17th World Conference on Earthquake Engineering, 17WCEE
Sendai, Japan - September 27th to October 2nd, 2021

CONCYTEC-WORLD BANK PROJECT: MACHINE LEARNING AND EARTH OBSERVATION TECHNOLOGIES FOR DISASTER MITIGATION

L. Moya[1,2], C. Gonzales[3], M. Diaz[4], B. Adriano[5], E. Mas[6], C. Zavala[7], S. Koshimura[8], and F. Yamazaki[9]

[1] *Principal Investigator, CISMID, National University of Engineering, Peru, lmoyah@uni.pe*
[2] *Visiting Researcher, IRIDeS, Tohoku University, Japan, lmoyah@irides.tohoku.ac.jp*
[3] *Research Associate, CISMID, National University of Engineering, Peru, cgonzalest@uni.edu.pe*
[4] *Associate Professor, CISMID, National University of Engineering, Peru, mdiazf@uni.edu.pe*
[5] *Postdoctoral researcher, Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, bruno.adriano@riken.jp*
[6] *Associate Professor, IRIDeS, Tohoku University, Japan, mas@irides.tohoku.ac.jp*
[7] *Professor, National University of Engineering, czavala@uni.edu.pe*
[8] *Professor, IRIDeS, Tohoku University, Japan, koshimura@irides.tohoku.ac.jp*
[9] *Research Fellow, National Research Institute for Earth Science and Disaster Resilience, Japan, fumio.yamazaki@bosai.go.jp*

Abstract

Currently, there are several types of sensors that are used for disaster science. Accelerometers for ground motion recording, tidal gauges for tsunami monitoring, and satellite images for change detection are such examples. Field surveys might also be considered as a form of earth observation from which essential products like fragility functions have been obtained. This paper summarizes the research framework of the project "Fusion of Machine Learning Algorithms and Earth Observation Technologies for Disaster Mitigation," which aims to merge all the referred sources of information to identify the affected area in the aftermath of a large-scale disaster in a fully automatic mode. For this purpose, in-place sensors that measure the intensity of the disaster in a given coordinate are used to calibrate a disaster's numerical model. The project focuses on earthquakes, and thus, ground motion networks are used to compute the peak ground acceleration (PGA), the peak ground velocity (PGV), and the seismic intensity at the station's coordinates. Then, a numerical interpolation method that considers attenuation laws is used to compute the referred seismic parameters in a uniform spatial grid, referred to as an intensity map. The intensity map is used with fragility functions to obtain prior information of the expected damage at an arbitrary coordinate. Finally, the prior information is used with satellite images to allocate real damaged areas. The fusion of the intensity map, the fragility functions, and the satellite imagery is performed using machine learning techniques. It is worth noting that the machine learning algorithm was adapted to use intensity maps and fragility functions instead of training data, which enables its application in a fully automatic form. The project is conducted within the framework and financial support of the Concytec – World Bank "Improvement and Extension of the Services of the National System of Science, Technology and Technological Innovation (SINACYT)" Project 8682-PE, through its executive unit Fondecyt (contract 038-2019).

Keywords: machine learning, earthquake, disaster mitigation, Concytec-World Bank project No. 8682-PE.

1. Introduction

On February 8, 2017, the Government of the Republic of Peru signed the Loan Agreement No.8682-PE with the World Bank (WB) to fund the Project called: "Improvement and extension of the services of the National System of Science, Technology and Technological Innovation (SINACYT)," aimed at contributing to the economic and competitive diversification of Peru, which shall help to reduce the vulnerability of the productive sector and finally achieve a sustainable development based on knowledge. The National Fund for the Development of Science, Technology, and Technological Innovation - FONDECYT is the Project Implementing Entity that has, in one of its components, the objective of strengthening the applied research and/or technological development capacity through training and attraction of human capital, infrastructure improvement in research, development and innovation (R&D+i) project funding [1].

Research proposals in the following sectors were invited to apply for funding: Farming, energy, telecommunications, health, education, environment, metallurgy, housing and sanitation, agroindustry and food processing, wood forest, textile and apparel, mining and its manufacturing, advanced manufacturing, ecotourism, restoration, and creative industries. A proposal, submitted by the present authors, entitled Fusion of machine learning algorithms and earth observation technologies for disaster mitigation [contract No. 038-2019], was granted as one of the funded projects in the housing and sanitation sector in November 2019. This project aims to research techniques for early disaster response, focusing on, but not limited to, the effects of earthquakes. The research team is composed of the authors of this paper. The project formally started on January 1, 2020, and will continue for two years until December 2021. This paper describes the overall objectives, research plan, and project progress of the SINACYT project number 038-2019, focusing on earthquakes as the primary disaster.

