAXONOMY	DG0	DG1	DG2	DG3	DG4	DG5
M3_L	99.90	0.10	0.00	0.00	0.00	0.00
M3_M	98.18	1.73	0.06	0.01	0.01	0.00
M4_L	100.00	0.00	0.00	0.00	0.00	0.00
M4_M	98.89	1.00	0.09	0.02	0.00	0.00
M5_L	99.83	0.17	0.00	0.00	0.00	0.00
M5_M	98.48	1.45	0.06	0.00	0.00	0.00
M6_L	99.63	0.36	0.01	0.00	0.00	0.00
M6_M	98.81	1.16	0.03	0.01	0.00	0.00
RCW_L	99.90	0.10	0.00	0.00	0.00	0.00
RCW_M	97.85	2.09	0.06	0.00	0.00	0.00

In the end, we can show a comparison between the distribution of damage grade of Neuchatel and Yverdon-Les-Bains in Figure 27. As we can see, the majority of buildings is subjected to a damage grade DG0. Apart from that, a difference can be observed between the distribution of the damage degree of the two cities, as the DG1 in Neuchatel is higher than the DG1 in Yverdon-Les-Bains. This may be due to the fact that Neuchatel is closer to the source of the earthquake (Basel) than Yverdon-Les-Bains. Another reason might be that in Neuchatel there is a higher percentage of buildings from construction periods 1 and 2 than in Yverdon-Les-Bains, as shown in Figure 9. The other damage grades are only present in small percentages

in both cities. For instance, the most severe degree of damage (i.e., DG5) is only present in Neuchatel, in a small percentage (0.0036%).

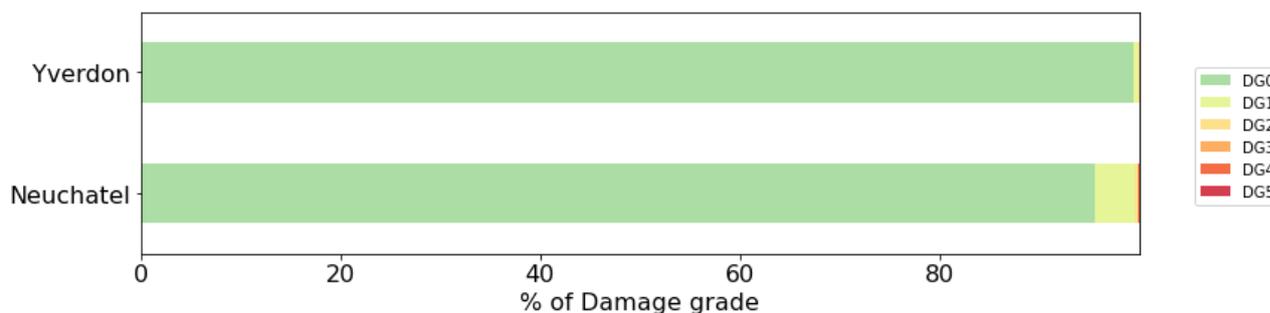


Figure 27 - Distribution of damage grade: comparison between Neuchatel and Yverdon-Les-Bains

5.5 DAMAGE ASSESSMENT OF SOLOTHURN: REAL AND PREDICTED EXPOSURE MODELS

In this section, the damage assessment of Solothurn is carried out. In particular, the exposure models used are both the real ones and those predicted by the RF model. Therefore, two exposure models will be subjected to damage assessment: (1) Solothurn - reality; (2) Solothurn – prediction.

First, we can take a look at the number of buildings in each class of building in the two cases (Table 16). For both Solothurn – reality and Solothurn – prediction, the major classes are: (1) unreinforced buildings with rigid floors – low-rise (i.e., M6_L); (2) masonry buildings with simple stones – low-rise (i.e., M3_L); (3) reinforced concrete buildings - low-rise (i.e., RCW_L); (4) masonry buildings with simple stones – mid-rise (i.e., M3_M). The other building types represent a minority and mainly concern mid-rise buildings, stone buildings (i.e., M4) and unreinforced buildings with flexible floors (i.e., M5).

Table 16 – Building types in Solothurn: reality and prediction

TAXONOMY	REALITY		PREDICTION	
	Number of buildings	% of buildings	Number of buildings	% of buildings
M3_L	640	19.97	663	20.71
M3_M	433	13.51	411	12.84
M4_L	0	0	39	1.22
M4_M	1	0.03	58	1.81
M5_L	263	8.21	269	8.40
M5_M	44	1.37	6	0.19

M6_L	885	27.61	843	26.33
M6_M	174	5.43	189	5.90
RCW_L	555	17.32	527	16.46
RCW_M	210	6.55	197	6.15

Figure 28 shows the distribution of damage grade of Solothurn, for the real and predicted exposure model, as an average of percentage of damage between the different buildings. As it is possible to see, as the grade of damage increases, the percentage of buildings decreases. Moreover, despite slight differences of taxonomy between the actual and predicted model, the damage degree distribution is almost identical. Indeed, for the grade of damage DG0, there is a difference of 0.18% between reality and prediction, whereas for DG5, there is no difference.

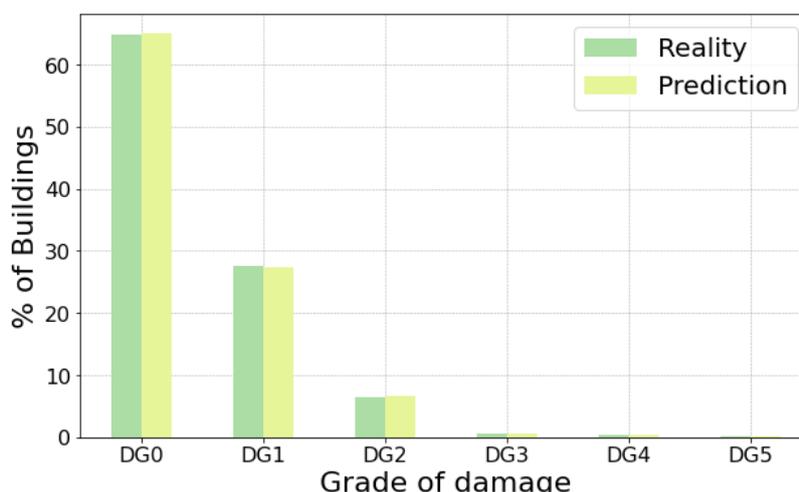


Figure 28 - Distribution of damage grade of Solothurn based on the real and predicted exposure models

The distribution of damage grade can also be represented according to the taxonomy of buildings, as presented in Figure 29. The results are also shown in Table 18 and Table 19 of the Appendix. Figure 30 shows the difference in damage grade, as the difference between the damage evaluated from the real and the predicted exposure model. For all the classes of buildings, the difference decreases as the damage grade increases from the least severe (i.e., DG0) towards the most severe (i.e., DG5). The biggest difference is observed for M4 (mid-rise) and M5 (mid-rise), that are the classes with less elements (please see Table 16); this demonstrates once again that the fewer elements in the sample, the lower the accuracy of the model's prediction. For the other building types, the difference between the actual and predicted damage grade distribution is limited to 1%.

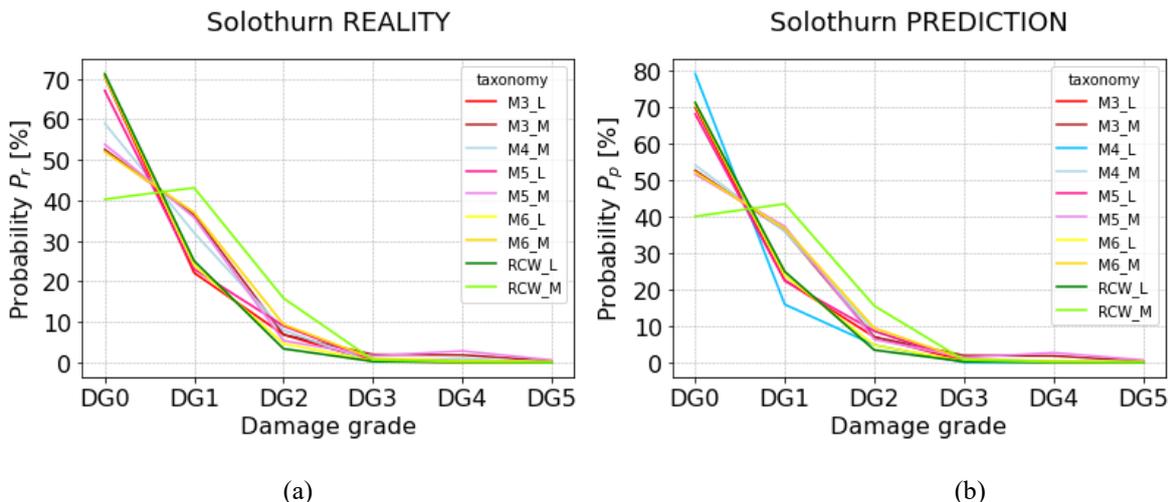


Figure 29 - Distribution of damage grade for different building types in Solothurn: (a) reality; (b) prediction.

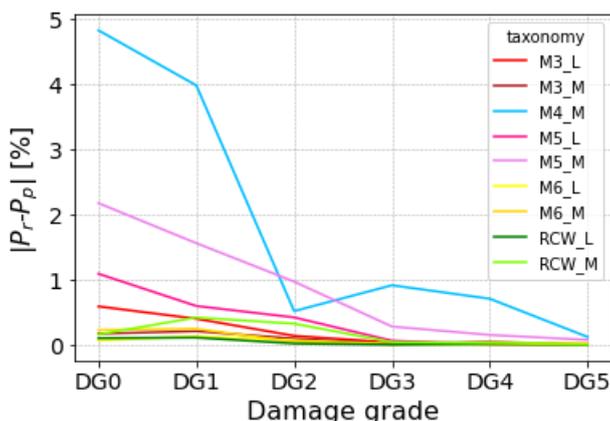


Figure 30 - Differences of the expected damage evaluated from real and predicted exposure model in Solothurn

Finally, we can show a comparison between the distribution of damage grade of Solothurn with the real exposure model and the one with the predicted exposure model in Figure 31. Here we can see that more than half of the buildings have an average probability of damage grade 0; about a third of the buildings have probability of damage grade 1; the remaining percentage of buildings have higher damage grades. The difference between the actual and predicted distribution of damage grade is very small.

