Subash Ghimire
Philippe Guéguen

Evaluation of Machine Learning Models for Average Annual Losses Assessment

Comparison with ESRM20 Results in France

ISTerre @ Université de Grenoble Alpes
philippe.gueguen@univ-grenoble-alpes.fr

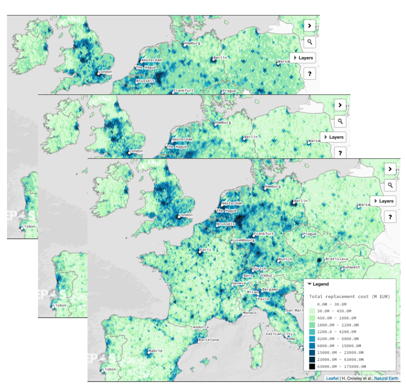
Seismic Exposure and Vulnerability Models

Hazard → Exposure model → Taxonomie → Capacity/ Damage scale → Fragility function → Loss function

Hazard

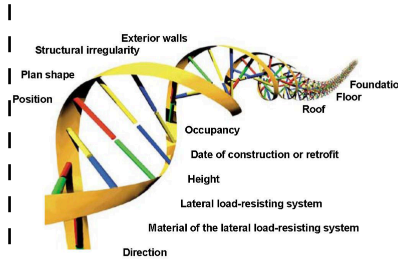
Scenario-based
Intensity-based
Frequency-based

Exposure model



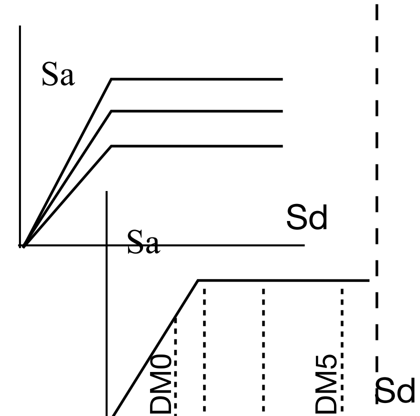
Population
Assets
Urban fabrics
$ metrics (repl. cost)

Taxonomie



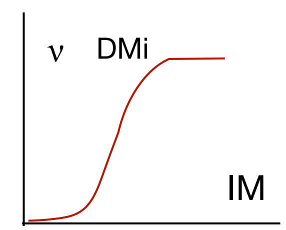
Materials, Lateral load resisting systems, Code Level or Ductility, Height…

Capacity/ Damage scale



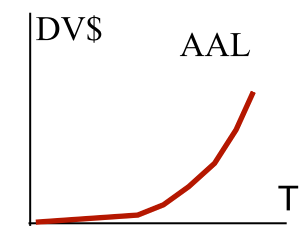
Engineering-based models (nonlinear dynamic analysis), Empirical-based analysis

Fragility function

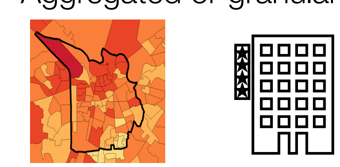


P(DMi|Classe,IM)

Loss function



Aggregated or granular



Damage ratio (related to replacement cost) + human vulnerability (related to entrapment) +uncertainties

Scientific issues

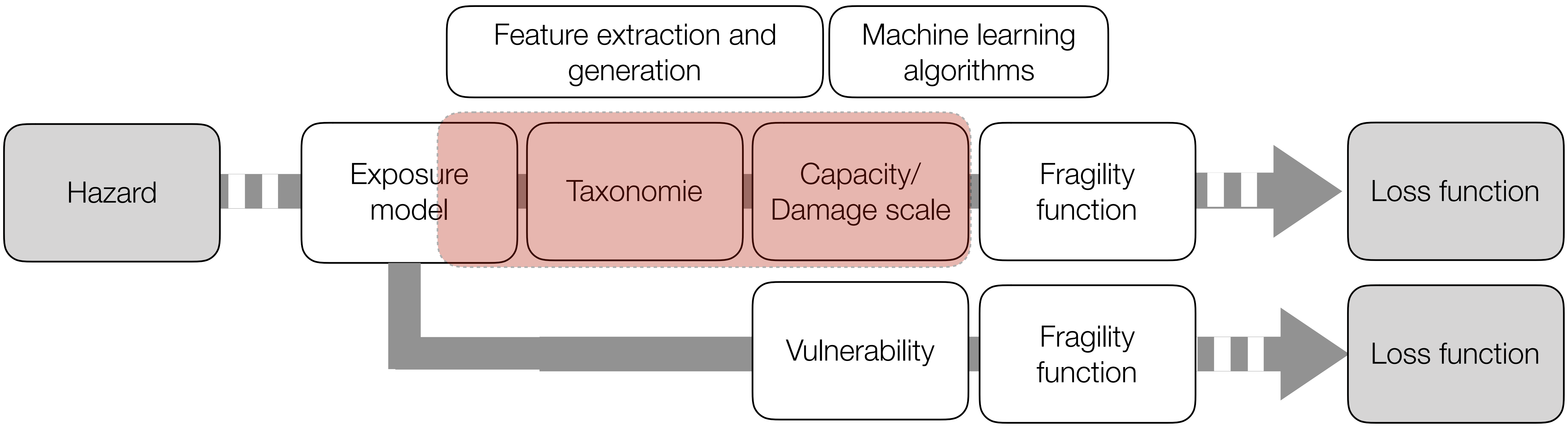
Exposure model - How to quantify the physical exposure model (buildings) and their vulnerability ?

Capacity/damage scale - How to assess the degree of damage based on building types, construction designs, location …?

Host-to-target adjustment of fragility function?

Machine learning in seismic risk models

Vulnerability-based methodology



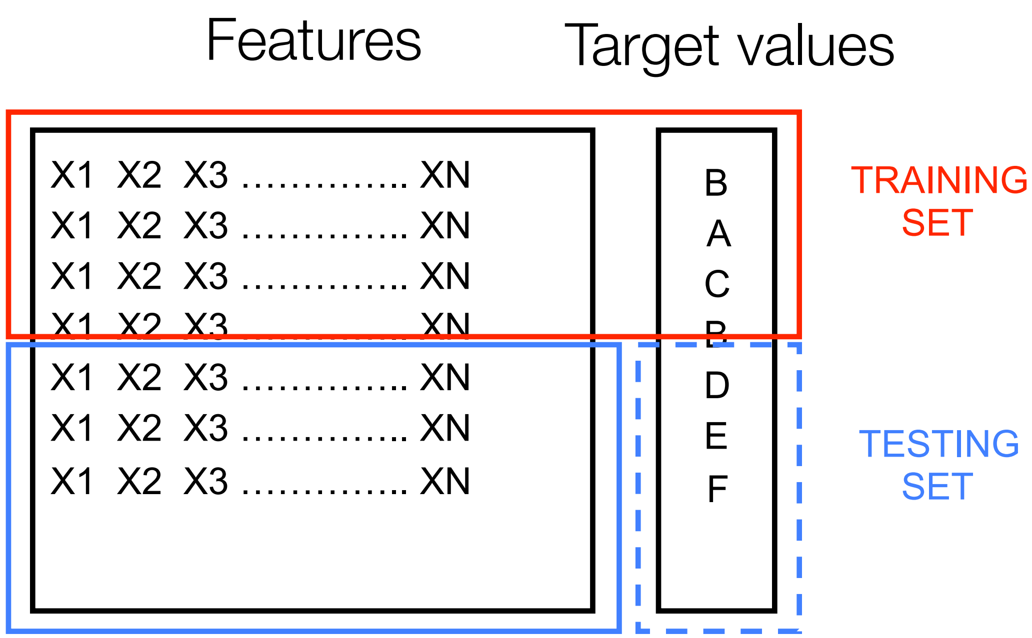
Predictive analytics: use Machine Learning to predict vulnerability function based on available metadata (features) and spatial patterns