2. Background and objectives of the project

Disasters produced by natural phenomena are common in Peru, and their frequency is increasing in recent years. Floods, debris flow, and landslides occur every year during rainfall periods. Peru is located in the fire belt, and thus, the country is located in an earthquake and tsunamis-prone area. Events such as the 2007 Pisco earthquake, the 1970 Ancash earthquake, where thousands of people suffered personal, property, and material losses, remind us of the importance of increasing society's resiliency against disasters. Therefore, a disaster reduction initiative based on current technology is necessary.

Currently, we live in an era where data is collected and mined in every area of science, and disaster science is not the exception. For instance, the Japanese ground motion networks KiK-net and K-NET [2] provide a massive amount of seismic records to investigate Earth's structure, the physics of earthquakes, and the relationship between round motion and damage. In recent years, new ground motion networks, such as the REDACIS [3], Red Acelerográfica UNI-CIP-SENCICO [4], have been established in Peru for open access. Another important source of information for disaster science is remote sensing. Remote sensing can be defined as acquiring information about an object or phenomenon without making physical contact with that object or phenomenon. Remote sensing products include satellite, aerial, and UAV images. Satellite imagery plays an essential role in disaster mitigation because it is the only technology used to identify damage on a regional scale [5]. The European Space Agency provides open access to the products of his satellite constellations Sentinel-1, Sentinel-2, Sentinel-3, and Sentinel-5P [6], which provides an unprecedented opportunity to implement sustainable remote-sensing-based solutions to certain problems that countries with limited resources face.

In the aftermath of a large-scale disaster, it is crucial to identify damaged buildings as early as possible to facilitate efficient search-and-rescue activities. The need for an early response is related to the life expectancy of occupants trapped under collapsed buildings [7]. It is, therefore, necessary to optimize all steps in the process of building damage mapping. Machine learning algorithms have been used to construct damage maps with high accuracy. However, when machine learning techniques are employed for building damage classification, significant time is spent on retrieving training data. Training data obtained from field surveys require significant time to establish the logistics, conduct the survey, and digitize it. Collecting training data from visual interpretation of optical images are influenced by the person performing the analysis. Official reports from the government provide some of the most reliable information used as training data; however, such reports usually become available only after several weeks or months. To reduce the time required to retrieve training data, Wieland et al. [8] performed a quantitative evaluation of the influence of the number of training samples on the accuracy of damage classification using the support vector machine (SVM) technique. These authors concluded that a minimum number of training samples from a small study area could be used to map building damage in a much wider region. A potential solution to the lack of training data is to transfer training data that have been collected from one disaster event to another disaster event [9,10]; however, this option increases the problem's complexity. Another solution is the combination of remote sensing data with in-place sensors, and numerical simulations proved to be efficient to map damage [11-14], which is the approach this project aims to implement.

3. Overall Research Plan

This project aims to perform comprehensive research on machine learning applications on earth observation technologies to near-real-time damage mapping. We aim to integrate satellite imagery, ground motion sensor networks, and risk analysis concepts to achieve a fully automatic building damage detector. Fig. 1 shows the system's processing chain, which would be applied after an alert is activated. It requires three different inputs. The first input denotes remote sensing data, which should contain post-event and pre-event imagery to identify changes associated with the disaster's effect. The most common approach, termed change detection, aims to identify changes between a pair of images recorded before and after a disaster, from which changed samples are associated with severely damaged buildings and nonchanged samples are associated with non-severely damaged buildings. Note that the whole process will begin once a post-event image is available. The second input is a land-use map to filter out changes due to natural processes. We are interested only in changes in human-made structures, not in changes in, for instance, forest. Note the land use map must be prepared in advance and be ready to use for real-time purposes. The third input is a demand parameter map, which is computed using in-place sensors and numerical simulation. A demand parameter refers to a quantitative measure of the intensity to which a building is subjected. For earthquakes, the peak ground acceleration (PGA), the peak ground velocity (PGV), the Modified Mercalli Intensity, and the spectral response can be used as demand parameters. For tsunamis or floods, the inundation depth has commonly been used as a demand parameter because it can be measured from post-disaster field surveys.