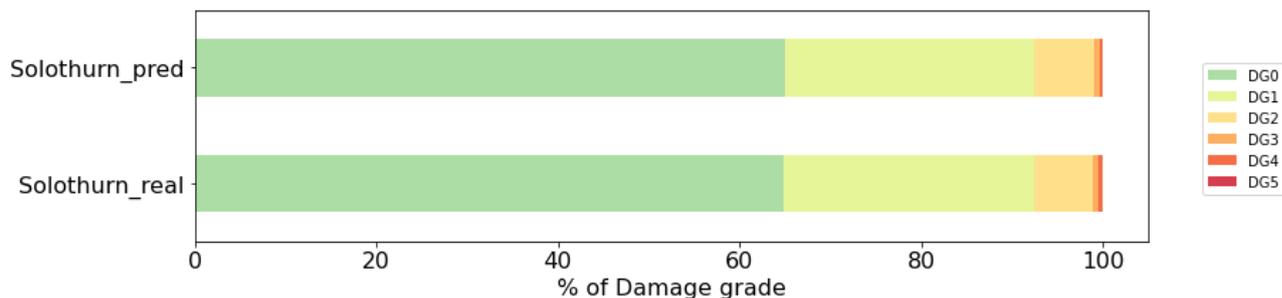


Figure 31 - Distribution of damage grade: comparison between Solothurn reality and Solothurn prediction

5.6 DAMAGE ASSESSMENT OF VISP: REAL AND PREDICTED EXPOSURE MODELS

In this section, the damage assessment of Visp is carried out. In particular, the exposure models used are both the real ones and those predicted by the RF model. Therefore, two exposure models will be subjected to damage assessment: (1) Visp - reality; (2) Visp – prediction. First, we can give a look at the number of buildings subjected to damage assessment for each class of building in each of the four cases. Most of the buildings in Visp used for the damage assessment are RC (mid-rise), with 33.57% in the real classification and 40.07% in the predicted classification. After that, there are the M6 (mid-rise and low-rise), with percentages ranging from 16.61% to 20.58%. The remaining buildings are M3, M4, M5. Although, we can see that M4 and M5 are present in a very small amount in the predicted distribution and are not present in the real distribution.

Table 17 - Building types in Visp: reality and prediction

TAXONOMY	REALITY		PREDICTION	
	Number of buildings	% of buildings	Number of buildings	% of buildings
M3_L	30	10.83	24	8.66
M3_M	21	7.58	12	4.33
M4_M	0	0	1	0.36
M5_M	0	0	1	0.36
M6_L	52	18.77	52	18.77
M6_M	57	20.58	46	16.61
RCW_L	24	8.66	30	10.83
RCW_M	93	33.57	111	40.07

Figure 32 shows the distribution of damage grade of Visp, for the real and predicted exposure model, as an average of percentage of damage between the different classes of buildings. As

it is possible to see, despite slight differences in taxonomy between the actual and the predicted model, the damage degree distribution is almost identical. Furthermore, we can see that more than 90% of the buildings fall into the damage category 0; less than 10% are DG1 and only a very small number of buildings fall into the higher damage categories. This could be due to the large amount of reinforced concrete RC and M6 buildings present (please see Table 17).

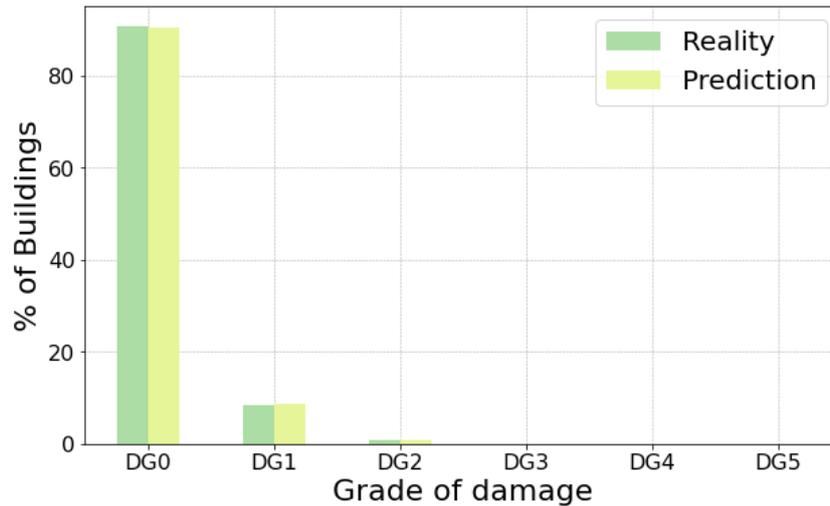


Figure 32 - Distribution of damage grade of Visp based on the real and predicted exposure models

The distribution of damage grade can also be represented according to the taxonomy of buildings, as presented in Figure 33. The results are also shown in

Table 20 and Table 21 of the Appendix. In Figure 34, the difference of distribution of damage grade between the reality and prediction is depicted. We can see that as the severity of the damage increases (from DG0 to DG5), the difference between the two cases - reality and prediction - decreases. The biggest difference is present in the M3 (mid-rise) building type, which is the building type with the fewest elements (please see Table 17). This confirms that the prediction accuracy is higher for building types with a larger number of elements.

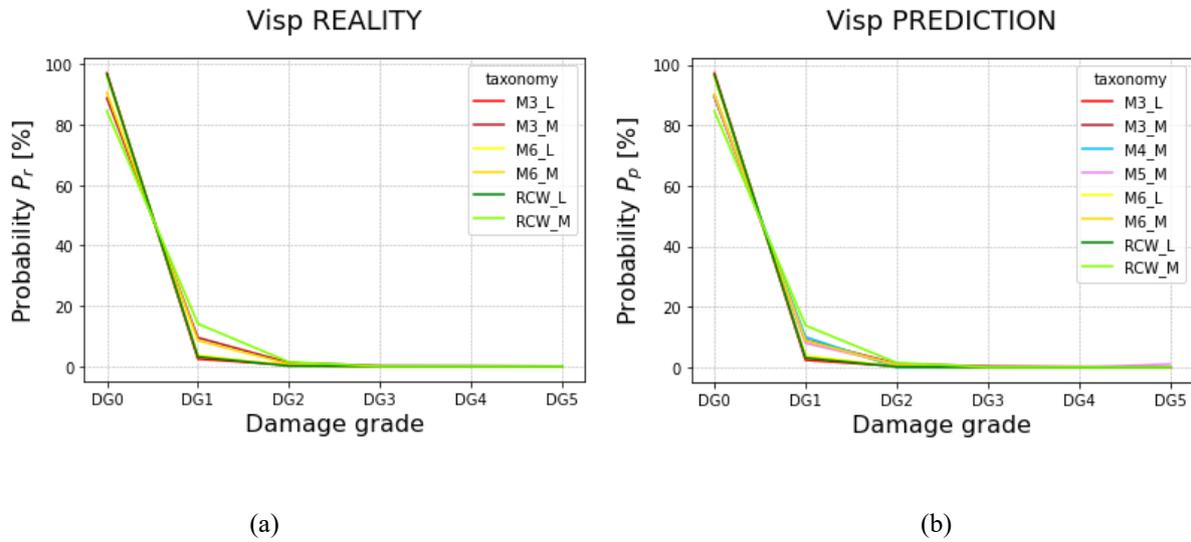


Figure 33 - Distribution of damage grade in Visp resulted from: (a) real exposure model; (b) predicted exposure model

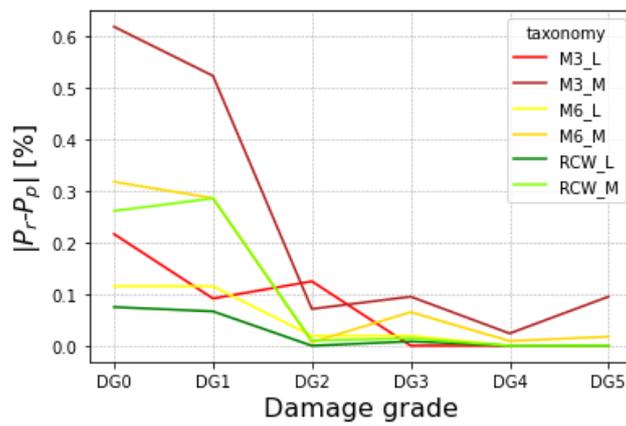


Figure 34 - Differences of damage grade evaluated from real and predicted exposure model for Visp

Finally, we can show a comparison between the distribution of damage grade of Visp with the real exposure model and the one with the predicted exposure model.

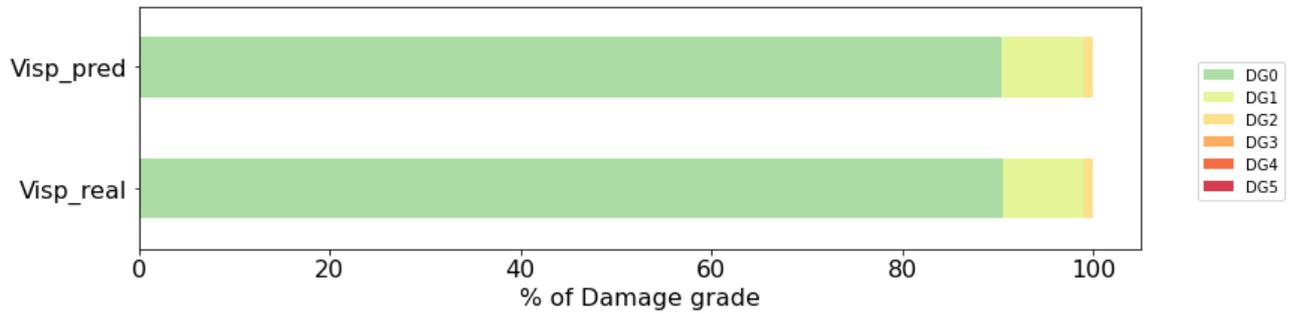


Figure 35 - Distribution of damage grade: comparison between Visp reality and Visp prediction

Speaking of the difference in probability between the real and predicted exposure models, we can also illustrate it in the maps of Figure 36 and Figure 37. Indeed, Figure 36 shows a map of buildings in Visp considering the difference between reality and prediction of DG0, Figure 37 shows the same map, considering the difference between reality and prediction of DG1. A third map, referring to DG2 is included in the Appendix (Figure 43).