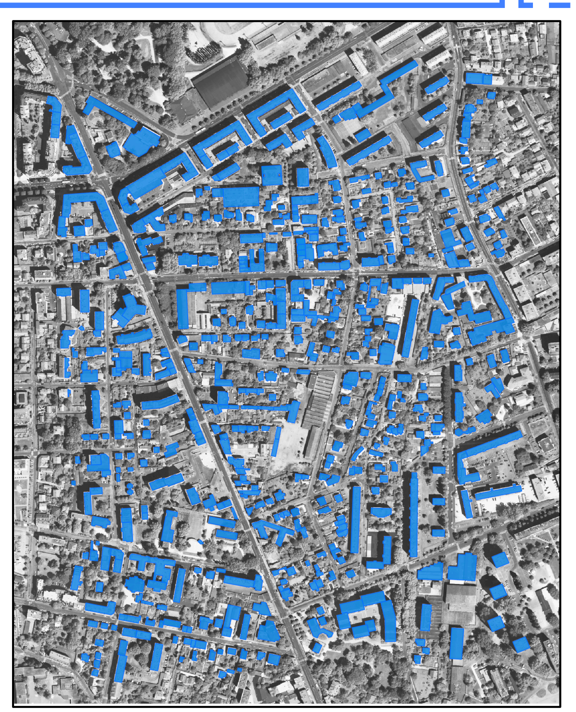
Machine learning in seismic risk models

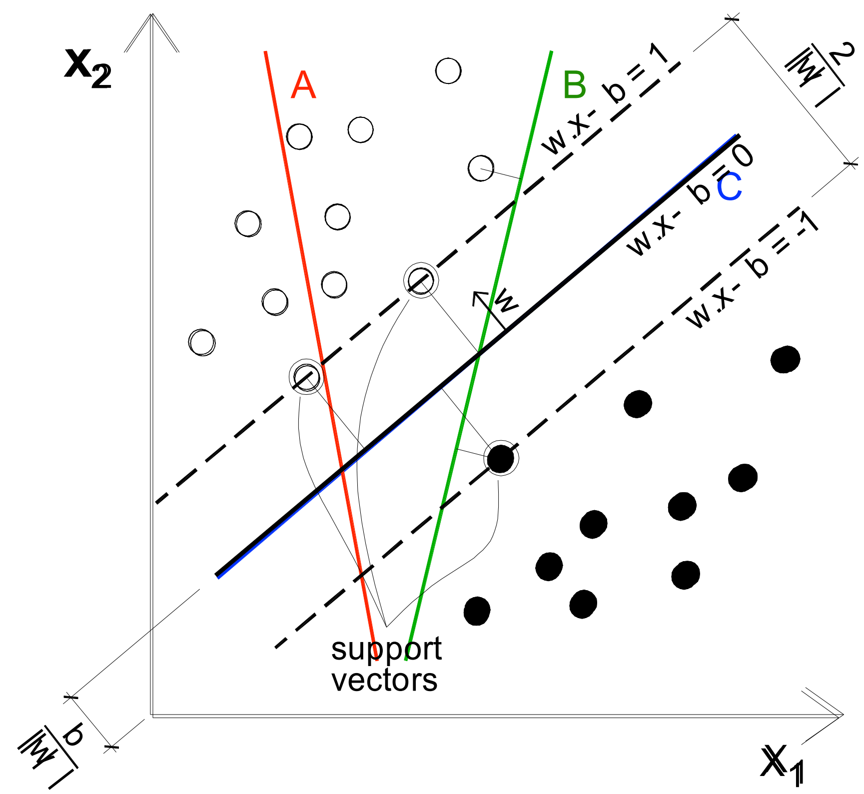
Vulnerability-based methodology - Riedel, Guéguen et al., 2017

Support Vector Machine (SVM)

Supervised learning models with associated learning algorithms that ANALYSE data and RECOGNIZE patterns, used for CLASSIFICATION.





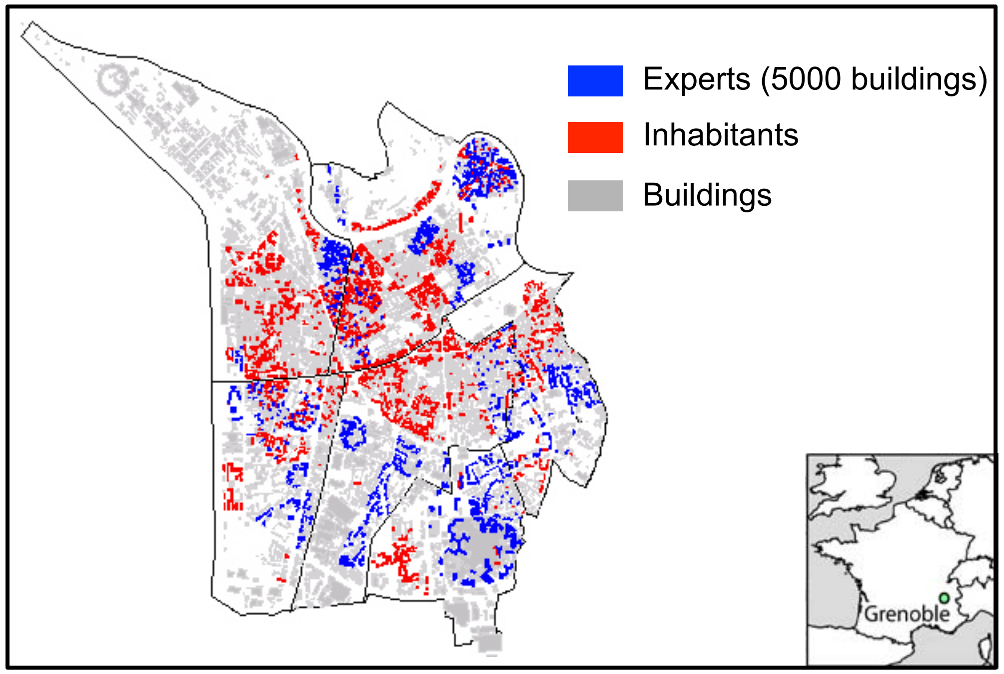


Binary and Linear Classification

Minimize (in **w**, b) ||**w**|| ; subjected to (for any i = 1…n) yi (**w**.**x**i – b) >= 1