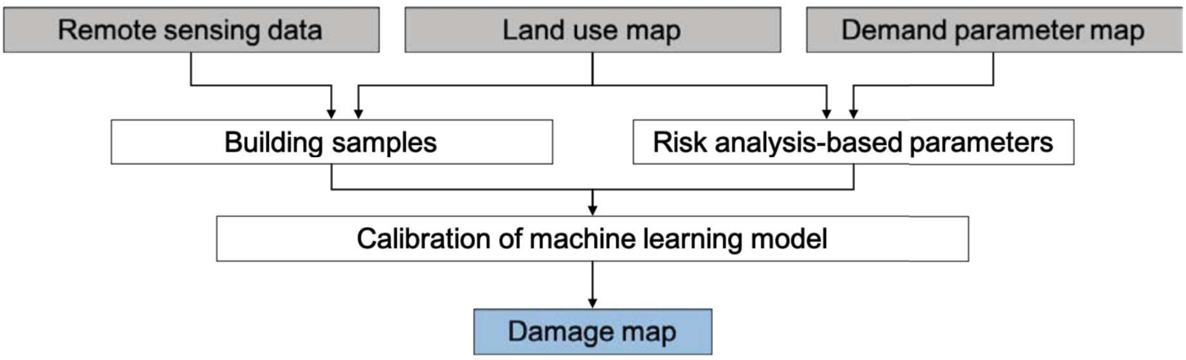

Fig. 1 – Processing chain for damage mapping.

3.1 Remote sensing data

The project is designed, but not limited to, to exploit open-access satellite imagery, such as the constellations from the European Space Agency (ESA) and the Landsat program [15]. Product from Sentinel-1 constellation is of particular importance because it provides synthetic aperture radar (SAR) imagery. Sentinel-1 contains an active sensor, and thus, it operates during day and night. As an active sensor, it has its radiation source, which is designed to obtain cloud-free images. That is, it operates under all weather conditions. Sentinel-1 has different acquisition modes: Stripmap (SM), Interferometric Wide swath (IW), Extra-Wide swath (EW), and Wave (WV). The most commonly available product is IW, which has about 10 m pixel resolution. The sensor can transmit/receive electromagnetic radiation in vertical (V) and horizontal (H) polarizations. Thus, for near-real-time applications, SAR images exhibit fewer constraints than the traditional optical images (See Fig. 2). Under could-free conditions, images from passive sensors provide rich information as well. They can be used not only for damage mapping but also for developing land-use maps, which is the subject of the following subsection. Fig. 3a shows the tile codes of the Sentinel-2 multispectral images, and Fig. 3b and c show some images recorded in 2020.

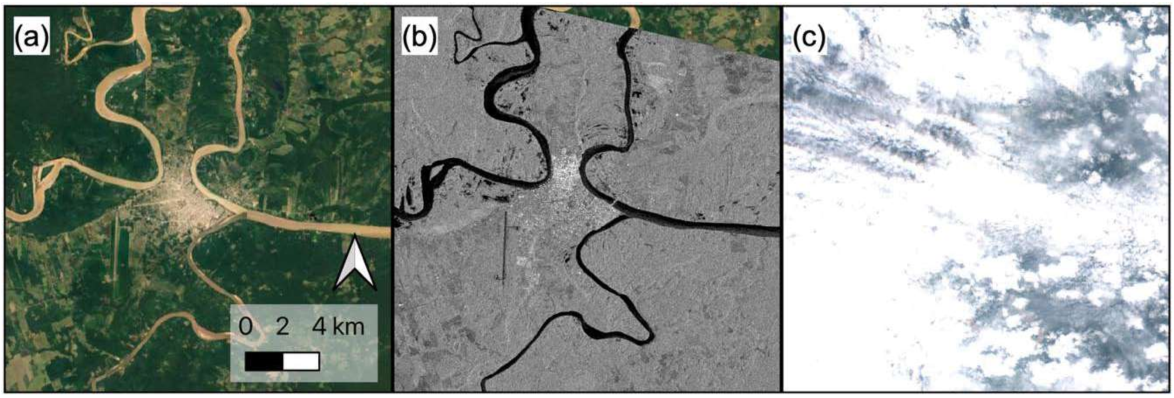

Fig. 2 – Puerto Maldonado city, Madre de Dios, Peru. (a) Image from GoogleMaps. (b) SAR image recorded in February 21, 2021 (c) Optical image recorded in February 20, 2021.