Looking at these maps, we can see that the very interesting aspect of the RF model is that it provides predictions building by building. In fact, with the mapping scheme (aggregated data), we can only have a general idea of the percentage of buildings that belong to a certain class and we have no information about the location of the buildings. In the aggregated data, we know the number of buildings for each class, but we do not have precise information, and this could be a problem especially for small villages. On the other hand, with the RF model, we have great accuracy at the building level.

With these maps, although we have some discrepancies in the RF model, we can show that the level of damage is well predicted. The maximum difference between reality and prediction is in the order of 7% only for some buildings, depicted in purple. For all buildings coloured in green, the difference is minimal, on the order of 0-2%. However, for most buildings a good estimate of the damage level was achieved by the exposure model predicted by the RF model.

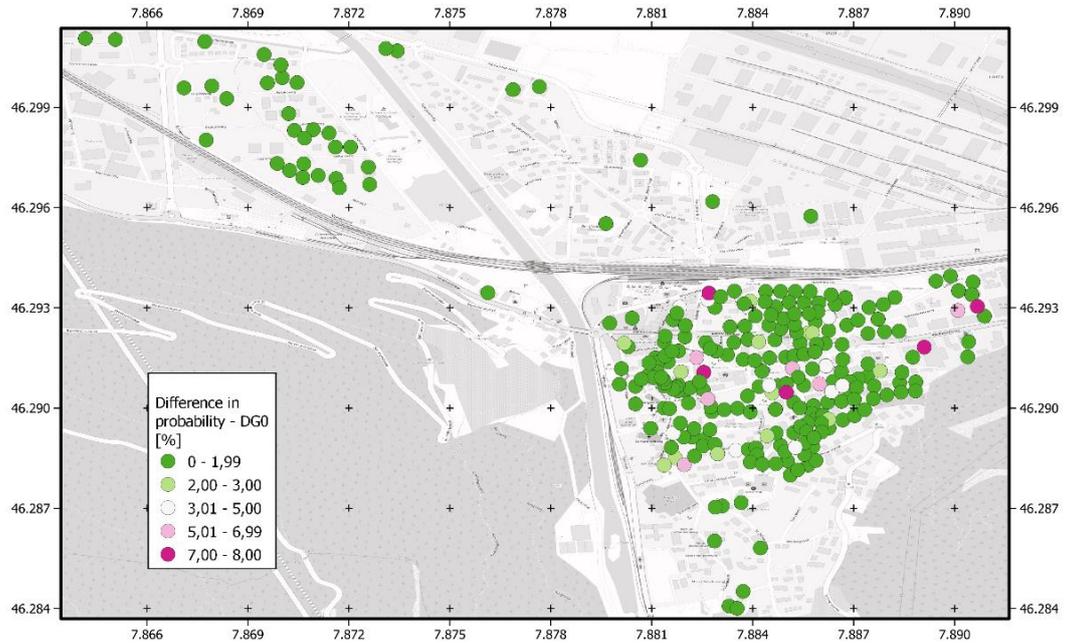


Figure 36 – Difference in probability between reality and prediction – DG0

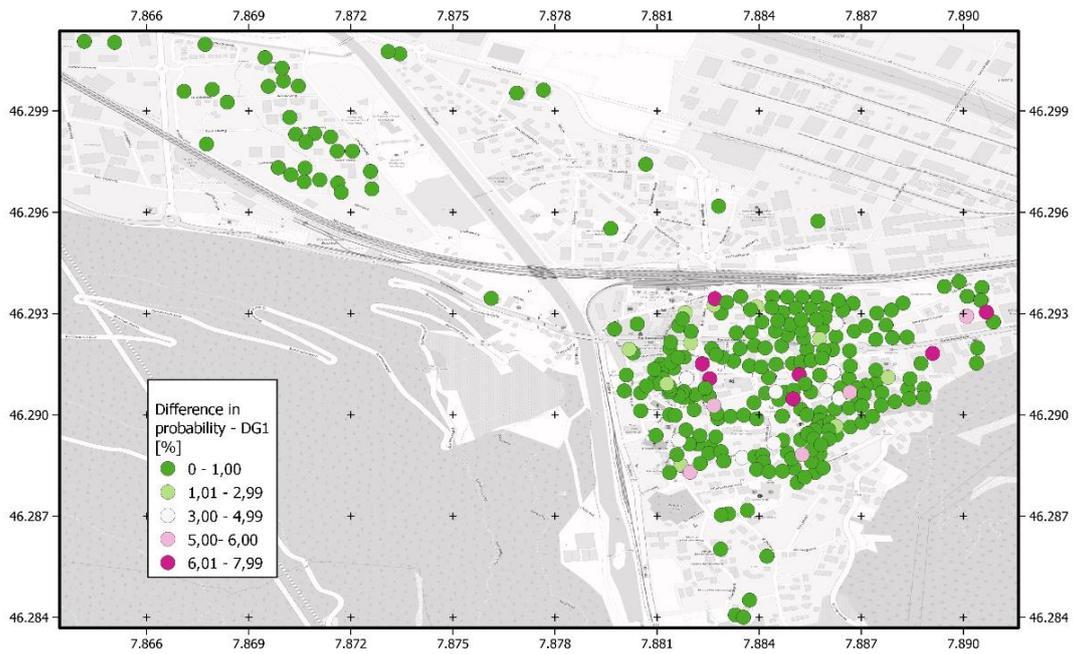
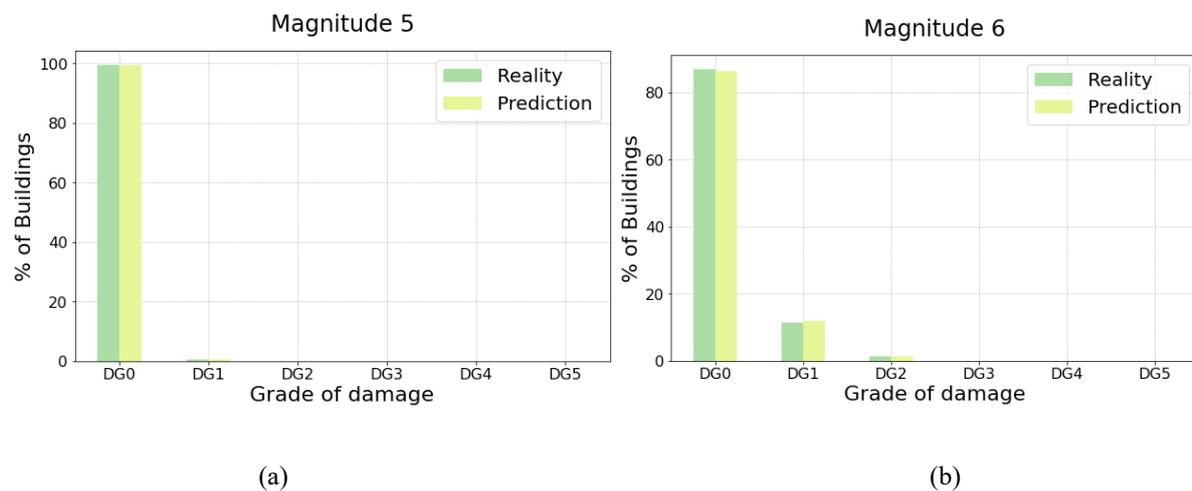


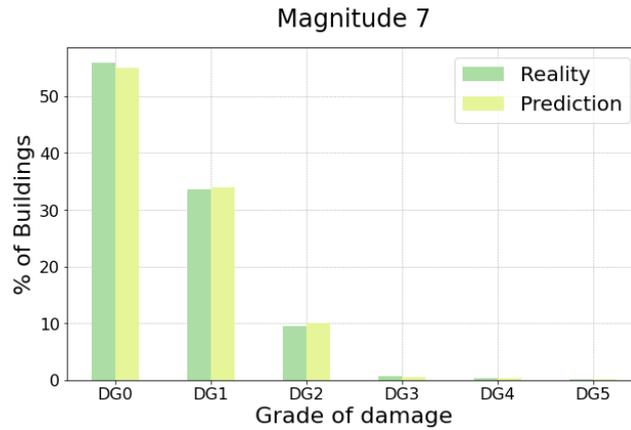
Figure 37 – Difference in probability between reality and prediction – DG1

5.7 DAMAGE ASSESSMENT IN VISP: COMPARISON BETWEEN HAZARD MODELS OF MAGNITUDE 5, 6 AND 7

In the previous section, a damage assessment on Visp has been carried out, by considering a comparison between the real and predicted exposure model, considering the earthquake of Sierre of 1946 with magnitude 5.8 (please see section 5.1). In this section, another study has been done, on the Visp real and predicted exposure models, subjecting it to magnitudes 5, 6 and 7 hazard model. The differences between the three cases will be presented.

The real distribution of building types in Visp for the current damage assessment only includes M3, M6 and RC buildings (please see Table 17), whereas the predicted exposure model also includes few elements of M4 and M5. Figure 38 shows the distribution of damage grade in the three cases – Magnitude 5, Magnitude 6, Magnitude 7. It is clear that as the magnitude increases, also the percentages of buildings in higher grade of damages increase. In fact, for Magnitude 5, grade of damages after DG0 are almost absent, while for Magnitude 6 and Magnitude 7, they increase in percentage of buildings. Furthermore, we can see that as the magnitude increases, the difference between the damage grade distribution with the real and predicted exposure model increases. Indeed, for Magnitude 5, the difference of DG0 between reality and prediction is 0.04%; for Magnitude 6, this difference is 0.42%; finally, for Magnitude 7, the difference is 0.81%.