Machine learning in seismic risk models

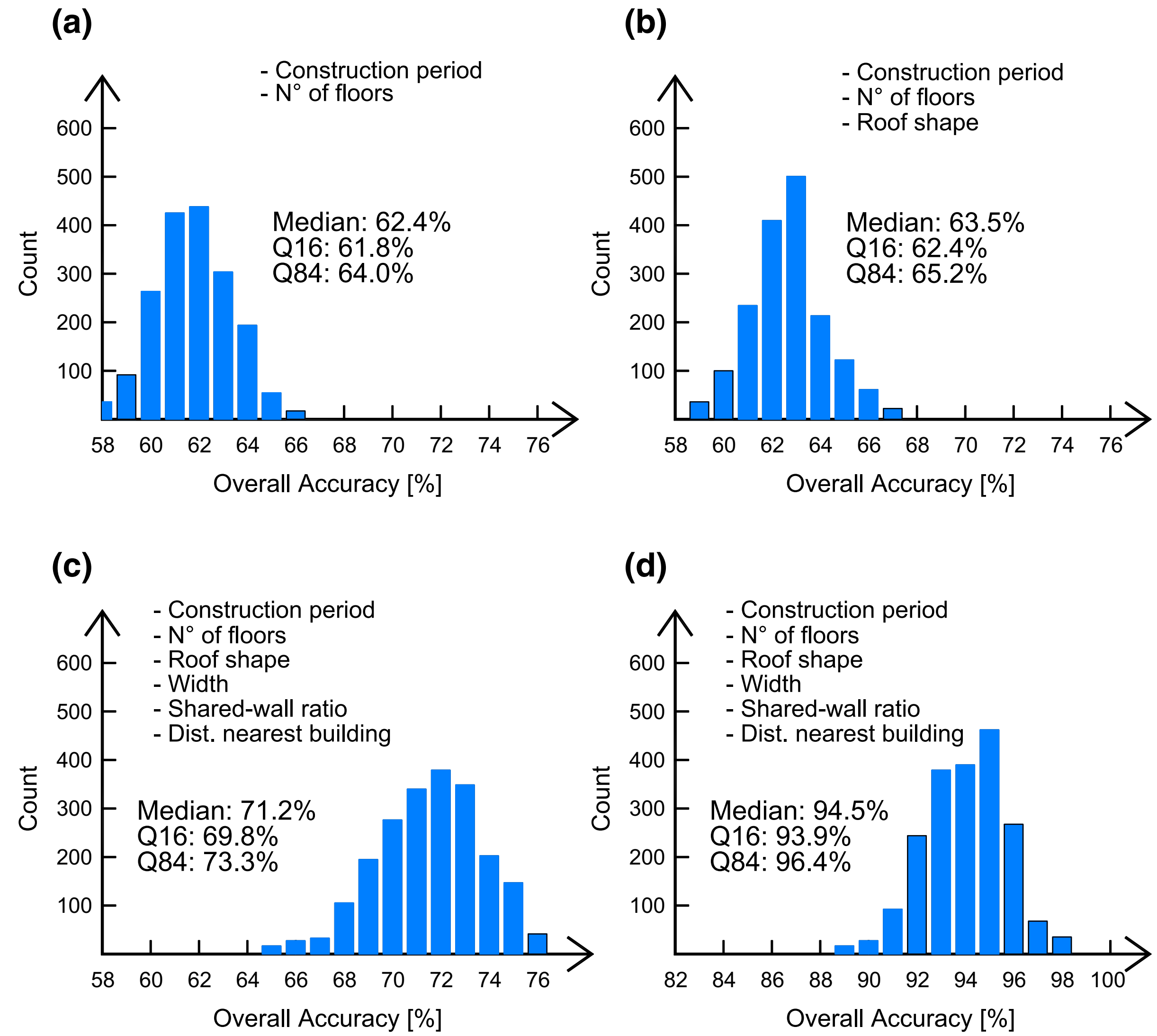
Vulnerability-based methodology - Riedel, Guéguen et al., 2017



(Gueguén et al., 2007)

Confusion matrix : Construction period - Num of floors

	A	B	C	D	E	F	
A	**131**	121	37	0	0	0	
B	111	**941**	78	21	0	0	
C	29	86	**571**	395	43	0	
D	9	7	107	**249**	193	0	
E	0	0	8	32	**331**	0	
F	0	0	0	0	0	**0**	
	280	1155	801	697	567	0	3500
Acc.	0.629						