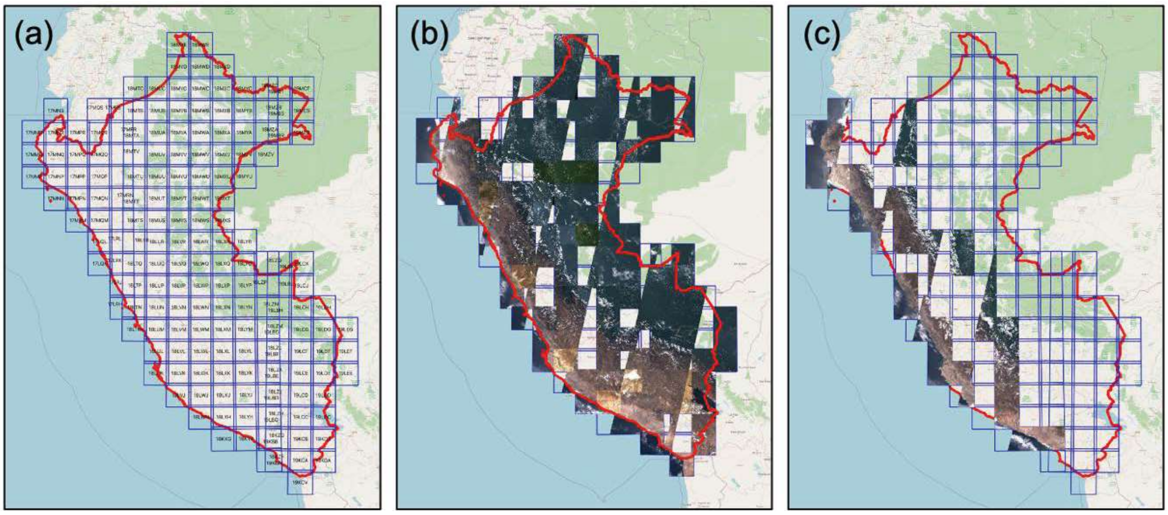

Fig. 3 – Sentinel-2 products in Peru. (a) Grid codes, (b) Images recorded during the year 2020 that are 10%, or less, cloud covered, (c) Images recorded during November 2020 that are 20%, or less, cloud covered.

3.2 Land use map

Supervised machine learning algorithms are employed to identify urban and non-urban areas. For the collection of training samples, geocoded building footprints are retrieved from the OpenStreetMaps (OSM) platform. OpenStreetMap is built by a community of mappers that contribute and maintain data about buildings, roads, trails, cafés, railway stations, and much more worldwide. Building footprints were collected in the following cities of Peru: Lima, Arequipa, Trujillo, Chiclayo, Piura, Huancayo, Cusco, Chimbote, Iquitos, Tacna, Juliaca, Ica, Cajamarca, Pucallpa, Sullana, Ayacucho, Chincha, Huanuco, Huacho, Tarapoto, and Puno (Fig. 4a). Fig. 4b shows a closer look at the building footprints within the city of Lima. As can be observed, like in the other Peruvian cities, the OSM data set is incomplete; however, it is large enough to be used as training samples to calibrate a machine learning classifier and identify all urban areas. Regarding the collection of non-urban areas, we use the National System of Natural Areas Protected by the State's dataset (SINANPE) [16]. SINANPE aims to preserve representative areas of Peru's biological diversity. Fig. 4a depicts the SINANPE's areas as blue polygons. Once urban and non-urban training samples are collected, supervised machine learning algorithms can calibrate a discriminant function. Fig. 5 shows the urban areas identified using the support vector machine algorithm.

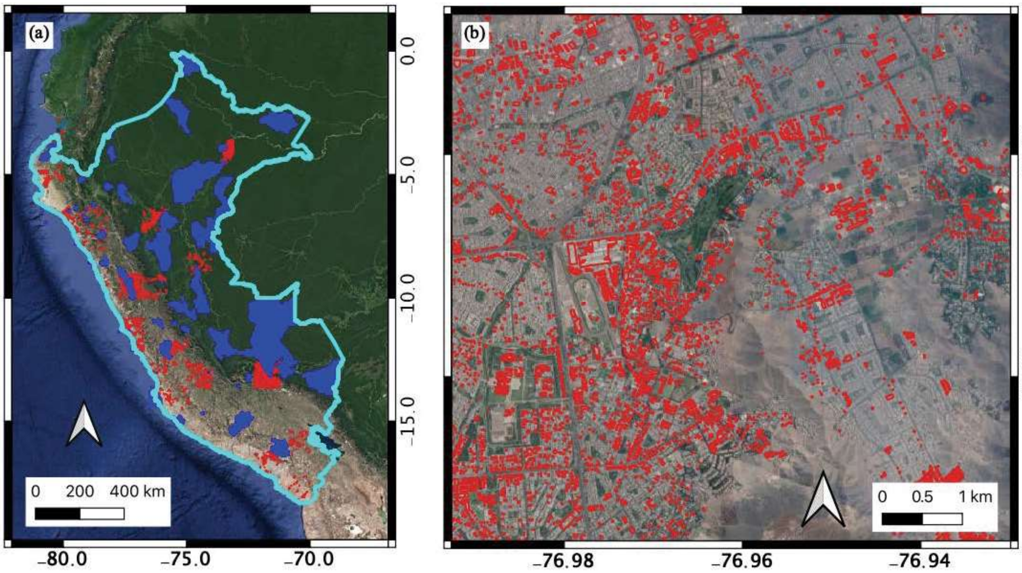

Fig. 4 – (a) Geolocation of training samples. Red polygons denote building units collected from OpenStreetMaps, and the blue polygons are the national protected areas. (b) Closer look of building footprints in Lima, Peru.