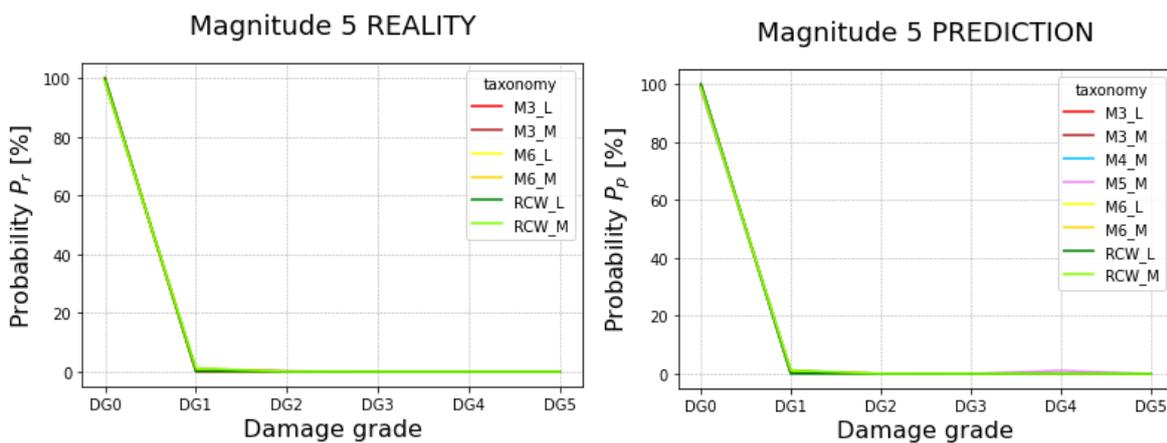




(c)

Figure 38 – Distribution of grade of damage: (a) Magnitude 5; (b) Magnitude 6; (c) Magnitude 7.

Figure 39 shows the comparison of distribution of damage grades evaluated the real and predicted exposure models, in the three cases – Magnitude 5, Magnitude 6 and Magnitude 7. First of all, we can see that the more the magnitude increases, the more the difference in the distribution of the damage grade between the different classes of buildings increases; it is clear that as the severity of the hazard model increases, each class of building reacts differently. Furthermore, we can see that although the magnitude increases, the distribution of the damage grade for the actual and predicted exposure models are still similar to each other, which again proves the validity of the RF model.



(a)

(b)

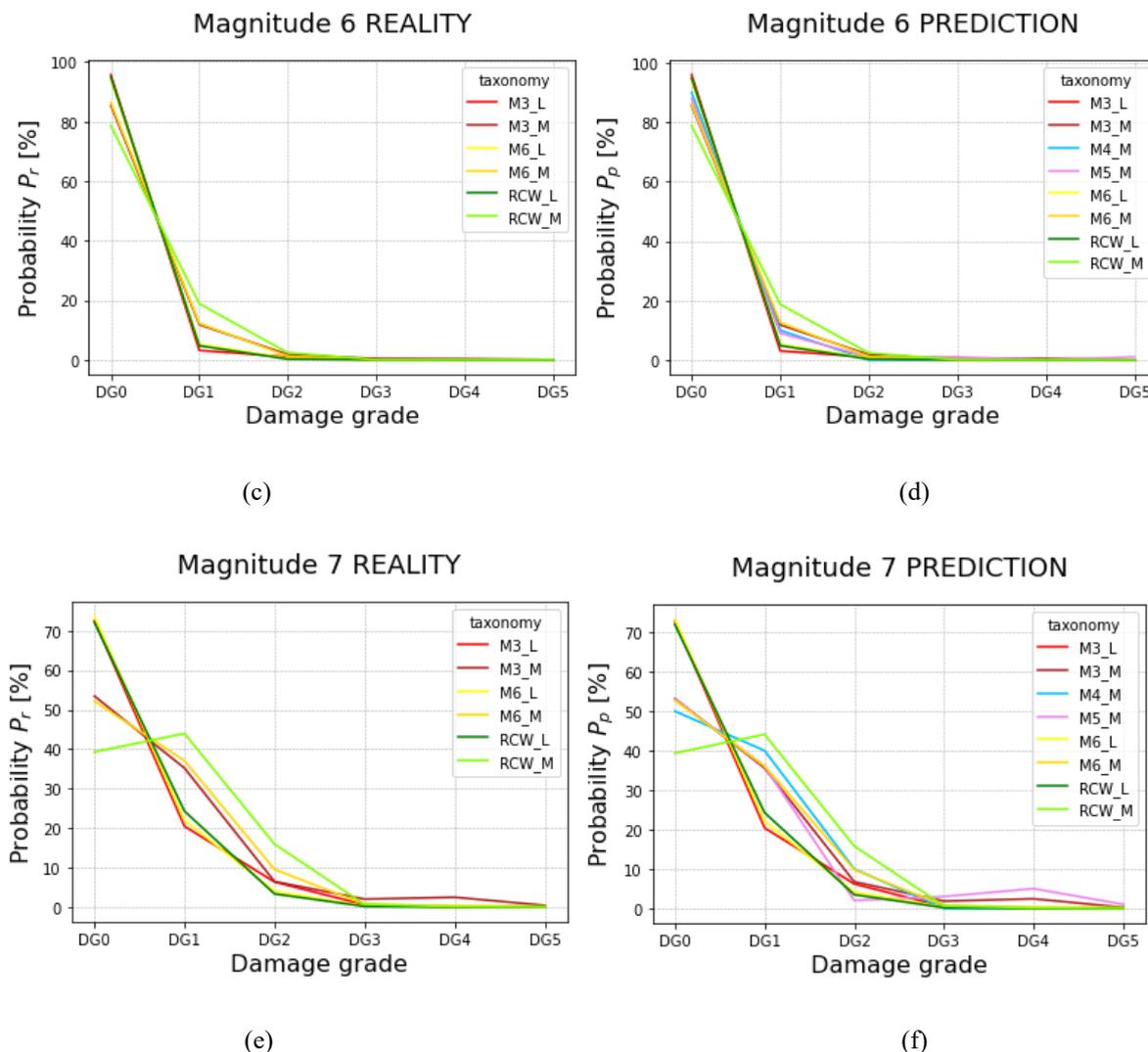


Figure 39 – Distribution of damage grade: (a) Magnitude 5 – reality; (b) Magnitude 5 – prediction; (c) Magnitude 6 – reality; (d) Magnitude 6 – prediction; (e) Magnitude 7 – reality; (f) Magnitude 7 – prediction.

Figure 40 shows the differences in damage grades for the three cases – Magnitude 5, Magnitude 6 and Magnitude 7. Although there is no net change between the cases, it can be seen that as the damage grade increases, the difference between the building types decreases. It is important to stress, however, that the difference represented is an absolute difference. This means that it is not a given that if the difference decreases, then accuracy is better. Indeed, when the damage grade increases, the percentages of buildings are much smaller, so the difference is also smaller. Also, as the magnitude increases, the difference between reality and prediction increases. In fact, we can see that while for Magnitude 5, the maximum difference between reality and prediction is about 0.14%, for Magnitude 7 it is about 1.0%. In this regard, we stress that in Magnitude 5, there seems to be a lot of instability, while in Magnitude 7 the

difference seems to change gradually; actually, the range of difference in the two cases is very different and this justifies the apparent instability of the graph in Figure 40(a).

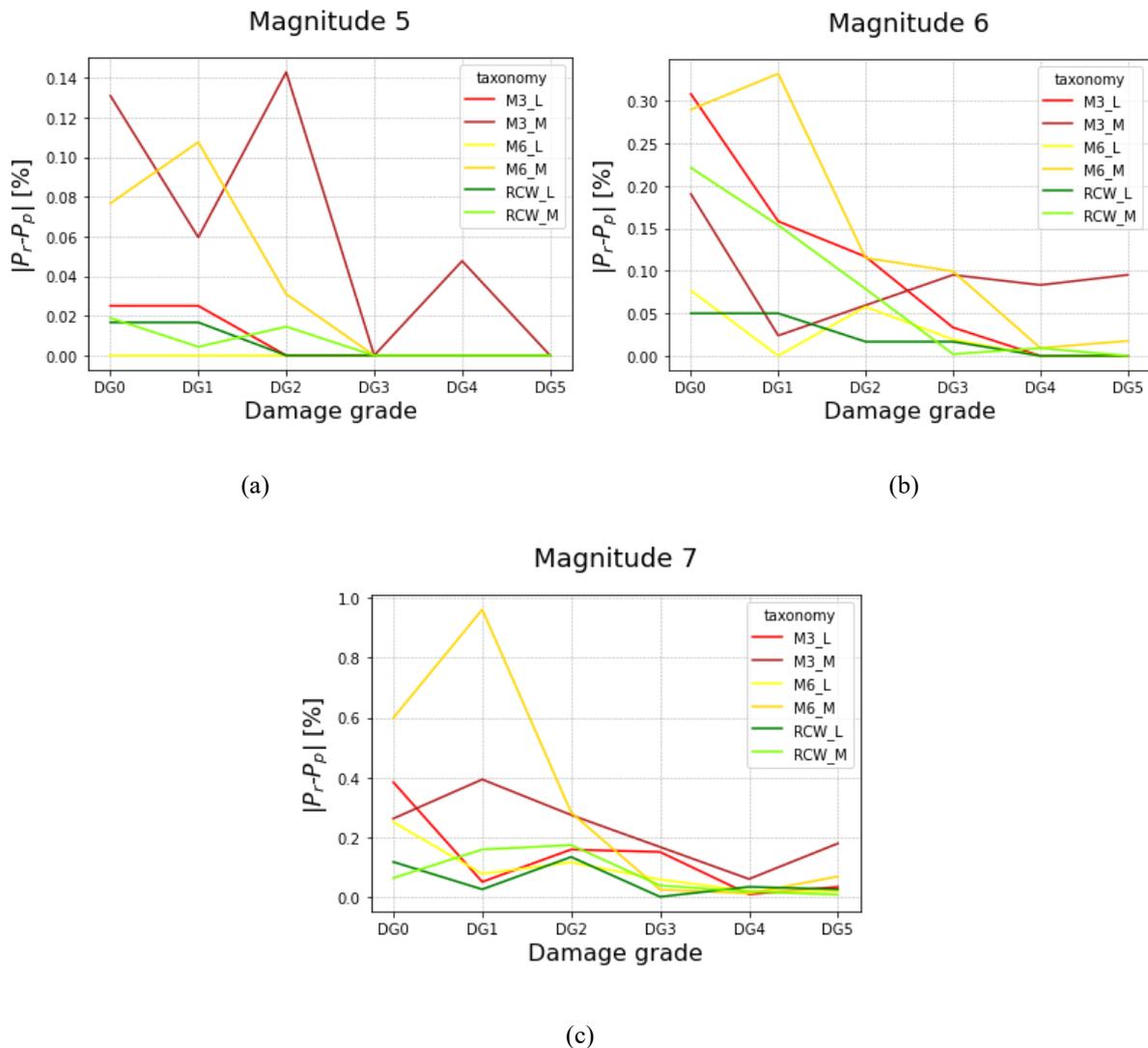


Figure 40 – Difference of damage grades in Visp: (a) Magnitude 5; (b) Magnitude 6; (c) Magnitude 7.

