Probabilistic inter-scheme compatibility matrices for buildings. An application using existing vulnerability models for earthquakes and tsunami from synthetic datasets constructed using the AeDEs form through expert-based heuristics

(https://doi.org/10.5880/riesgos.2022.003)

Juan Camilo Gómez Zapata[1, 2], Massimiliano Pittore[1, 3]

1. Seismic Hazard and Risk Dynamics, GFZ German Research Centre for Geosciences, Potsdam, Germany
2. Institute for Geosciences, University of Potsdam, Potsdam, Germany
3. Institute for Earth Observation, EURAC Research, Viale Druso 1, Bolzano, 39100, Italy

1. Licence

2. Citation

When using the data please cite:

Gomez-Zapata, Juan Camilo; Pittore, Massimiliano (2022): Probabilistic inter-scheme compatibility matrices for buildings. An application using existing vulnerability models for earthquakes and tsunami from synthetic datasets constructed using the AeDEs form through expert-based heuristics. GFZ Data Services. https://doi.org/10.5880/riesgos.2022.003

The data are supplementary material to:

Gomez-Zapata, J.C., Pittore, M., Brinckmann, N., Cotton, F., Lizarazo-Marriaga J., Medina S., Tarque, N. 2022. Scenario-based multi-risk assessment from existing single-hazard vulnerability models. An application to consecutive earthquakes and tsunamis in Lima, Peru (Under review).

Table of contents

3. Data Description

This data publication contains the scripts, input and output files required to calculate the inter-scheme conversion matrices for building types and the implicit damage states of their respective fragility models for two selected vulnerability schemes: one for earthquakes and the other for tsunamis. They were used in previous studies to characterize the residential building stock of Lima. The outcomes generated in this data repository are valuable inputs to then calculate the disaggregated and cumulative damage and losses expected for cascading hazard scenarios.

4. Files Description

4.1. Inter-scheme building conversion matrices

There are two folders:

- "**Exposure_fuzzy_scores**": In the subfolder "*lib_fuzzy_scoring*" there are the python modules implementing the basic properties of fuzzy numbers (file fuzzy.py) and the functions for the fuzzy scoring (file scoring.py) as proposed in the original method of Pittore et al., (2018). In the subfolder "*fuzzy_scored_schemes*" there are two JSON files that represent the two schemes of interest (sets of building classes) that were used to classify the building stock of Lima to two types of hazard-related vulnerabilities: seismic ground shaking (SARA, (Yepes-Estrada et al., 2017)) and tsunami (Medina, 2019). For such a purpose, each building class within the two schemes has been disaggregated into attribute types, and values contained in the GEM v.2.0 taxonomy (Brzev et al. 2013). The fuzzy scores in the JSON file assign the corresponding compatibility between building classes and attribute values. These files also included is a weighting scheme that ranks the importance of the set of attribute types.
- **"Inter-scheme_mapping"**: it contains the script "*Tax_Conv_SARA-MEDINA-Lima.py*" written in Python3. It can be used to calculate inter-scheme compatibility matrices using the method originally proposed in Gómez Zapata et al., (2022, 2021). The output is a CSV file. In this code, the weighting scheme that ranks the importance of the set of attribute types can be modified. Contrary to the aforementioned publications, we do not present any sensitivity analysis on the selection of these weights. This script also provides the option to generate the graphical representation of the inter-scheme building matrix between SARA (source scheme) and Medina (target one) as presented in Figure 6 of the associated paper for which this data repository is a supplement.

4.2. Inter-scheme damage conversion matrices

This folder contains the main Python script used to obtain the outcomes described in the title. It is called: "Prob_inter-scheme_damage_conv_matrices_SARA_Medina_2019.py". Also, there are three subfolders:

- "Generate_DS_OD_comb_heuristics"
- "Prob_likelihoods_OD"
- "Conversion_matrices"

The first folder contains two other subfolders: SARA and Medina_2019. These are the names of the schemes (set of building classes with corresponding fragility functions) that were used to classify the residential building stock of Lima to assess their physical vulnerability to seismic ground motion and tsunami inundation respectively. In every folder there has been provided a Jupyter Notebook (Python programming language) to create a synthetic dataset based on the expected observable damage on four building elements (vertical structure (VS), infills and partitions (IP), roof (RF) and floor (FL)) using the associated scale (0=L to 9=A) proposed by the AeDES form of the Italian Civil Protection (Baggio et al., 2007). These differential scores are assigned to each damage state (defined in their associated fragility model) and to each building type per hazard-dependent vulnerability scheme. These Jupyter Notebooks are named:

- "Generate_DS_OD_comb_heuristic_SARA.ipynb"
- "Generate_DS_OD_comb_heuristic_Medina_2019.ipynb"

In these scripts, a set of scores (0 to 9) were already proposed through expert elicitation, but can be changed by others. The reader should note that Figure 7 of the paper for which this repository is a supplement represents such a heuristic system. These scripts generate a set of synthetic datasets (with a .CSV format) as outputs (one per each building type) that are inputs for a subsequent process. These .CSV files are also provided in their corresponding folder (per scheme of interest). The length of each dataset (number of combinations of observed damage per damage state) fully depends on the assigned range. For instance, using the proposed scores, the SARA scheme can vary from 258 possible combinations (for the CR-LWAL-DUC-H4-7 type) to a more limited set of observable damage (110 combinations for the MCF-DNO-H1-3 type). It is worth noting that these "permissible" ranges of observed damage extension can be modified on the basis of future empirical data as well as when more robust numeral models get available. For the selection of these scores, it is recommended that the expert should have a clear knowledge of the observable damage per damage state and per building type. In this particular case, some of these features are described in the associated publications of these fragility functions (Medina et al., 2019; Villar-Vega et al., 2017).

The Python3 script "Prob_inter-scheme_damage_conv_matrices_SARA_Medina_2019.py" makes use of all of the synthetic datasets (CVS files) previously obtained to generate the probabilistic inter-scheme compatibility matrices between sets of damage states belonging to a pair of building types of two different hazard-related vulnerability schemes. The script requires of the definition of the source and target scheme. In this case, they are the classes proposed by SARA (earthquake) and Medina, 2019 (tsunami) respectively. For each of them, 10,000 exhaustive possible combinations according to the AeDES-based damage scale are simulated. Details about the theoretical framework and mathematical expressions contained in this script are provided in the paper for which this repository is a supplement.

To execute the script, the user should select the preferred machine learning technique for the classification problem to solve. The definition "*compute_damage_conversion*" is set up to handle two algorithms: *"Gaussian Naïve Bayes"* and *"Logistic regression (multinomial)"* both contained in the library Scikit-learn for the Python programming language (Buitinck et al., 2013). As an intermediate output, the probabilistic description of observed damages in terms of the four selected building elements can be obtained as .CSV files. The generation of these files is optional (no specific files are used for further calculations). Nonetheless, such files are provided in the folder "*Prob_likelihoods_OD*" for both algorithms (files starts with "PB" for the Gaussian Naïve Bayes case, and with "LR" for the Logistic regression case). The generation of these outcomes is ensured through the execution that is written right after the comment *"to obtain the probabilistic description of observed damages in .CVS format, uncomment the next line*."

The Python script *"Plot_3D_heuristics_for_SARA_and_MEDINA.py"* is provided in each of the two aforementioned folders, and can be used to generate similar 3D plots as the one shown in Fig. 8 of the paper associated to this data repository. This is useful to graphically visualise the probability of having certain combination of damage extension for a given damage state and building type described onto other building elements (i.e. FL and RF), which are different from the ones the paper displays (i.e. VS and IP). The reader should note that although quite contrasting likelihood probabilities terms can be obtained from defining various slightly similar damage-based scores, it seems that the final compatibility matrices for similar AeDES-based scores is mostly controlled by the selection of the machine learning algorithm.

Although the generation of aforementioned set of CSV files that express the differential likelihood of the damage levels on building elements is optional and are not explicitly used to obtain their associated inter-scheme probabilistic compatibility matrices, their related variables must be calculated (i.e. *B_given_od* or *A_given_od*) as they are inputs to then calculate the compatibility matrix through the definition "*get_damage_conv*", function that emulates Eq. 6 of the associated paper to this repository. Thus, the conditional probability of expressing the inter-scheme damage states solved through a Bayesian approach, (i.e. employing the aforementioned likelihood distributions, as well as assuming that the marginal probability is the proportion between one observation and the exhaustive combinations (1/10,000)).

With the former, we can obtain the final outputs, which constitute a set a files with all of the possible combinations between building types of the two schemas (i.e., 21 SARA classes by 6 Medina classes = 126 conversion matrices). These outputs can be expressed as .CSV files (useful for rapid interpretation and plotting), and as .JSON files (required by the software *DEUS* (Brinckmann et al., 2021) for further physical vulnerability calculations (damage and loss forecast), as documented in the related to this repository.