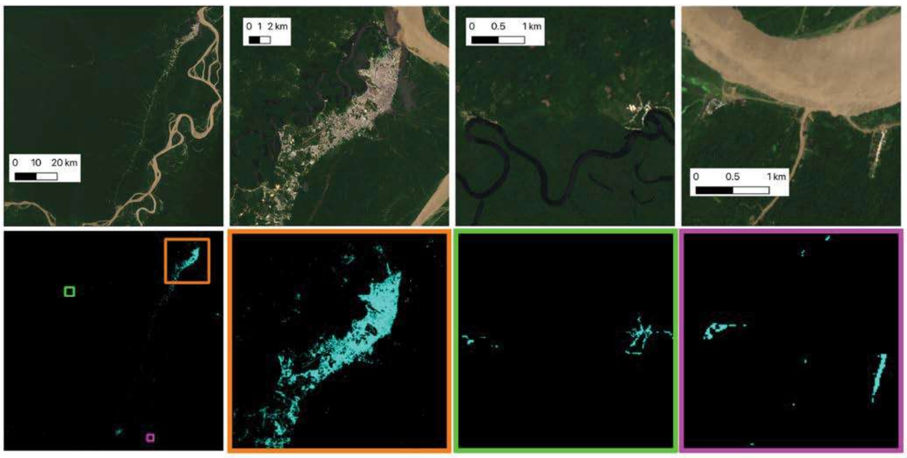

Fig. 5 – Urban areas in Iquitos city, Peru. Top: Sentinel-2 image. Bottom: Urban areas.