5.8 DAMAGE ASSESSMENT IN VISP: PROBABILITY OF DAMAGE GRADE FOR DIFFERENT CLASSES OF DAMAGE

In this section, we present the map of Visp with the buildings subdivided according to the taxonomy (Figure 41), the map illustrating the percentage of damage grade DG0 of individual buildings (Figure 42) and the map illustrating the percentage of damage grade DG1 of

individual buildings (Figure 43). In the Appendix, we also attach the map illustrating the percentage of damage grade DG2 of individual buildings (Figure 47).

Looking at the maps, we can see for example that M3 - low-rise are close to 1 as a probability of DG0, while they have a lower probability of DG1. On the contrary, M3 - mid-rise have a probability around 0.85 for DG0 and a higher probability (e.g. 0.15) for DG1. A parallelism can also be seen with M6 and RCW: low-rise buildings have a higher probability of DG0 and a lower probability of DG1, while mid-rise buildings follow the opposite law. This phenomenon can be clearly seen for the North-West buildings, where both RCW - low-rise and RCW - mid-rise are present. RCW - low-rise have a probability close to 1 for DG0, while they have a probability of about 0.05 for DG1. RCW - mid-rise have a probability of about 0.80 for DG0, while they have a probability of about 0.20 for DG1.

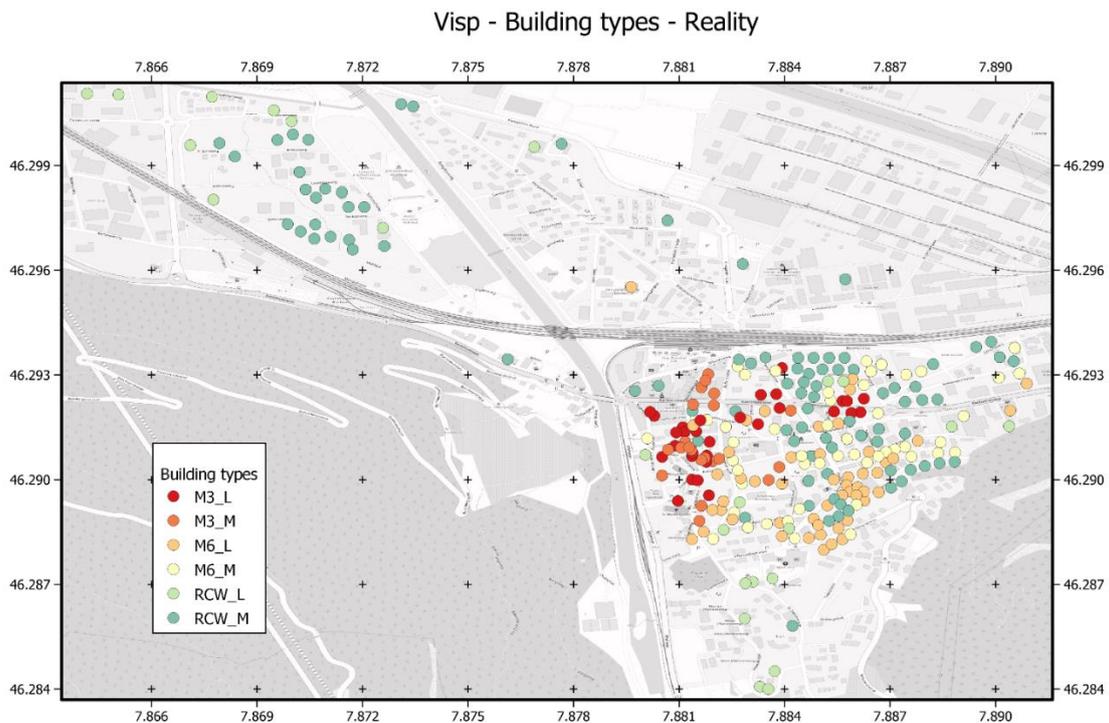


Figure 41 - Visp - Distribution of building types

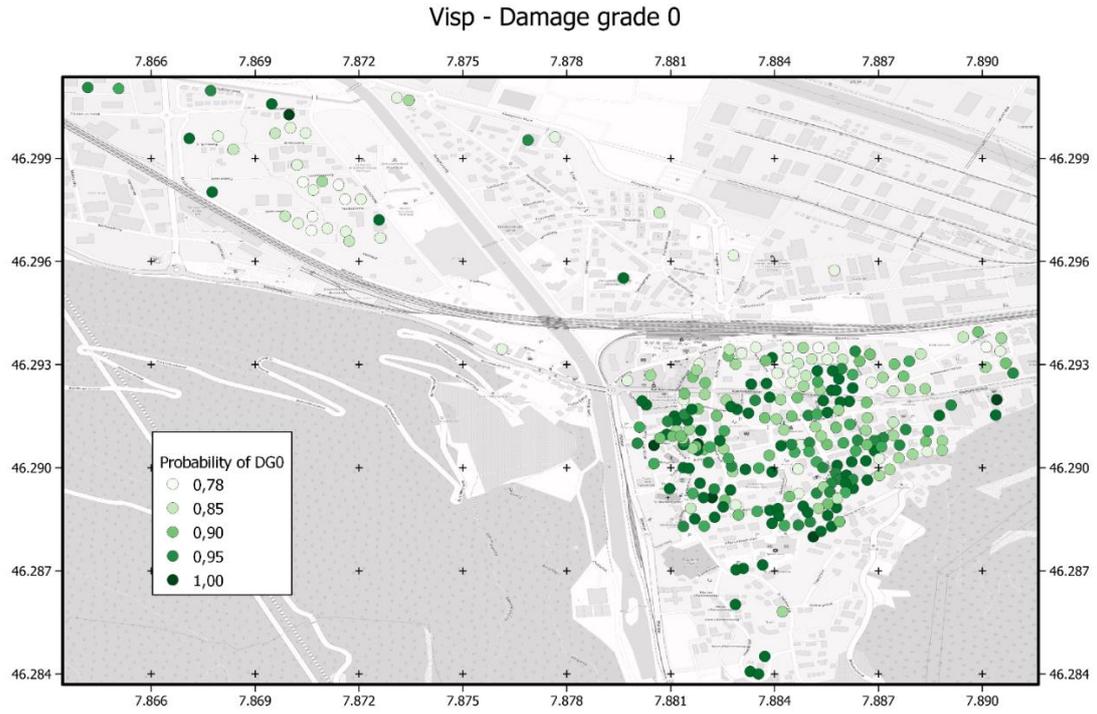


Figure 42 - Visp - DG0

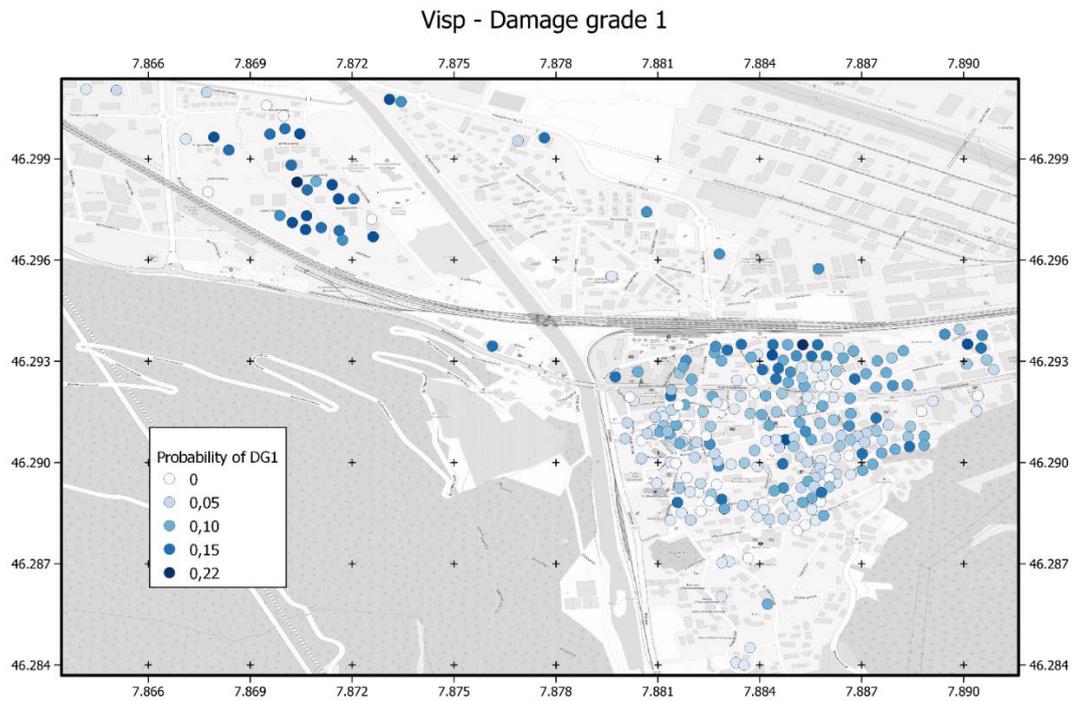


Figure 43 - Visp - DG1



6 CONCLUSION

Natural disasters have always represented a danger to human beings, and among these are earthquakes. Recent examples in Europe include the L'Aquila earthquake (6th of April 2009), which caused more than 300 deaths, 1600 injuries and 10 billion euros of damage. In order to prevent the tragic consequences of natural disasters, it is important to assess the risk to which cities are exposed and to plan, if necessary, retrofitting measures. In this respect, several risk assessment studies are carried out every year, using both traditional and advanced methods (e.g. remote sensing, machine learning etc.). Assessing seismic risk consists of evaluating existing buildings and their expected response in the event of an earthquake. To do this, it is very important to define the exposure model of the area to be assessed. Defining the exposure model consists in classifying buildings according to a building taxonomy (e.g. masonry structures, reinforced concrete structures etc.), based on their material characteristics and construction process. The most traditional method is a visual survey, where buildings are examined one by one and assigned a building type. However, this method is very time and labour intensive, as it has to be carried out by a team of experts. Therefore, in recent years, various techniques have been developed to speed up and automate the work. For example, in the field of Data mining, a computational process used to discover patterns within large datasets through a combination of machine learning, statistics, and dataset systems, we find Association rule learning (ARL) and support vector machine (SVM). Both methods reveal interesting relationships among variables within a large dataset and facilitate the prediction of building types. Furthermore, remote sensing, that is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance has been used for derivation of building vulnerability related features, such as shape characteristics, height, roof material, period of construction, structure type, and spatial context. In addition, LiDAR points, orthophotos and satellite images have been proposed as easily usable and reproducible means to create an exposure and earthquake vulnerability dataset. Despite the development of these innovative methods, it is important to remember that the visual survey is essential for having a reliable exposure model to validate the results of the automated methods.