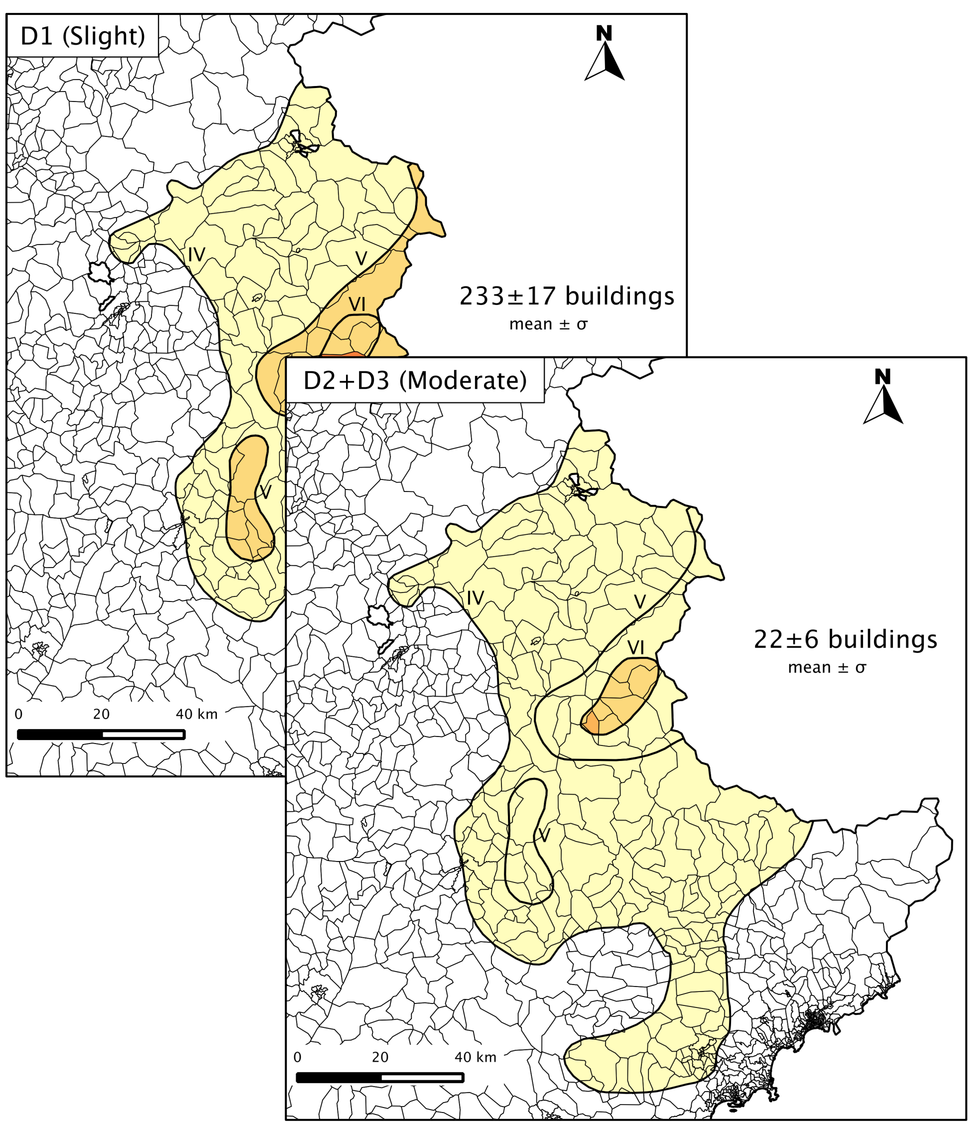
Machine learning in seismic risk models

Vulnerability-based methodology - Riedel, Guéguen et al., 2017

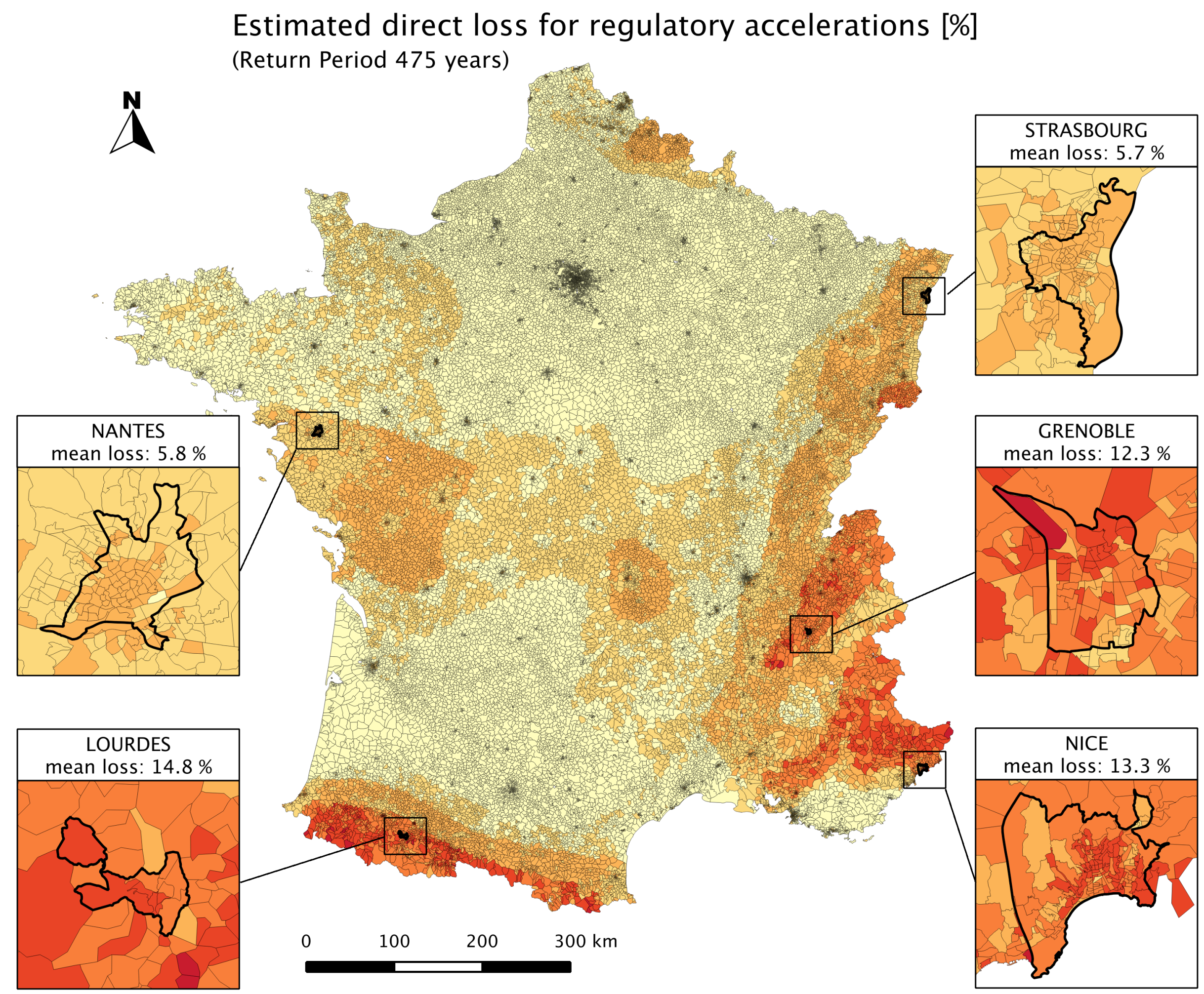
Scenario-based testing

Ubaye earthquake (M 4.9 - 2014)

Observed: **272** damaged buildings (macroseismic field)
Predicted: **255 +/- 33** D1/D2/D3 buildings



Intensity-based application



$$C^{direct} = \sum_{i=1}^{i=5} P_{Di} * C_{Di}$$

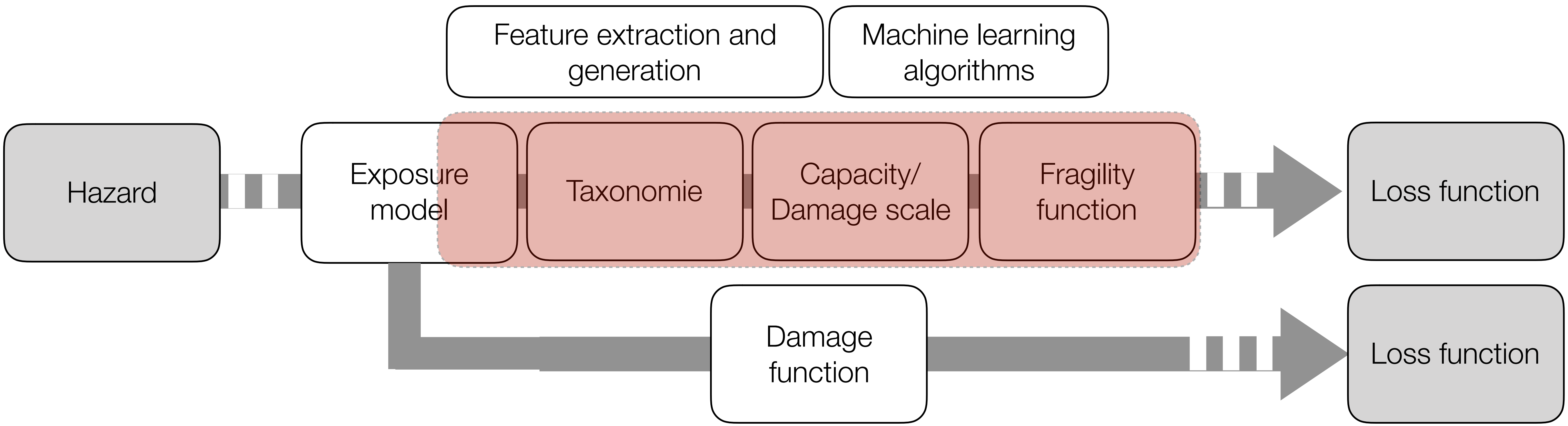
Mean Direct Loss
(% of total building stock value)

Percentage of buildings
(with damage Di)

Loss Ratio
(for damage Di)

Machine learning in seismic risk models

Damage-based methodology



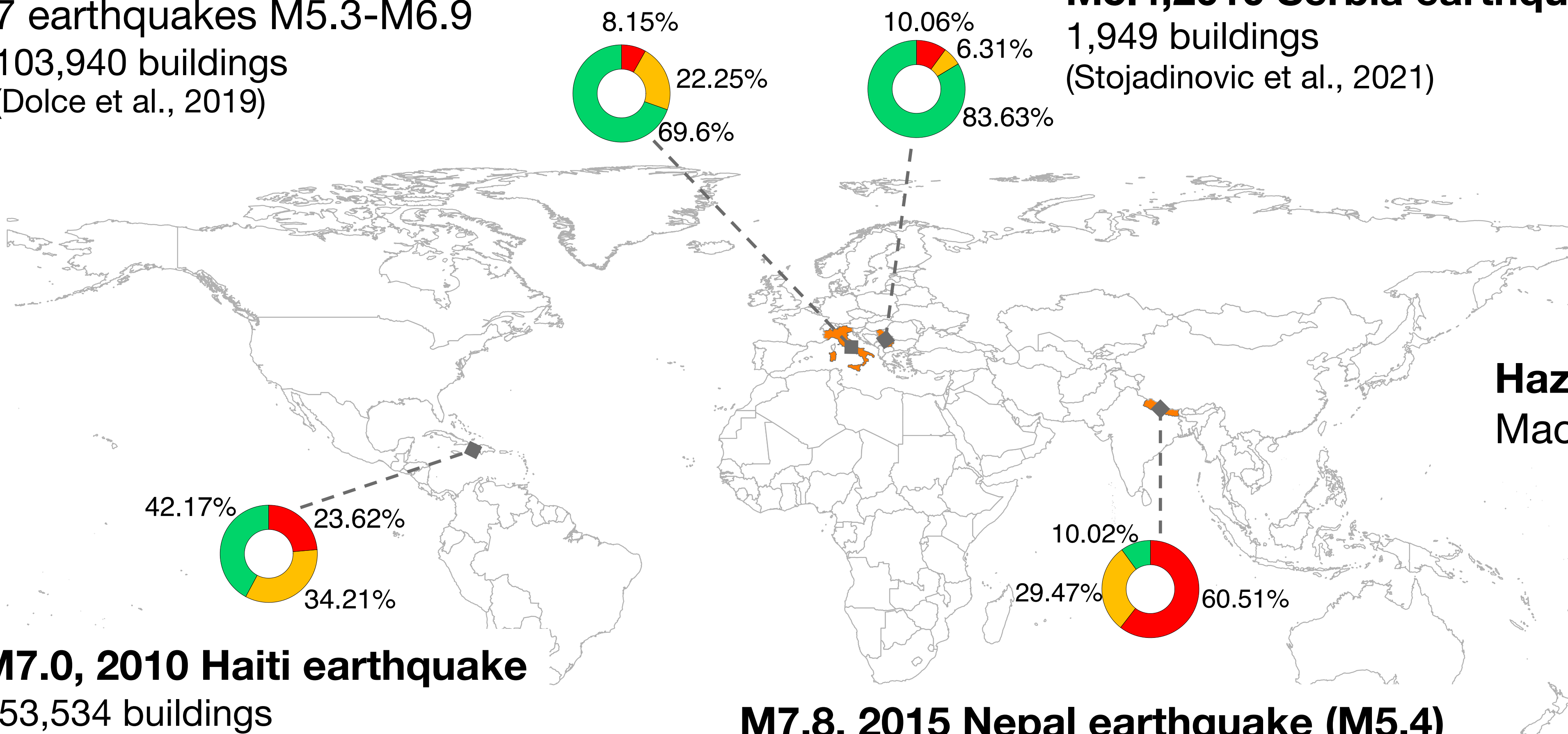
Predictive analytics: use Machine Learning to predict damage function based on available metadata (features) and spatial patterns