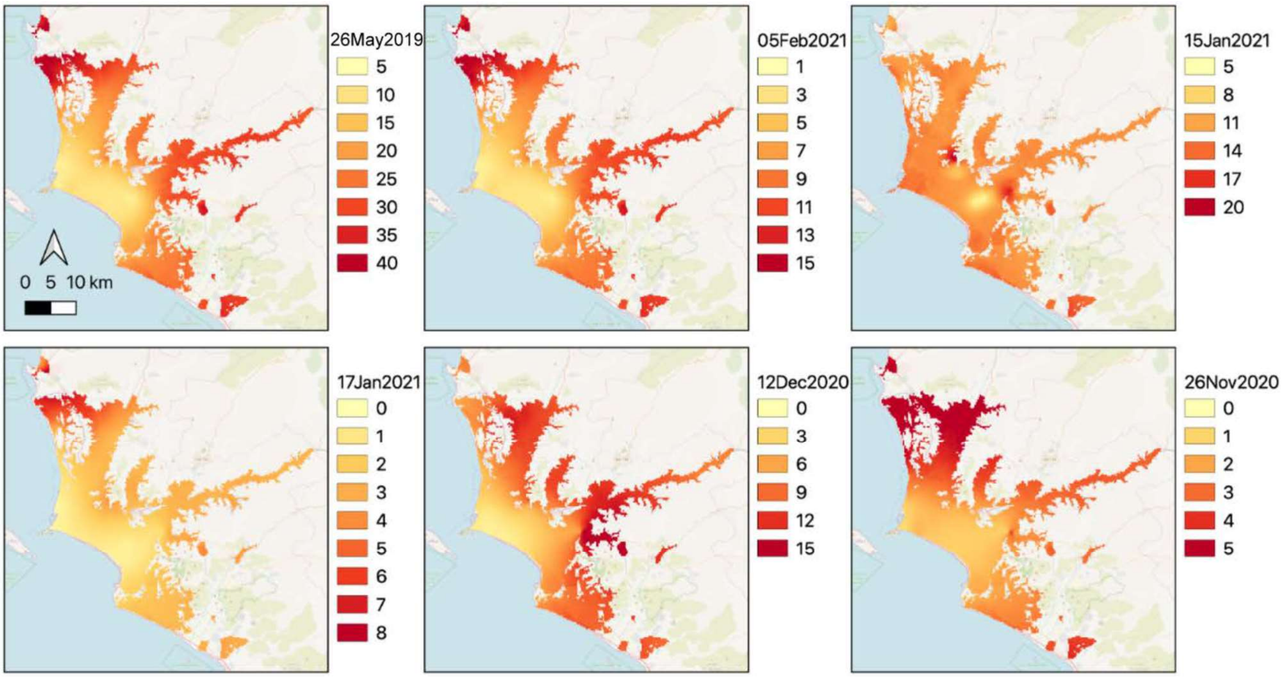

Fig. 6 – PGA maps of recent earthquakes in Lima, Peru. Units are in gals

3.3 Demand parameter map

For the case of earthquakes, the PGA is used as a demand parameter. We have adopted the methodology reported in Matsuoka and Yamamoto [18]. We interpolate the PGA recorded on ground motion stations using the Kriging interpolation method. We use the stations from the REDACIS network in near real-time. Later on, an update of the estimates is performed with additional information from the UNI-CIP-SENCICO network. The PGA is constructed as follows: First, the recorded PGA at the stations were estimated at the engineering bedrock. Then, the Kriging interpolation is performed. Finally, the PGA map at the surface is computed using the amplification factors. The amplification factors for Lima city proposed in [19,20] were used in this project. Fig. 6 shows PGA maps that occurred recently in Lima, Peru. It is worth pointing out that there is a controversy on the accuracy of the method adopted in this project. However, in the following sections, we show that the method for damage mapping requires only an estimation of the order of magnitude of the demand parameter.

As additional activities, the project has scheduled the construction of nine ground motion stations, the acquisition of five low-cost ground motion sensors, and performing in-situ tests to obtain mechanical and dynamic soil properties.

3.4 Risk analysis parameters

Seismic risk analysis refers to the relationship between frequency of exceedance and damage/loss. Two other relationships are necessary: (i) the Probabilistic seismic hazard analysis (PSHA), which express the relation between frequency of exceedance and ground motion, and (ii) Damage/Loss functions, which denotes the relationship between ground motion and damage/loss [21]. Given that our target is to identify actual damage, probabilistic information on exceedance frequency is not necessary. However, damage functions and their relation to ground motion are highly relevant. A fragility function, often idealized by a sigmoid function, is defined as the relationship between the probability that an asset reaches or exceeds a damaged state and the demand it experiences [22]. Fragility functions can be constructed either from field surveys [23], from numerical simulations [24], or a combination of both [25].

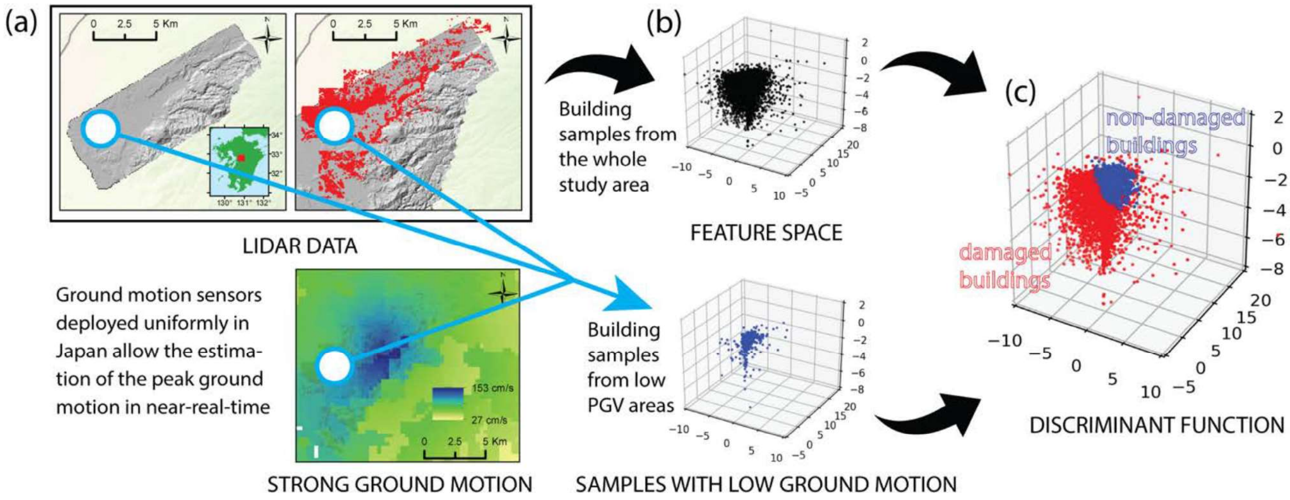

Fig. 7 – Scheme of the automatic samples collection proposed in Moya et al. [13]. (a) Remote sensing data (Lidar data) and strong ground motion map produced during the 2016 Kumamoto earthquake. (b) Feature samples extracted from remote sensing data. Top: whole building samples. Bottom: Samples located at areas with low PGV. (c) Classification of samples in damage (red marks) and non-damaged (blue marks) buildings.