In our research, we combined the traditional visual survey with the implementation of a deep learning model for building type identification. First, two Swiss cities were investigated with a visual survey: 3537 buildings in Neuchatel and 2808 buildings in Yverdon-Les-Bains. The main characteristics of the buildings were considered, such as the roof shape, the façade aspect, the presence of balconies, and each building has been assigned to a building type. This step has been very time consuming and was carried out according to our engineering judgment. The taxonomy used refers to Lagomarsino et al. (2006). Then, using the buildings' dataset of the Federal Office of Buildings and Logistic of Switzerland, the dataset was enriched with more than 18 building characteristics. An evaluation of the importance of the features and their correlation for the classification of the buildings was carried out and the results of the survey

were analysed according to the most important features (e.g. period of construction, number of floors, number of housing units). As a first outcome of survey, the mapping schemes of the cities have been obtained, with a building type classification according to period of constructions and height classes.

After that, Random Forest (RF), a supervised learning algorithm that can be used for both classification and regression, was applied. RF has been used in the areas of medicine, economics, engineering, for different purposes (e.g. extraction of geometrical features of buildings, classification of images...). In our study, we implemented this method for getting a building type for each building, by exploiting the building attributes and applying the RF for classification. This algorithm is an ensemble of random decision Tree classifiers, that makes predictions by combining the predictions of the individual trees. It considers the buildings' features for making a prediction on the classification of buildings. First, the method was trained and tested on the cities of Neuchatel and Yverdon-Les-Bains. The classification results were evaluated with three measures of accuracy: AM1, based on the confusion matrix (also known as the error matrix); AM2 is calculated by weighting the value of precision of each category on the number of buildings of the respective category. Finally, a third accuracy AM3, that is based on the number of buildings for different types, was considered. Discrete accuracies have been reported for both cities.

Then, a new RF model was created by training and testing on the concatenated datasets of Neuchatel and Yverdon-Les-Bains. In this model, a better performance was observed, compared to the models trained and tested on the individual datasets, as an increase in data provides an improvement in accuracy. The latter model was applied to two other Swiss cities: 3238 buildings in Visp and 307 buildings in Visp, which were also visually surveyed by other means (e.g. Google street and QGIS). This made it possible to obtain both a real and a predicted exposure model. Different accuracies were obtained in these two cities (in Visp the model performed better than in Visp) and this is probably due to the fact that the two cities are different: Visp is more similar to Neuchatel and Yverdon-Les-Bains, while Visp is smaller and has a different distribution of building types. Nevertheless, it is possible to say that the RF model, trained and tested on the joint datasets of Yverdon-Les-Bains and Neuchatel, applied on two cities with a different distribution of building types, provided a good prediction of building types.

As a final step, the seismic damage assessment was performed considering the real and predicted building exposure models to study the detection performance of the proposed building type in the final seismic risk assessment. For the damage assessment, two different hazard models were used. For the Visp damage assessment, Sierre earthquake of 1946 with magnitude 5.8 was considered. It resulted in 3500 damaged buildings and CHF 26 million in today's money in damages. For Visp, the Basel earthquake of 1356 with magnitude 6.6 was considered. A negligible discrepancy was reported between the damage assessment of the real and predicted exposure models. This demonstrates the robustness of the method and paves the way for its application to other cities. After that, another study has been done, on the Visp real and predicted exposure models, subjecting it to magnitudes 5, 6 and 7 hazard model. It is clear that as the magnitude increases, also the percentages of buildings in higher grade of damages

increase. Furthermore, it resulted that as the magnitude increases, the difference between the damage grade distribution with the real and predicted exposure model increases.

In conclusion, we can say that a new model for detecting building types has been developed, where all characteristics of buildings have been considered and the output obtained is at the level of the individual building, not at the aggregate level. Furthermore, it could be extended according to the functionality of the buildings (commercial or residential). Quite good accuracy in detecting building types and higher accuracy in assessing damage were found.



7 FUTURE WORKS

In the present work, a deep learning model for the classification of buildings has been developed and applied on more than 6000 buildings. After that, a damage assessment has been carried out on the real and predicted exposure model. A negligible discrepancy reported between the damage assessment of the real and predicted exposure models demonstrated the robustness of the prediction method and paves the way for its application to other cities.

In the future, we can apply this model to other cities and we can extend it by investigating different types of cities (e.g. big cities, rural areas). Also, since in the previous model we considered all building types, we could focus on only one category (e.g. commercial buildings).

Speaking of machine learning techniques, it would be interesting to develop building classification models that use not the RF method (RF), but other algorithms (e.g. Neural networks, SVM). In fact, machine learning techniques have been developing more and more in recent years and include a variety of algorithms and applications.

Another possibility is to develop models that connect images with building datasets. Indeed, automating the process of recognising and classifying buildings could speed up the creation of the exposure model and make this method applicable to many cities.



8 BIBLIOGRAPHY

- Agrawal, R., Imieliński, T., & Swami, A. (1993). *Mining association rules between sets of items in large databases*. 207–216. <https://doi.org/10.1145/170035.170072>
- Bosch, A., Zisserman, A., & Muñoz, X. (2007). Image Classification using Random Forests and Ferns. *Undefined*. <https://doi.org/10.1109/ICCV.2007.4409066>
- Brando, G., Matteis, G. De, & Spacone, E. (2017). Predictive model for the seismic vulnerability assessment of small historic centres : Application to the inner Abruzzi Region in Italy. *Engineering Structures*, 153(December 2016), 81–96. <https://doi.org/10.1016/j.engstruct.2017.10.013>
- Breiman, L. (2001). Random Forests. *Machine Learning 2001* 45:1, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- C. Cortes, V. V. (1995). Support-Vector Networks. *IEEE Expert-Intelligent Systems and Their Applications*, 7(5), 63–72. <https://doi.org/10.1109/64.163674>
- Diana, L., Thiriot, J., Reuland, Y., & Lestuzzi, P. (2019). Application of association rules to determine building typological classes for seismic damage predictions at regional scale: The case study of basel. *Frontiers in Built Environment*, 5(April), 1–17. <https://doi.org/10.3389/fbuil.2019.00051>
- Donges Niklas. (2021). *Random Forest Algorithms: A Complete Guide | Built In*. <https://builtin.com/data-science/random-forest-algorithm>
- Erdik, M. (2017). Earthquake risk assessment. In *Bulletin of Earthquake Engineering* (Vol. 15, Issue 12). Springer Netherlands. <https://doi.org/10.1007/s10518-017-0235-2>
- Fan, S., Zhang, A., Hu, S., & Sun, W. (2013). A method of classification for airborne full waveform LiDAR data based on random forest. *Zhongguo Jiguang/Chinese Journal of Lasers*, 40(9). <https://doi.org/10.3788/CJL201340.0914001>
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge Discovery in Databases: An Overview. *AI Magazine*, 13(3), 57–57. <https://doi.org/10.1609/AIMAG.V13I3.1011>
- Fritsche, S., & Donat, F. (2009). *The 1946 magnitude 6 . 1 earthquake in the Valais : site-effects as contributor to the damage*. 102, 423–439. <https://doi.org/10.1007/s00015-009-1340-2>
- Geiß, C., Jilge, M., Lakes, T., & Taubenböck, H. (2016). *Estimation of Seismic Vulnerability Levels of Urban Structures With Multisensor Remote Sensing*. 9(5), 1913–1936.
- Geiß, C., & Taubenböck, H. (2013). Remote sensing contributing to assess earthquake risk: From a literature review towards a roadmap. *Natural Hazards*, 68(1), 7–48. <https://doi.org/10.1007/s11069-012-0322-2>
- Greco, A., Pluchino, A., Barbarossa, L., Calì, I., Martinico, F., & Rapisarda, A. (2018). *Seismic and Energy Renovation for Sustainable Cities A simplified model based on self-*

organized criticality framework for the seismic assessment of urban areas.