Machine learning in seismic risk models
Damage-based methodology

Post-earthquake macroseismic surveys

DaDO Italian data
7 earthquakes M5.3-M6.9
103,940 buildings
(Dolce et al., 2019)

M5.4,2010 Serbia earthquake
1,949 buildings
(Stojadinovic et al., 2021)



M7.0, 2010 Haiti earthquake
353,534 buildings
(MTPTC, 2010)

M7.8, 2015 Nepal earthquake (M5.4)
757,362 buildings
(NPC, 2015)

Hazard:
Macroseismic intensity (MSI)

ShakeMaps
USGS
science for a changing world

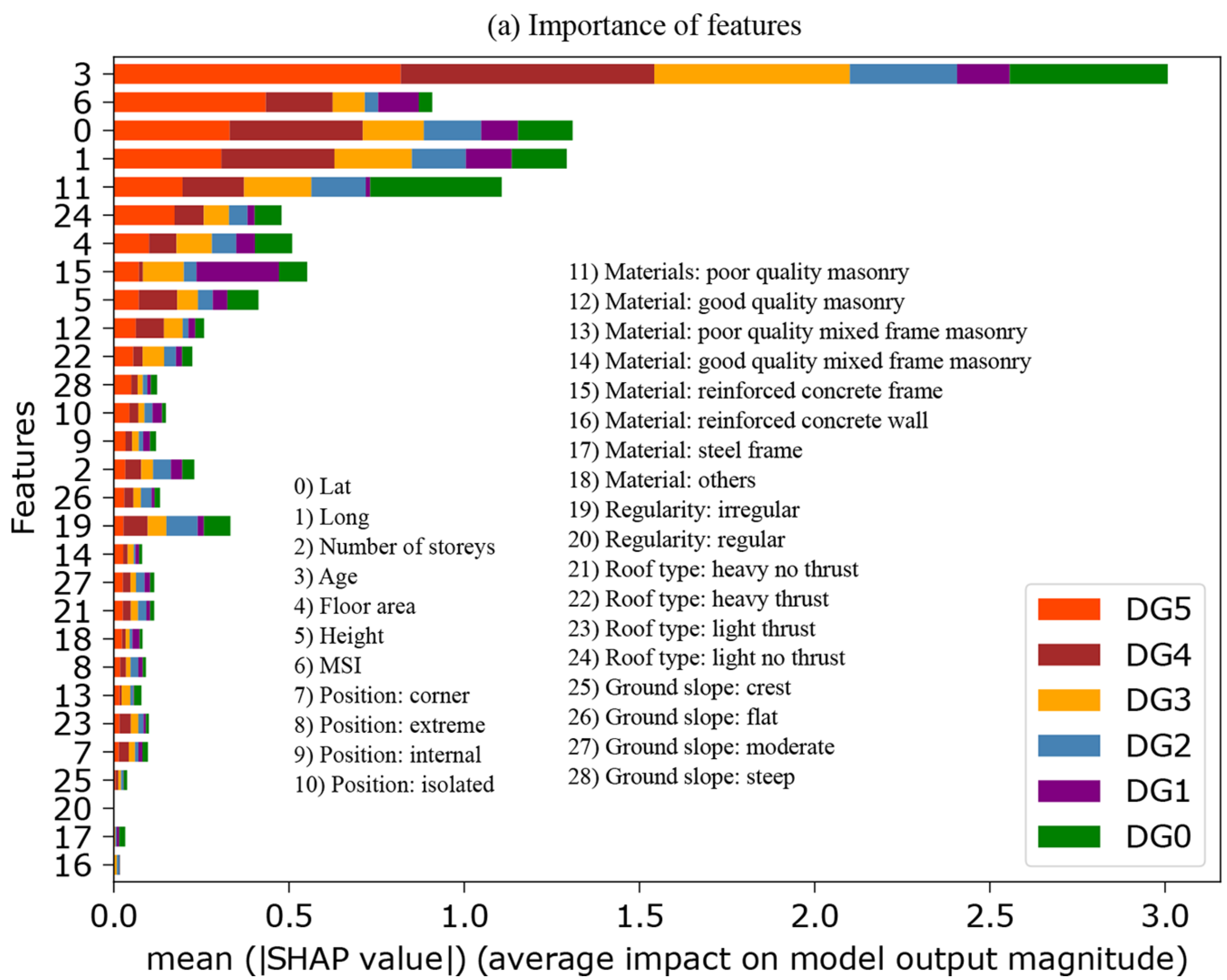
Damage scale:
EMS98 - DG0-DG5
Traffic Light classification -
DG0+DG1, DG2+DG3, DG4+DG5

Machine learning in seismic risk models

Damage-based methodology - **Ghimire et al.**, Earthq. Spec. 2022; NHESS2023

DaDO Italian earthquakes

Training/Testing = 60%/40%



Methods

Six methods tested (3 classification, 3 regression)
random forest, gradient boosting, extreme gradient boosting

Classification models performed slightly better

The most efficient methods: Extreme gradient boosting classification (XGBC) (Chen and Guestrin, 2016).

Imbalance issue

Four methods tested
Random undersampling, random oversampling, synthetic minority oversampling technique (SMOTE) and SMOTE-ENN.

Random oversampling method by rectifying the skewed distribution of the target features (DGs).

Features

Weight of the most important building feature evolves according to DG.