3.5 Calibration of machine learning classifier

Consider a discriminant function whose input denotes a feature vector, and output is a binary response, where 0 denotes a non-damage building, and 1 denotes a damaged building. Such binary response is often referred to as label. Calibration denotes the process of estimating the parameters of the calibration function. In supervised machine learning, the parameters are calibrated using a set of samples whose label is known in advance. The application of supervised machine learning in remote sensing data to construct a damage map has proven efficient in terms of accuracy [5, 8, 27]. There is, however, an important issue regarding the availability of training samples in the aftermath of a disaster. A common solution is to use training samples from previous disasters [9,28]. In this project, we adopt another solution that does not require gathering training data manually or from previous events. It has been shown in [11] and [12] that, instead of training samples, a discriminant function can be calibrated using a demand parameter map and fragility functions. Recently, our project partially funded a study that used only a demand parameter map [13]. The strategy was to collect training samples of non-damaged buildings in areas with low demand parameters. Then, they generalize the discriminant function observing the sample distribution in areas with intermediate/large demand parameters (Fig. 7).

4. Conclusions

A research program called "Improvement and extension of the services of the National System of Science, Technology and Technological Innovation (SINACYT)" started under the agreement between the Peruvian Government and the World Bank. This paper describes the SINACYT project "Fusion of machine learning algorithms and earth observation technologies for disaster mitigation" The project involves multidisciplinary research that includes seismic motion, remote sensing, buildings, and machine learning. The interaction of these fields aims to develop a near-real-time damage mapping system. Currently, the project is at the beginning of the last year. During this period, the project contributed to new damage mapping methods that can work in near-real-time. In order to achieve the objective, sub-systems to create the required inputs are being implemented as well. We have implemented a framework to build a demand parameter map currently working only for the capital city. However, it is expected to be applied to the whole country. Another contribution is a system to automatically collect open-access satellite imagery and training samples of urban and non-urban areas. This information is then used to construct maps of urban areas. The project will continue until the end of this year (2021), and we expect the outputs will contribute to Peru's early disaster response in future disasters.

5. Acknowledgements

This work was supported in part by the Concytec–World Bank research program "Improvement and Extension of the services of the National System of Science, Technology and Technological Innovation" 8682-PE through its executing unit Fondecyt [contract 038-2019].

6. References

[1] Concytec, Fondecyt, and The World Bank (2019) Integrated rules for incorporation of researchers. Lima Peru.

[2] National Research Institute for Earth Science and Disaster Resilience (2019): K-NET, KiK-net https://www.kyoshin.bosai.go.jp/.

[3] Gonzales C, Lazares F, Aguilar Z, Alva J (2021): Development of a University Seismic Network in Metropolitan Lima, Peru. *Seismological Society of America Annual Meeting*. Session: Strong-Motion data Processing and Dissemination: State-of-Art and Outlook.

[4] UNI, CIP, SENCICO (2017): Red Acelerográfica UNI-CIP-SENCICO, Lima, Peru, http://www.red-acelerografica-peru.uni.edu.pe/es/main/home.

[5] Koshimura S, Moya L, Mas E, Bai Y (2020): Tsunami Damage Detection with Remote Sensing: A Review. *Geosciences* 10, no. 5: 177.

[6] European Space Agency (last accessed: 2021): Copernicus Open Access Hub, https://scihub.copernicus.eu/

[7] Ohta Y, Murakami H, Watoh Y, Koyama M (2004): A Model for Evaluating Life Span Characteristics of Entrapped Occupants by an Earthquake. *In Proceedings of the 13th World Conference on Earthquake Engineering*, Vancouver, BC, Canada.

[8] Wieland M, Liu W, Yamazaki F (2016): Learning Change from Synthetic Aperture Radar Images: Performance Evaluation of a Support Vector Machine to Detect Earthquake and Tsunami-Induced Changes. *Remote Sens*. 2016, 8, 792.

[9] Moya L, Mas E, Koshimura S (2020): Learning from the 2018 Western Japan Heavy Rains to Detect Floods during the 2019 Hagibis Typhoon. *Remote Sens*. 12, no. 14: 2244.

[10] Adriano B, Yokoya N, Xia J, Miura H, Liu W, Matsuoka M, Koshimura S (2020): Learning from Multimodal and Multitemporal Earth Observation Data for Building Damage Mapping. *arXiv:2009.06200*.