The reader should note that although this set of files with all of the compatibility values between damage states and for all of the possible combinations of pairs of building types may appear to be very extensive, when they are used ultimately used for risk assessment following the method of the associated paper (i.e. following Eq. 11 & Fig 1b), these matrices are differentially weighted according to the compatibility values of the inter-scheme building type matrices (i.e. Fig. 6b in the related paper to this data repository), and hence, hierarchized. Notably, the user can obtain two different sets of compatibility matrices (i.e. SARA-Medina (BtoA_conv_matrix) and assuming Medina as the source scheme and SARA as the target one (Medina-SARA, *AtoB_conv_matrix)*. The numerical values within those sets of matrices are different, which tells that the outlined approach generate matrices that are not invertible. In our case, we only care about SARA-Medina, due to the multi-hazard risk scope of the study, for which the change of reference schemes is strictly aligned to the order of succession in the hazard scenarios.

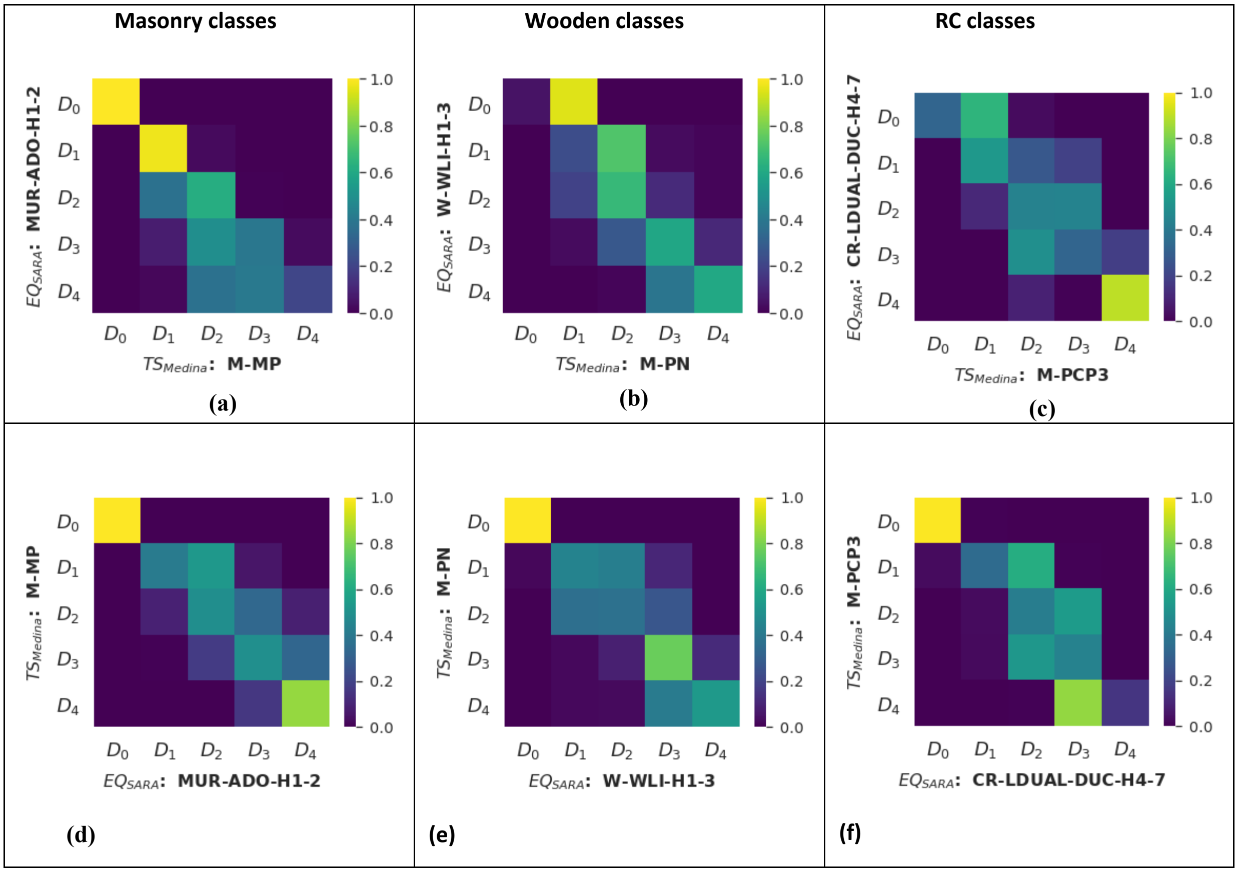

Figure 1. Probabilistic inter-scheme damage compatibility matrices for three pairs of building classes whose fragility functions of the SARA and Medina vulnerability schmes. The plots on the first row (a, b, c) display the results when the source scheme is SARA (EQ: earthquake-oriented) and the target one is Medina (TS: Tsunami-oriented) vulnerability schemes; whilst the plots on the second row (d, e, f) display the results when the source and target schemes are interchanged. All of these results were obtained making use of the Logistic Regression through the free software library Scikit-learn for the Python programming language (Buitinck et al., 2013), contrary to the associated paper for which this data repository is a supplement, which made use of the Gaussian Naïve Bayes classification algorithm.

5. References

Baggio, C., Bernardini, A., Colozza, R., Corazza, L., Della Orsini, M., Di Pascuale, G., Dolce, M., Goretti, A., Martinelli, A., Orsini, G., Papa, F., Zuccaro, G., 2007. Field Manual for post-earthquake damage and safety assessment and short-term countermeasures (AeDES), EUR 22868 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen. ed, EUR – Scientific and Technical Research series – ISSN 1018-559. Office for Official Publications of the European Communities, Luxembourg. URL: https://www.eeri.org/images/archived/wp-content/uploads/Italy/EUR%2022868%20(2007)%20Field%20Manual%20for%20post-earthquake%20damage%20assessment.pdf