- Ham, J. S., Chen, Y., Crawford, M. M., & Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), 492–501. <https://doi.org/10.1109/TGRS.2004.842481>
- Han, J., & Kim, J. (2019). GIS-Based Seismic Vulnerability Mapping and Assessment Using AHP: A Case Study of Gyeongju, Korea. *Korean Journal of Remote Sensing*, 35(2), 217–228. <http://doi.org/10.7780/KJRS.2019.35.2.2>
- Harirchian, E., Lahmer, T., Kumari, V., & Jadhav, K. (2020). *Application of Support Vector Machine Modeling for the Rapid Seismic Hazard Safety Evaluation of Existing Buildings*. <https://doi.org/10.3390/en13133340>
- Ho, T. K. (1995). *Random Decision Forests*. 47, 4–8.
- Ho, T. K. (1998). *The Random Subspace Method for Constructing Decision Forests*. 20(8), 832–844.
- Huang, Y. (2017). *on Integrated LiDAR and High-Resolution Images*. <https://doi.org/10.3390/rs9070679>
- Jackson, J., & Conway, G. R. (2006). Fatal attraction: living with earthquakes, the growth of villages into megacities, and earthquake vulnerability in the modern world. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 364(1845), 1911–1925. <https://doi.org/10.1098/RSTA.2006.1805>
- Josip Atalić, Uroš, M., Novak, M. Š., Demšić, M., & Nastev, M. (2021). *The Mw5.4 Zagreb (Croatia) earthquake of March 22, 2020: impacts and response*.
- Lagomarsino, S., & Giovinazzi, S. (2006). *Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings*. 415–443. <https://doi.org/10.1007/s10518-006-9024-z>
- Lallemant, D., Kiremidjian, A., & Burton, H. (2015). Statistical procedures for developing earthquake damage fragility curves. *Earthquake Engineering & Structural Dynamics*, 44(9), 1373–1389. <https://doi.org/10.1002/EQE.2522>
- Lee, S., Panahi, M., Pourghasemi, H. R., & Shahabi, H. (2019). *applied sciences SEVUCAS : A Novel GIS-Based Machine Learning Software for Seismic Vulnerability Assessment*.
- Linoff, & Berry. (1997). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship ... - Michael J. A. Berry, Gordon S. Linoff - Google Libri*. https://books.google.it/books?hl=it&lr=&id=Ni5nMDO1OfEC&oi=fnd&pg=PR19&ots=va78nqEIKN&sig=YcL4-afbK9m_XGPFv1CxTfNORto&redir_esc=y#v=onepage&q&f=false
- Liu, Y., Li, Z., Wei, B., Li, X., & Fu, B. (2019). *Seismic vulnerability assessment at urban scale using data mining and GIScience technology: application to Urumqi (China)*. *Seismic vulnerability assessment at urban scale using data mining and GIScience technology: application to Urumqi (China)*. January. <https://doi.org/10.1080/19475705.2018.1524400>
- Mumtaz, A. (2020, August 13). *How to Develop a Credit Risk Model and Scorecard | Towards Data Science*. <https://towardsdatascience.com/how-to-develop-a-credit-risk-model-and-scorecard-91335fc01f03>

- Navlani, A. (2018). *Sklearn Random Forest Classifiers in Python - DataCamp*.
<https://www.datacamp.com/community/tutorials/random-forests-classifier-python>
- Noble, W. S. (2006). What is a support vector machine? *NATURE BIOTECHNOLOGY*, 24.
<http://www.nature.com/naturebiotechnology>
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222.
<https://doi.org/10.1080/01431160412331269698>
- Pavić, G., Hadzima-Nyarko, M., Bulajić, B., & Jurković, Ž. (2020). Development of Seismic Vulnerability and Exposure Models—A Case Study of Croatia. *Sustainability 2020, Vol. 12, Page 973, 12(3), 973*. <https://doi.org/10.3390/SU12030973>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
<https://arxiv.org/abs/1201.0490v4>
- Porter, K. A., Kiremidjian, A. S., & LeGrue, J. S. (2001). Assembly-Based Vulnerability of Buildings and Its Use in Performance Evaluation: *Https://Doi.Org/10.1193/1.1586176, 17(2), 291–312*. <https://doi.org/10.1193/1.1586176>
- Riedel, I., Guéguen, P., Dalla Mura, M., Pathier, E., Leduc, T., & Chanussot, J. (2015). *Seismic vulnerability assessment of urban environments in moderate-to-low seismic hazard regions using association rule learning and support vector machine*. 1111–1141.
<https://doi.org/10.1007/s11069-014-1538-0>
- Riedel, I., Gueguen, P., Dunand, F., Cottaz, S., Riedel, I., Gueguen, P., Dunand, F., & Macro-scale, S. C. (2014). *Macro-scale vulnerability assessment of cities using Association Rule Learning To cite this version : HAL Id : hal-01015723*.
- Torres, Y., Juan, J., Gaspar-escribano, J. M., Haghi, A., Martínez-cuevas, S., Benito, B., & Carlos, J. (2019). Int J Appl Earth Obs Geoinformation Integration of LiDAR and multispectral images for rapid exposure and earthquake vulnerability estimation . Application in Lorca , Spain. *Int J Appl Earth Obs Geoinformation*, 81(May), 161–175.
<https://doi.org/10.1016/j.jag.2019.05.015>
- Vens. (2013). *Random Forest*.
https://scholar.google.it/scholar?hl=it&as_sdt=0%2C5&q=random+forest+vens+2013&oq=vens
- Wei, B., Su, G., & Liu, F. (2013). *Public response to earthquake disaster : a case study in Yushu Tibetan Autonomous Prefecture*. 441–458. <https://doi.org/10.1007/s11069-013-0710-2>
- Wu, H., Cheng, Z., Shi, W., & Miao, Z. (2014). *An object-based image analysis for building seismic vulnerability assessment using high-resolution remote sensing imagery*. 151–174. <https://doi.org/10.1007/s11069-013-0905-6>
- Zuccaro, G., & Cacace, F. (2015). Seismic vulnerability assessment based on typological characteristics . The first level procedure “ SAVE .” *Soil Dynamics and Earthquake Engineering*, 69, 262–269. <https://doi.org/10.1016/j.soildyn.2014.11.003>



9 ACKNOWLEDGMENTS

At the end of this thesis, I would like to express my deep gratitude to Politecnico di Torino and Ecole Polytechnique Fédérale de Lausanne, which allowed me to spend two years of my Master's degree course in a stimulating and international environment.

I would like to thank my supervisors Prof. Rosario Ceravolo and Prof. Pierino Lestuzzi, who gave me the opportunity of taking part to this project and who supported me during its planning and development. I thank Dr. Alireza Khodaverdian for his patient guidance, enthusiastic encouragement and useful critiques of this work.

Furthermore, I thank Collegio Einaudi, my second family, where I met brilliant and deserving people, who have always encouraged me to do my best, who have given me so much and to whom I will always be fond.

After that, I would like to thank everyone I met during these five years, who somehow motivated me and who supported me, leading me to see the good side even in the bad moments. It is also thanks to you that I have become the person I am today.

Finally, I thank my family, thanks to whom I have been able to realise at least part of my dreams.



10 APPENDIX

The building-by-building visual survey is performed in the city of Neuchatel, by associating each building to a class of Table 1. In order to provide a better understanding of what has been observed during the visual survey, a flowchart is presented below. It includes the main characteristics of buildings that have been taken into account during the survey and their correlations to the assigned building type. The period of construction allows to establish some clear distinctions between the building types, since it gives an indication on the materials that were used during the era of construction of the building, but this was only observed after the survey. The roof shape, the presence of balconies and the façade aspect are the main features that were taken into account during the survey. Moreover, even the shape and the ratio of the windows may be indicative, since RC structures usually allows to have a massive presence of openings. This does not happen for the other building types.

