



Integrated Seismic Risk Modelling and Parametric Disaster Financing in Indonesia Using Probabilistic Hazard Analysis, Machine Learning, and Actuarial Simulation

Adhitya Ronnie Effendie^{1*}, Gunardi Gunardi¹, Nanang Susyanto¹, Wiwit Suryanto²,
Wan Zuki Azman Wan Muhamad³, Anas Fauzi Masykuri², Hartati Hartati¹

¹ Department of Mathematics, Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Yogyakarta 55281, Indonesia

² Department of Physics, Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Yogyakarta 55281, Indonesia

³ Institute of Engineering Mathematics, Universiti Malaysia Perlis, Arau 02600, Malaysia

Corresponding Author Email: adhityaronnie@ugm.ac.id

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.130308>

ABSTRACT

Received: 20 October 2025

Revised: 7 December 2025

Accepted: 15 December 2025

Available online: 10 April 2026

Keywords:

actuarial simulation, calibrated earthquake probability, disaster risk financing, Indonesia, machine learning, parametric insurance, probabilistic seismic hazard analysis, tail value-at-risk

This study examines whether combining machine-learning models with seismicity indicators can improve the calibration and skill of short-term earthquake probabilities for parametric disaster financing in Indonesia. An integrated framework is developed that combines probabilistic seismic hazard analysis (PSHA), machine learning, and actuarial simulation. Using the United States Geological Survey (USGS) catalog for 1925–2025, Gutenberg–Richter (GR) parameters and long-term Poisson exceedance rates are estimated for events with $M \geq 5.0$ on a $0.5^\circ \times 0.5^\circ$ grid. These features, together with geographic coordinates, are used as inputs to tree-based models that predict the probability of at least one event with $M \geq 5.0$ occurring within the next five years. In walk-forward validation, Random Forest and Gradient Boosting achieve modest discrimination (Area Under the Receiver Operating Characteristic Curve (AUROC) ≈ 0.60) but yield small yet well-calibrated improvements over the GR–Poisson baseline, with lower Brier scores and more reliable probabilities. For the financial component, a parametric payout rule is applied in which payouts increase linearly with magnitude between 5.0 and 7.5. Monte Carlo simulations based on the long-term rates yield estimates of expected annual loss and the 95% tail value-at-risk (TVaR), highlighting skewed risk concentrated along the Sumatra subduction corridor. The framework assumes independent Poisson occurrence within each grid cell, uses a coarse spatial resolution, and does not explicitly model spatial correlation or aftershock clustering.

1. INTRODUCTION

Situated at the meeting point of the Pacific, Indo-Australian, and Eurasian plates, Indonesia is among the world's most seismically active regions [1]. Subduction zones such as the Sunda Megathrust and major strike-slip faults like the Sumatra and Palu–Koro faults generate frequent moderate-to-large earthquakes, often with devastating consequences [2]. Historical events such as the 2004 Aceh–Andaman earthquake and tsunami, the 2009 Padang earthquake, and the 2018 Palu–Donggala earthquake resulted in severe casualties and economic losses, highlighting the urgent need for robust hazard assessment and sustainable disaster financing mechanisms in Indonesia [3, 4].

Classical probabilistic seismic hazard analysis (PSHA) is typically based on the Gutenberg–Richter (GR) frequency–magnitude law [5] and the Poisson model of earthquake occurrence [6]. These approaches remain the backbone of regional and global seismic hazard studies, including those by Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) and the global earthquake model [7, 8]. However, classical models

often assume temporal and spatial stationarity, which may overlook nonlinear patterns and clustered behaviors in earthquake catalogs. In parallel, actuarial methods such as Average Annual Loss (AAL) and tail value-at-risk (TVaR) have been increasingly applied to quantify financial exposure to catastrophes and to support the design of insurance contracts and catastrophe bonds [9, 10]. Recent advances in machine learning (ML), including Random Forests (RFs) and Gradient Boosting (GB), have demonstrated their ability to capture complex nonlinearities in spatio-temporal hazard data and improve earthquake prediction and risk classification [11]. At the global scale, initiatives such as the global earthquake model and recent regional hazard and risk studies have provided consistent probabilistic estimates for many regions, including Southeast Asia [7–9]. However, the explicit integration of such hazard information with machine-learning-based probabilities and parametric pricing tailored to the Indonesian context remains limited.

Despite these advances, research gaps remain. First, most studies in Indonesia have focused either on seismic hazard quantification or financial loss modeling, with few attempts to

integrate these perspectives. Second, existing hazard models rarely incorporate ML-based predictive layers, and actuarial applications often rely on simplified assumptions without explicit hazard linkage. Finally, validation strategies in many studies are limited, with little emphasis on temporal walk-forward evaluation or spatial holdout testing to ensure generalization.