Brinckmann, N., Gomez-Zapata, J.C., Pittore, M., Rüster, M., 2021. DEUS: Damage-Exposure-Update-Service. V. 1.0. GFZ Data Services. https://doi.org/10.5880/riesgos.2021.011

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G., 2013. API design for machine learning software: experiences from the scikit-learn project, in: ECML PKDD Workshop: Languages for Data Mining and Machine Learning. pp. 108–122. https://doi.org/10.48550/arXiv.1309.0238

Gómez Zapata, J.C., Pittore, M., Cotton, F., Lilienkamp, H., Simantini, S., Aguirre, P., Hernan, S.M., 2022. Epistemic uncertainty of probabilistic building exposure compositions in scenario-based earthquake loss models. Bulletin of Earthquake Engineering. https://doi.org/10.1007/s10518-021-01312-9

Gómez Zapata, J.C., Shinde, S., Pittore, M., Merino-Peña, Y., 2021. Scripts to generate (1) attribute-based fuzzy scores for SARA and HAZUS building classes, and (2) probabilistic inter-scheme compatibility matrices. An application on the residential building stock of Valparaiso (Chile) for seismic risk applications. GFZ Data Services. https://doi.org/10.5880/riesgos.2021.002

Medina, S., 2019. Zonificación de la vulnerabilidad física para edificaciones típicas en San Andrés de Tumaco, Costa Pacífica Colombiana (Master thesis in Civil Engineering). Universidad Nacional de Colombia Facultad de Ingeniería, Departamento Ingeniería Civil y Ambiental, Bogotá, Colombia.

Medina, S., Lizarazo-Marriaga, J., Estrada, M., Koshimura, S., Mas, E., Adriano, B., 2019. Tsunami analytical fragility curves for the Colombian Pacific coast: A reinforced concrete building example. Engineering Structures 196, 109309. https://doi.org/10.1016/j.engstruct.2019.109309

Pittore, M., Haas, M., Megalooikonomou, K.G., 2018. Risk-Oriented, Bottom-Up Modeling of Building Portfolios With Faceted Taxonomies. Frontiers in Built Environment 4, 41. https://doi.org/10.3389/fbuil.2018.00041

Villar-Vega, M., Silva, V., Crowley, H., Yepes, C., Tarque, N., Acevedo, A.B., Hube, M.A., Gustavo, C.D., María, H.S., 2017. Development of a Fragility Model for the Residential Building Stock in South America. Earthquake Spectra 33, 581–604. https://doi.org/10.1193/010716EQS005M

Yepes-Estrada, C., Silva, V., Valcárcel, J., Acevedo, A.B., Tarque, N., Hube, M.A., Coronel, G., María, H.S., 2017. Modeling the Residential Building Inventory in South America for Seismic Risk Assessment. Earthquake Spectra 33, 299–322. https://doi.org/10.1193/101915EQS155DP