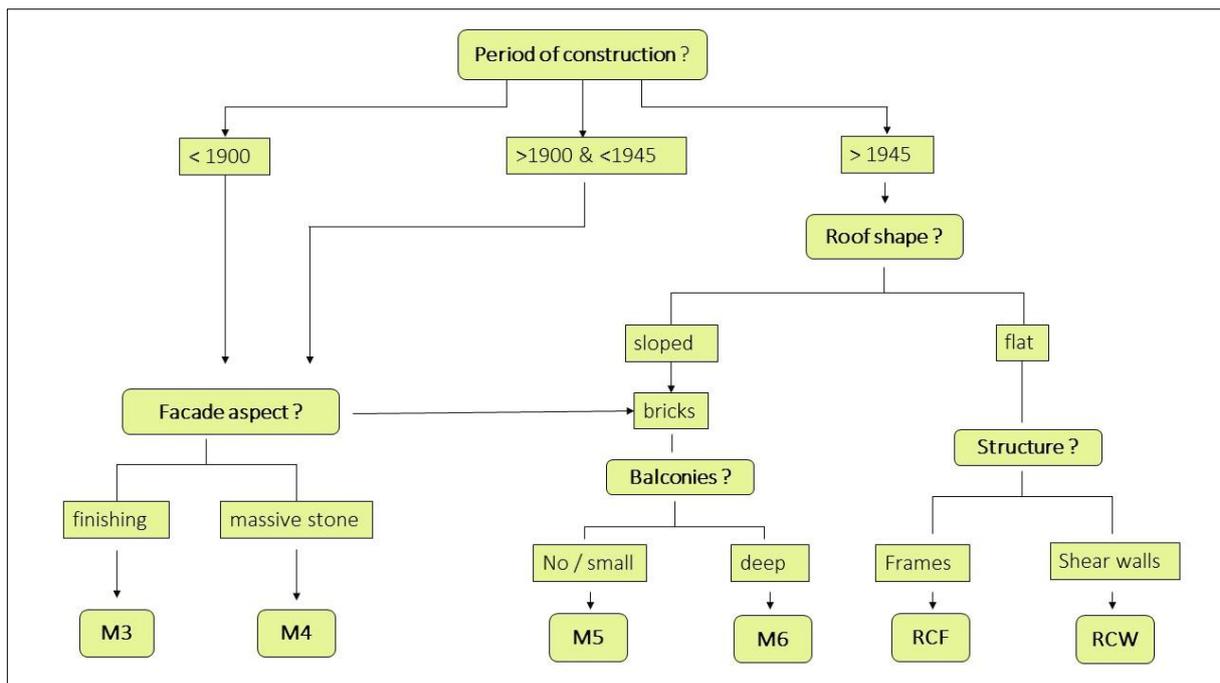


Figure 44 - Flowchart of visual survey based on experience

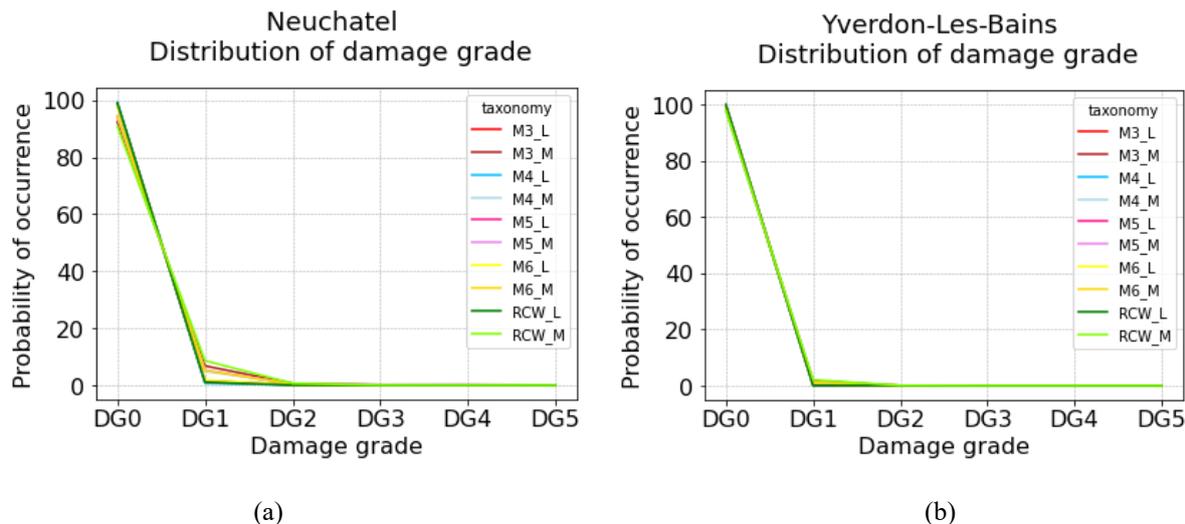


Figure 45 - Distribution of damage grade according to taxonomy of buildings: (a) Neuchatel; (b) Yverdon-Les-Bains.

Table 18 - Solothurn - reality: distribution of damage grade

SOLOTHURN - REALITY						
TAXONOMY	DG0	DG1	DG2	DG3	DG4	DG5
M3_L	70.60	22.13	6.86	0.36	0.04	0.02
M3_M	52.56	36.41	6.87	1.94	1.85	0.37
M4_M	59.00	32.00	8.00	0.00	1.00	0.00
M5_L	67.11	23.05	9.02	0.65	0.12	0.05
M5_M	53.84	35.77	5.36	1.61	2.82	0.59
M6_L	70.77	24.00	4.65	0.36	0.15	0.06
M6_M	52.00	37.03	9.41	0.95	0.39	0.22
RCW_L	71.24	25.06	3.37	0.25	0.06	0.02
RCW_M	40.27	43.12	15.86	0.64	0.10	0.02

Table 19 - Solothurn - prediction: distribution of damage grade

SOLOTHURN - PREDICTION						
TAXONOMY	DG0	DG1	DG2	DG3	DG4	DG5
M3_L	70.01	22.53	6.99	0.39	0.05	0.03
M3_M	52.73	36.20	6.97	1.92	1.81	0.37
M4_L	79.26	15.92	4.79	0.00	0.03	0.00
M4_M	54.17	35.98	8.52	0.91	0.29	0.12
M5_L	68.20	22.45	8.60	0.59	0.11	0.04

M5_M	51.67	37.33	6.33	1.33	2.67	0.67
M6_L	70.84	23.87	4.72	0.36	0.16	0.05
M6_M	52.23	36.79	9.46	0.94	0.35	0.24
RCW_L	71.35	24.95	3.39	0.24	0.05	0.02
RCW_M	40.11	43.54	15.53	0.68	0.13	0.02

Table 20 - Visp - reality: distribution of damage grade

VISP - REALITY						
TAXONOMY	DG0	DG1	DG2	DG3	DG4	DG5
M3_L	97.03	2.47	0.50	0.00	0.00	0.00
M3_M	88.71	9.52	1.24	0.24	0.19	0.10
M6_L	96.00	3.69	0.31	0.00	0.00	0.00
M6_M	90.58	8.56	0.79	0.00	0.05	0.02
RCW_L	96.63	3.17	0.17	0.04	0.00	0.00
RCW_M	84.39	14.10	1.49	0.02	0.00	0.00

Table 21 - Visp - prediction: distribution of damage grade

VISP - PREDICTION						
TAXONOMY	DG0	DG1	DG2	DG3	DG4	DG5
M3_L	97.25	2.38	0.38	0.00	0.00	0.00
M3_M	89.33	9.00	1.17	0.33	0.17	0.00
M4_M	90.00	10.00	0.00	0.00	0.00	0.00
M5_M	90.00	8.00	1.00	0.00	0.00	1.00
M6_L	95.88	3.81	0.29	0.02	0.00	0.00
M6_M	90.26	8.85	0.78	0.07	0.04	0.00
RCW_L	96.70	3.10	0.17	0.03	0.00	0.00
RCW_M	84.65	13.81	1.50	0.04	0.00	0.00

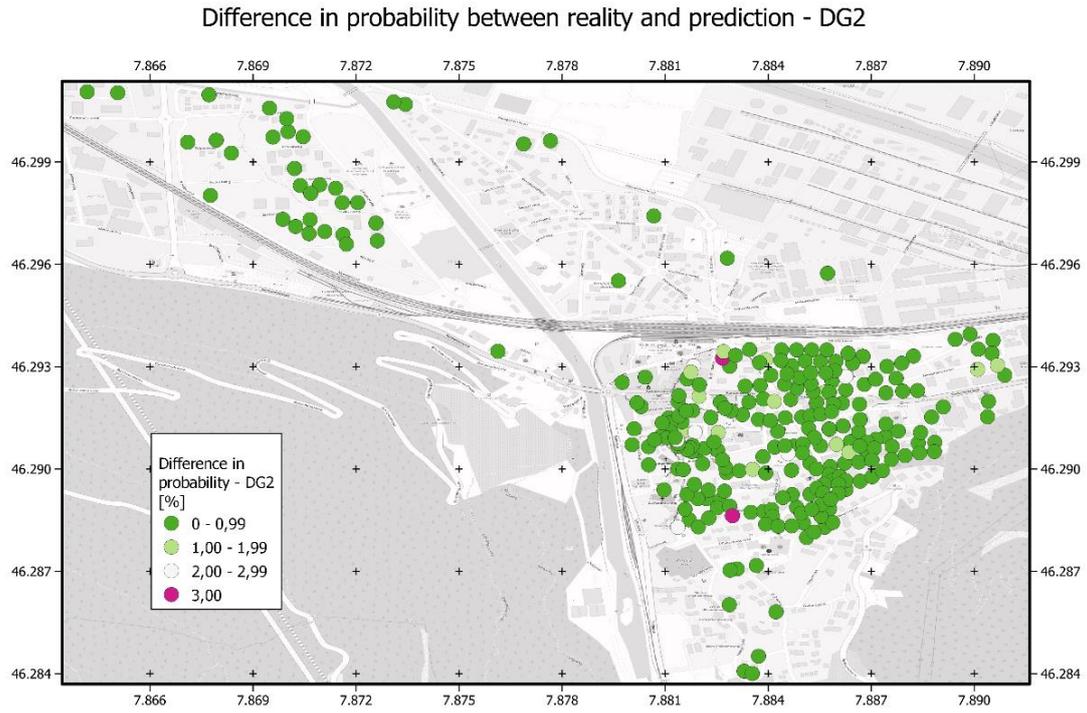


Figure 46 – Difference in probability between reality and prediction – DG2

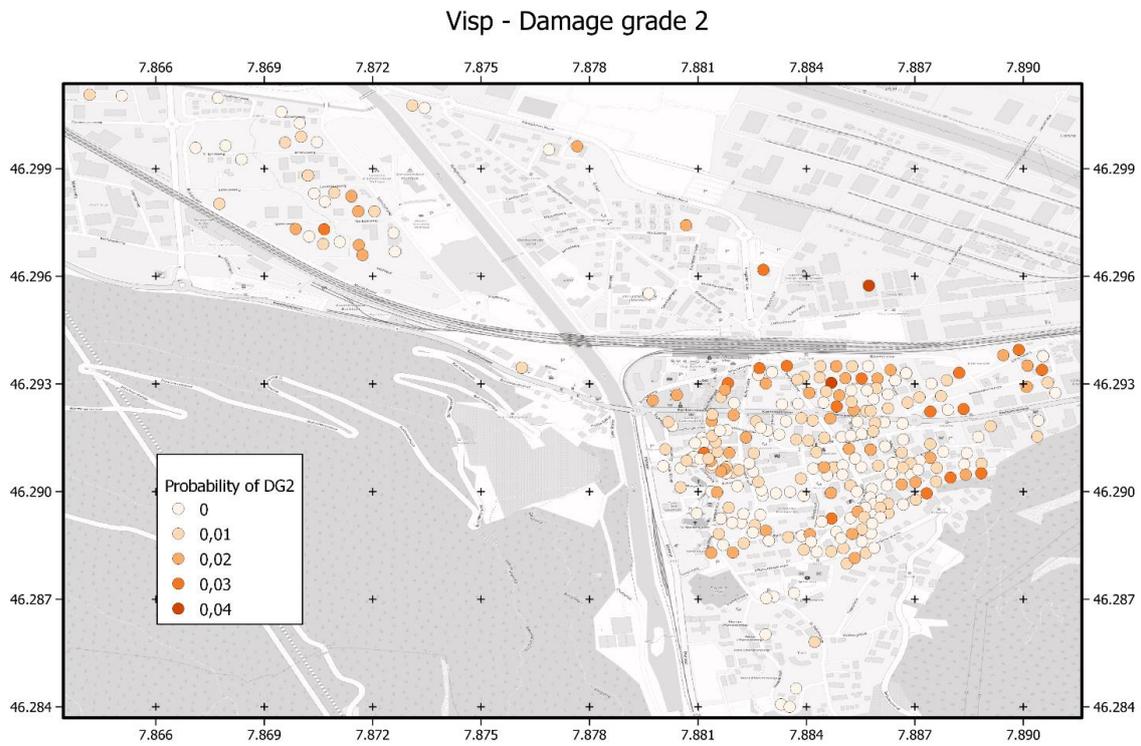


Figure 47 - Visp - DG2