This study addresses these gaps by developing an integrated end-to-end framework that links seismic hazard quantification, ML-based risk modeling, and actuarial premium design for parametric earthquake insurance in Indonesia. Using the United States Geological Survey (USGS) earthquake catalog (1925–2025), GR parameters and Poisson exceedance probabilities are estimated on a uniform $0.5^\circ \times 0.5^\circ$ spatial grid. ML models (RFs and GB) are trained on physics-informed features, and a parametric trigger (attachment $A = 5.0$, exhaustion $E = 7.5$) is applied in Monte Carlo simulations (10,000 synthetic years per grid) to compute AAL, TVaR, and insurance premiums at grid and provincial scales.

The novelty of this study lies in four aspects:

- (i) the integration of seismological models, ML, and actuarial methods into a single reproducible workflow;
- (ii) the use of long-term historical catalogs (≈ 100 years) aggregated at fine spatial resolution;
- (iii) the adoption of dual validation schemes (temporal walk-forward and spatial holdout) to enhance robustness; and
- (iv) the demonstration of premium differentiation across Sumatra provinces, providing actionable insights for risk-based disaster financing.

This study develops an integrated framework tailored to the Indonesian seismic context that links grid-based seismic hazard metrics to illustrative parametric insurance pricing. Concretely, this study asks whether combining simple machine-learning models with classical PSHA-based seismicity indicators can yield better-calibrated short-term exceedance probabilities and risk metrics that are useful for designing illustrative parametric earthquake financing schemes for Indonesia. Rather than claiming to be the first of its kind, our contribution lies in adapting and operationalizing this type of workflow using Indonesian earthquake catalogs, exposure proxies, and risk metrics relevant for disaster-risk financing.

2. LITERATURE REVIEW

2.1 Seismic risk and probabilistic seismic hazard analysis

In this study, the standard PSHA formulation is adopted in the sense of Cornell [12] and McGuire [10], where the GR frequency–magnitude relation is combined with a Poissonian occurrence model to obtain annual exceedance rates for selected ground-motion thresholds. Since the seminal formulation by Cornell, PSHA has underpinned building codes and regional hazard maps worldwide, including Indonesia [8, 10, 12]. Nonetheless, classical PSHA generally assumes temporal and spatial stationarity; such assumptions can underrepresent clustered and non-linear patterns observed in tectonically active subduction environments like Sumatra, motivating complementary approaches that enrich PSHA with data-driven signals.

2.2 Earthquake data modeling

Earthquake catalogs published by official agencies such as the USGS and BMKG provide the primary data sources for seismic analysis and monitoring [13]. These catalogs typically include event parameters such as time, location, magnitude, and depth, which, after preprocessing, are used to construct hazard models. Several studies emphasize the importance of careful data preprocessing to ensure reliable seismicity estimates. In particular, magnitude homogenization plays a key role when combining events from different magnitude scales and catalogs [14]. However, while catalog data have become increasingly standardized, spatial and temporal biases often persist, particularly for smaller-magnitude events. This indicates the limitations of relying solely on historical catalogs to build robust seismic risk models.

2.3 Machine learning in seismology

Advances in ML have opened new opportunities for seismic analysis. ML has been applied to enhance earthquake detection, aftershock prediction, and event probability estimation, often achieving higher accuracy than traditional methods [11, 15]. Models such as the Earthquake Transformer [11] and deep-learning approaches to aftershock sequences [15] demonstrate ML's capability to capture spatio-temporal complexities that conventional models struggle with.

Beyond detection and short-term forecasting, a growing body of work applies ML to risk quantification and catastrophe finance, for example, using tree-based ensembles to classify high-risk zones, predict portfolio losses, or support the pricing of catastrophe-linked securities and risk transfer instruments [9]. These studies demonstrate that ML can complement actuarial models by capturing nonlinear relationships between hazard indicators, exposure, and loss outcomes. In this study, a similar perspective is adopted, where RFs and GB are employed as calibrated risk models whose probability outputs can be interpreted within an actuarial framework, rather than as stand-alone detection tools.

2.4 Parametric insurance and disaster financing

From a financial perspective, parametric insurance has emerged as an innovative instrument for disaster risk financing [16]. Unlike traditional indemnity-based insurance, parametric schemes trigger payouts based on measurable physical parameters, such as earthquake magnitude or epicentral location. Premiums are typically derived from the expected payout adjusted for risk loads using actuarial measures such as TVaR [17]. Empirical studies have demonstrated the effectiveness of parametric insurance in accelerating claims settlement and strengthening fiscal resilience [18]. However, the literature remains relatively limited in integrating seismic hazard modeling with actuarial frameworks for parametric premium determination, particularly in the context of Southeast Asia. Addressing this gap is the key contribution of this study.

2.5 Related work and research gap

Several studies have contributed to advancing seismic hazard assessment and disaster risk financing, yet most have remained within disciplinary silos. Classical works on PSHA, building on the GR relation and Poisson models, continue to

provide the foundation for hazard quantification [12, 10]. Recent global and regional probabilistic hazard maps, including national maps produced for Indonesia (e.g., BMKG and Pusat Studi Gempa Nasional/PUSGEN), provide a basis for spatially explicit seismic risk assessment [19], and global earthquake models [20] extend these approaches into policy-relevant products. However, these studies often assume temporal stationarity and do not explicitly address predictive enhancement.