[11] Moya L, Mas E, Adriano B, Koshimura S, Yamazaki F, Liu W (2018): An integrated method to extract collapsed buildings from satellite imagery, hazard distribution and fragility curves. *International Journal of Disaster Risk Reduction*, Volume 31, Pages 1374-1384.

[12] Moya L, Marval-Perez LR, Mas E, Adriano B, Koshimura S, Yamazaki F (2018): Novel Unsupervised Classification of Collapsed Buildings Using Satellite Imagery, Hazard Scenarios and Fragility Functions. *Remote Sens*. 10, no. 2: 296.

[13] Moya L, Geiß C, Hashimoto M, Mas E, Koshimura S and Strunz G (2021): Disaster Intensity-Based Selection of Training Samples for Remote Sensing Building Damage Classification. *IEEE Transactions on Geoscience and Remote Sensing*, doi: 10.1109/TGRS.2020.3046004.

[14] Moya L, Mas E, Koshimura, S, Yamazaki F, Liu W, Geiß C (2020): Fusion of modeling and remote sensing for damage estimation. *17th World Conference of Earthquake Engineering*, Japan, 9 pages.

[15] Wulder MA et al. (2019): Current status of Landsat program, science, and applications, *Remote Sensing of Environment*, Volume 225, 2019, Pages 127-147.

[16] SINANPE (last accessed in March 2021) https://www.sernanp.gob.pe/el-sinanpe

[17] Maruyama Y, Yamazaki F, Mizuno K, Yogai H, Tsuchiya Y (2008): Development of Fragility Curves for Highway Embankment Based on Damage Data from Recent Earthquakes in Japan. *14th World Conference on Earthquake Engineering*, CD-ROM, Paper No. 01-1058, 8p.

[18] Matsuoka M, Yamamoto N (2012): Web-based quick estimation system of strong ground motion using engineering geomorphologic classification map and observed seismic records. *15World Conference on Earthquake Engineering*, Lisbon, Portugal, 10 pages.

[19] Sekiguchi T, Calderon D, Nakai S, Aguilar Z, and Lazares F (2013): Evaluation of Surface Soil Amplification for Wide Areas in Lima, Peru. Journal of Disaster Research. 8 (2). 259 – 265

[20] Calderon D, Calderon C, Gonzales C (2020): Estimation of Seismic Intensity for a Shake Map Development in Lima, Peru. Proceedings of the 17th World Conference on Earthquake Engineering. 1d-0112.

[21] McGuire (2004): *Seismic hazard and risk analysis*. Earthquake Engineering Research Institute EERI.

[22] Porter K, Kennedy R, Bachman R (2007): Creating Fragility Functions for Performance-Based Earthquake Engineering. *Earthquake Spectra*, 23(2), 471–489.

[23] Yamazaki F, Murao O (2000): *Vulnerability Functions for Japanese Buildings Based on Damage Data from the 1995 Kobe Earthquake. In Implications of Recent Earthquakes on Seismic Risk*. Imperial College Press: London, UK, Volume 2, pp. 91–102.

[24] Villar-Vega M, Silva V, Crowley H, Yepes C, Tarque N, Acevedo AB, Hube MA, Gustavo CD, María HS (2017): Development of a Fragility Model for the Residential Building Stock in South America. Earthquake Spectra, 33(2), 581–604. https://doi.org/10.1193/010716EQS005M

[25] Koshimura S, Oie T, Yanagisawa H, Imamura F (2009): Developing Fragility Functions For Tsunami Damage Estimation Using Numerical Model And Post-Tsunami Data From Banda Aceh, Indonesia. *Coast. Eng. J.* 51, 243–273.

[26] Torisawa K, Matsuoka M, Horie K, Inoguchi M, Yamazaki F (2021): *Study on Building Fragility Curves Based on Uki City's Disaster-Victim Certificate Data due to the 2016 Kumamoto Earthquake*. Journal of Japan Association for Earthquake Engineering, Volume 21, Issue 1, Pages 172-186.

[27] Moya L, Muhari A, Adriano B, Koshimura S, Mas E, Marval-Perez LR, Yokoya N (2020): Detecting urban changes using phase correlation and ℓ1-based sparse model for early disaster response: A case study of the 2018 Sulawesi Indonesia earthquake-tsunami. *Remote Sensing of Environment*, Vol 242, 111743.

[28] Wieland M, Martinis S (2019): A modular processing chain for automated flood monitoring from multi-spectral satellite data. *Remote Sens.*, vol. 11, no. 19, p. 2330.