Accuracy score (for TLS damage classification)

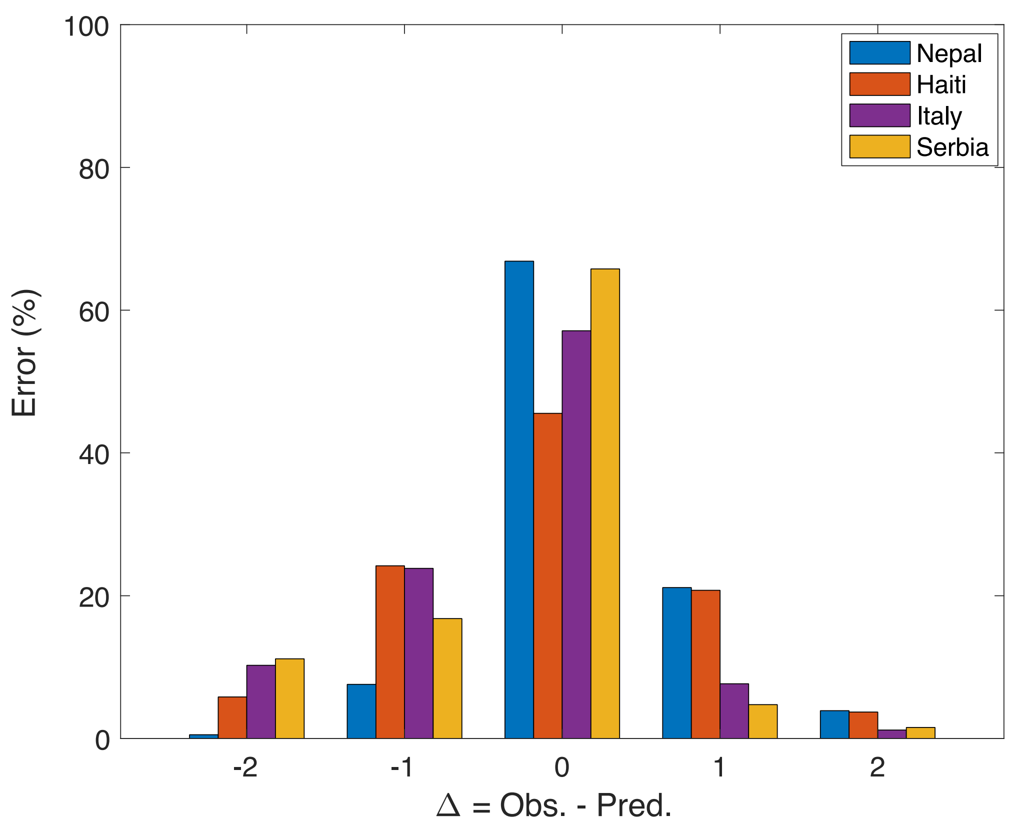
Basic-features setting : 0.68
Full-features setting : 0.72

Machine learning in seismic risk models

Damage-based methodology - **Ghimire et al.**, Nat. Haz. 2024

Training/Testing = 60%/40% -
Features = MSI, Nb of floor, Building age

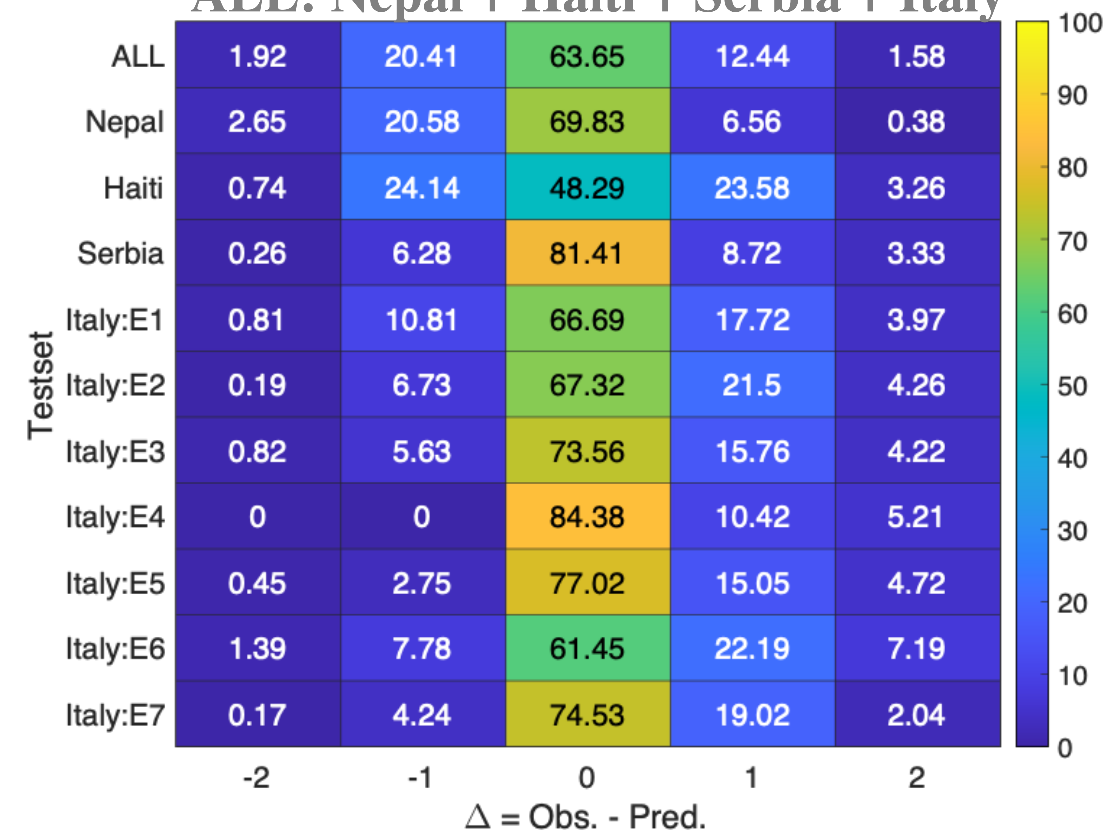
Given-earthquake



$$\text{Error} = \frac{\sum DG_{obs} - DG_{pred}}{N} \; x \; 100 \; (\%)$$

Aggregated-earthquake

ALL: Nepal + Haiti + Serbia + Italy



Accuracy score in other similar studies:

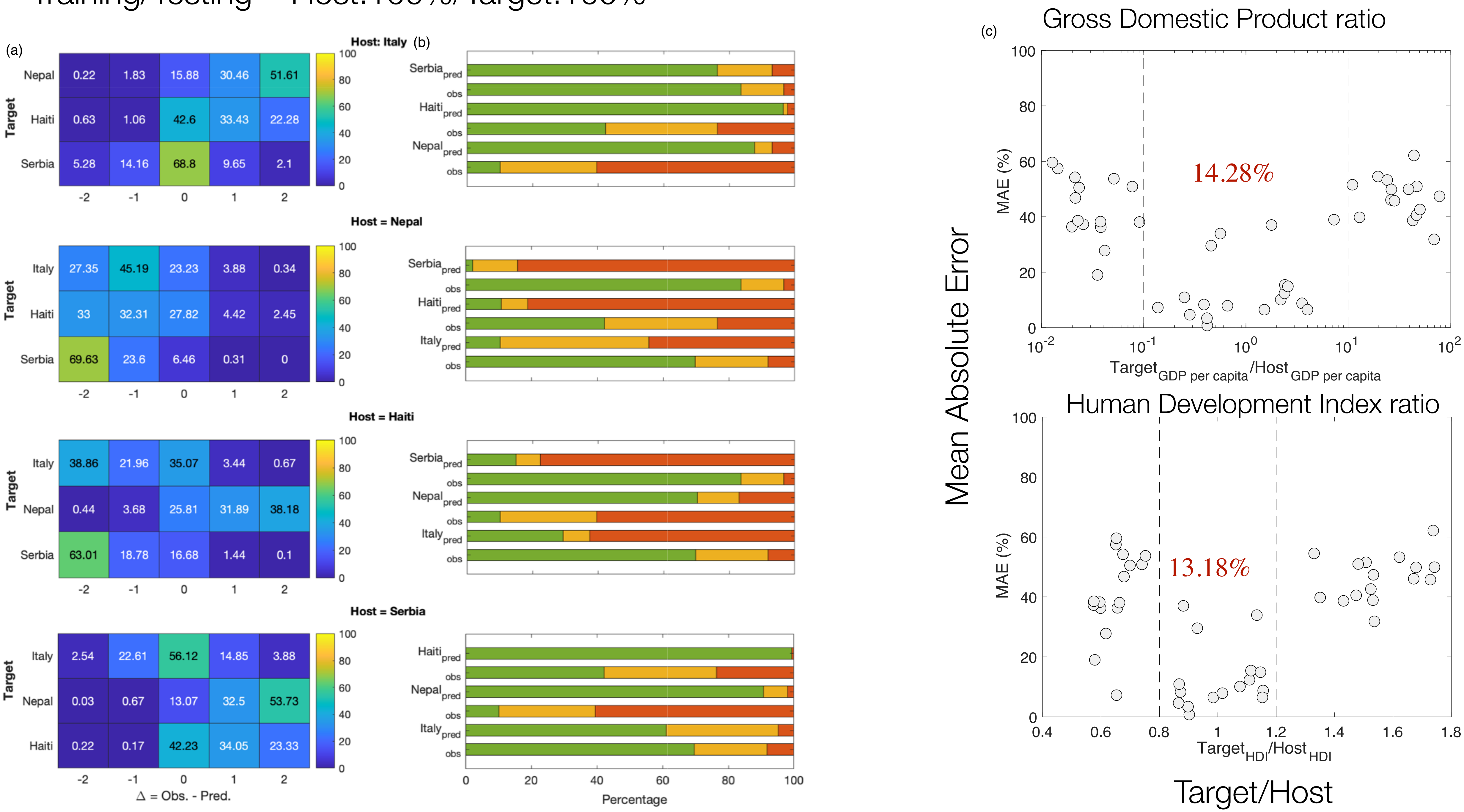
Mangalatheu et al. (2020): **66%**
Roslin et al. (2020): **67%**
Harirchian et al. (2021): **65%**
Ghimire et al. (2022): **68%**

Machine learning in seismic risk models

Damage-based methodology - **Ghimire et al.**, Nat. Haz. 2024

Model effectiveness : Host-to-Target adjustment

Training/Testing = Host:100%/Target:100%

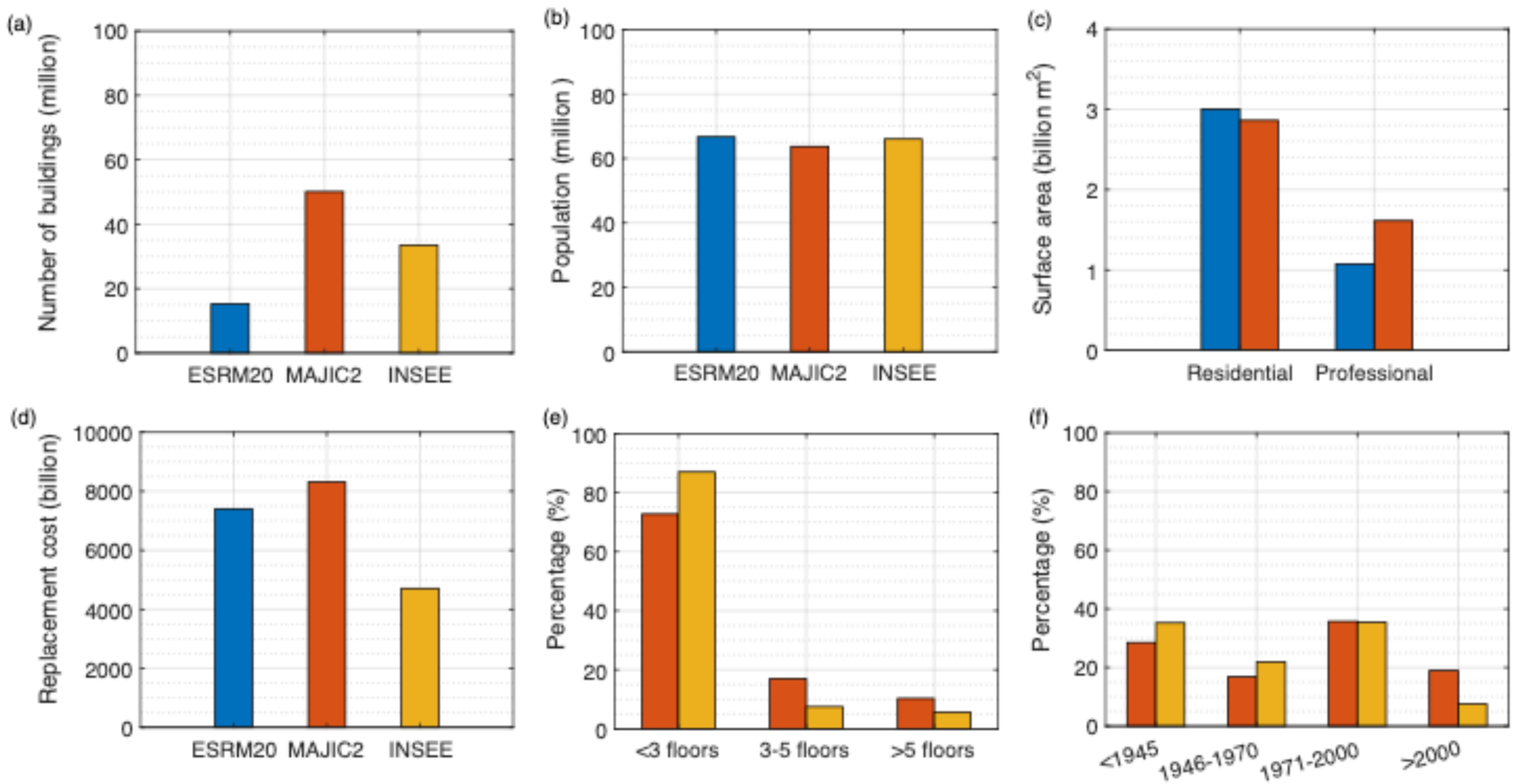


Testing ESRM20 with ML-based methods

For France

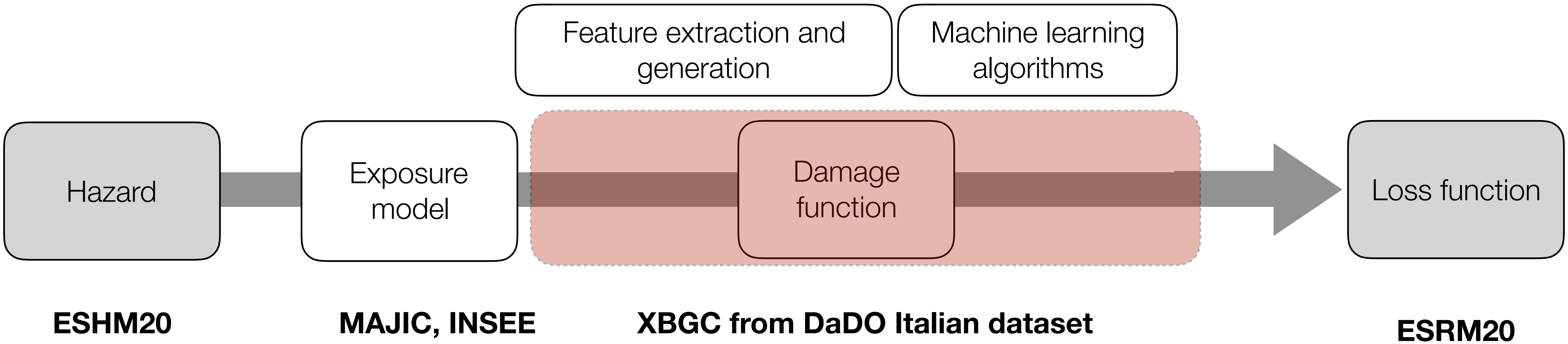
Exposure models for France

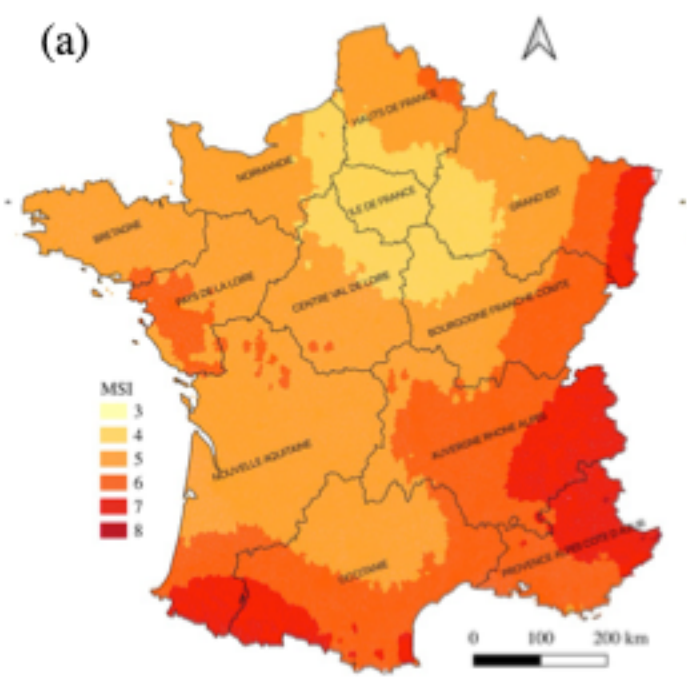
ESRM20, INSEE (National Census), MAJIC2 (French Cadastral Information)



Testing ESRM20 with ML-based methods

For France





Building classification (feature for ML): Number of floors, Age, MSI

DG1-DG5 damage classification

$$AAL = \frac{1}{T}\sum_{BC}(0 * N_{DG1} + 0.05 * N_{DG2} + 0.15 * N_{DG3} + 0.6 * N_{DG4} + 1.0 * N_{DG5})$$

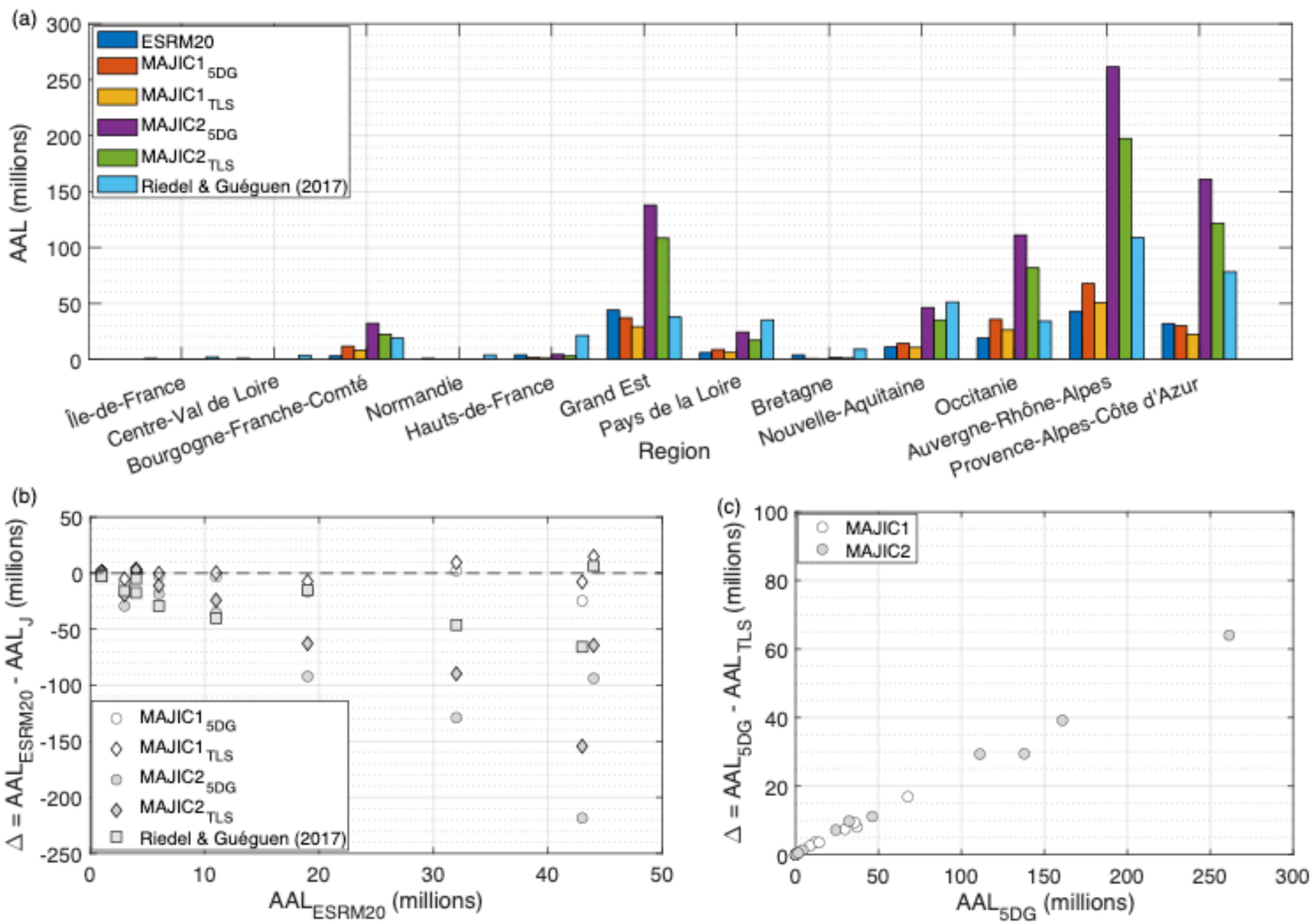
TLS damage classification

$$AAL = \frac{1}{T}\sum_{BC}(0 * N_{Green} + 0.1 * N_{Orange} + 0.8 * N_{Red})$$

Damage-to-loss ratio from ESRM20 (Crowley et al., 2021)

Testing ESRM20 with ML-based methods

For France



Summary

- Current models rely on engineering-based data and scenario simulations, but they are often limited by data availability, uncertainty, and computational complexity.

- Machine learning offers significant potential for improving the accuracy, scalability, and real-time capabilities seismic risk assessment, leading to more informed decisions for disaster preparedness and response.

- The future of seismic loss assessment lies in the integration of machine learning with traditional methods to create more dynamic, data-driven, and actionable models.

Opportunities

- Real-time, data-driven decision-making for risk reduction and emergency response.
- Test more accurate, adaptable vulnerability models using big data.

Challenges

- Data availability: The quality and granularity of exposure, and vulnerability data can limit model accuracy.
- Interpretability: Machine learning models often lack transparency, which is a challenge for regulatory use and policy-making.
- Integration: How to integrate machine learning models with existing risk assessment frameworks and decision-making processes.

Subash Ghimire
Philippe Guéguen

Evaluation of Machine Learning Models for Average Annual Losses Assessment

THANK YOU

Open the floor for questions and discussion

ISTerre @ Université de Grenoble Alpes
philippe.gueguen@univ-grenoble-alpes.fr