Parallel efforts in ML have shown promising results in earthquake detection, phase picking, and short-term aftershock modeling [21]. Ensemble methods such as RFs and GB have been applied for risk classification, but their integration into long-term hazard modeling frameworks remains limited. At the same time, actuarial approaches have been increasingly applied for catastrophe risk quantification, emphasizing AAL and TVaR as metrics for premium calculation [9]. Nonetheless, studies explicitly connecting seismic hazard parameters to parametric insurance pricing in Indonesia are scarce.

Taken together, prior literature establishes a strong foundation in hazard quantification, data-driven modeling, and actuarial finance. Yet, a gap remains in unifying these approaches into a reproducible, end-to-end workflow tailored to the Indonesian context. Specifically, there is limited research that:

1. Integrates GR–Poisson hazard modeling with ML predictors to improve forward-looking exceedance estimation.
2. Links hazard outputs directly to actuarial metrics (AAL, TVaR) within a parametric insurance framework.
3. Demonstrates provincial-scale premium differentiation in high-risk regions such as Sumatra.

This study addresses these gaps by combining seismic hazard quantification, ML-based predictive modeling, and actuarial premium design into a unified framework. The approach not only improves scientific understanding of earthquake risk but also provides actionable insights for disaster risk financing in Indonesia.

3. METHODOLOGY

This study develops an integrated end-to-end framework linking seismic hazard assessment, ML, and actuarial principles to quantify earthquake risk and design parametric insurance premiums. The methodology consists of the following stages:

3.1 Data collection and preprocessing

The primary dataset is the earthquake catalog provided by the USGS, covering events within the Indonesian region [22]. Each record includes event date, time, latitude, longitude, magnitude, and depth. The catalog undergoes standardized preprocessing consisting of:

- a. Deduplication: removing duplicate entries across overlapping records.
- b. Missing value imputation: addressing incomplete metadata, particularly for magnitude and depth.
- c. Consistency checks: validating magnitude types and depth values to ensure data quality.
- d. Timestamp synchronization: standardizing time zones and formats.

- e. Spatial aggregation: mapping earthquake events to a uniform spatial grid ($0.5^\circ \times 0.5^\circ$) [22].

The resulting dataset contains 71,340 events, of which 14,466 are $M \geq 5.0$, distributed across 2,193 active grid cells (see Section 4.1).

3.2 Seismic hazard quantification

Seismicity within each spatial grid is summarized using the GR relation [23]:

$$\log_{10} N(M \geq m) = a - bm$$

where, a represents seismic productivity and b denotes the relative frequency of small versus large earthquakes.

Assuming a Poisson process, the annual exceedance rate λ for earthquakes above threshold magnitude $M_0 = 5.0$ is [5]:

$$\lambda(M \geq M_0) = \frac{10^{(a-b \cdot M_0)}}{T_{cat}}$$

where, T is the observed catalog duration in years. The exceedance probability over the time horizon t is:

$$P(N \geq 1|t) = 1 - \exp(-\lambda * t)$$

Here, T_{cat} denotes the catalog duration in years used to estimate (a, b) , while t denotes the prediction horizon (e.g., 10, 25, or 50 years) used to compute exceedance probabilities.

This enables construction of probabilistic hazard surfaces for 10-, 25-, and 50-year horizons (see Section 4.2) [24].

The catalog duration T_{cat} is distinguished from the prediction horizon t ; parameters a, b are estimated from T_{cat} , while exceedance probabilities are computed for any t .

In this study, the GR parameters (a, b) and the associated annual exceedance rate λ for $M \geq 5.0$ are estimated from the full 1925–2025 catalog for each grid cell. These quantities are treated as long-term, stationary descriptors of local seismicity and are not re-estimated on short rolling windows, in order to avoid unstable and physically questionable parameter values. In the present implementation, earthquake occurrence in each grid cell follows a homogeneous Poisson process with constant annual rate λ , and that grid cells are independent of each other. Spatial correlation of ground motion and temporal clustering (e.g., aftershock sequences) across grids are not explicitly modeled. This simplifying assumption keeps the simulation tractable but may under-represent multi-grid event dependence in a tectonically complex setting such as Indonesia.

3.3 Machine learning modeling

On top of physics-based features (a, b, λ) computed per grid from the GR–Poisson hazard model, machine-learning models are trained to improve forward-looking risk quantification. The GR parameters (a, b) and the associated annual exceedance rate λ for $M \geq 5.0$ are estimated from the full 1925–2025 catalog for each grid cell and treated as long-term, stationary descriptors of local seismicity; they are not re-estimated on short rolling windows. The analysis considers a single prediction task:

- Risk-proxy classification: predict the probability of at least one exceedance within a 5-year horizon ($t = 5$ years) (i.e., $P(N \geq 1 | t)$) for a chosen magnitude threshold M_0 ,

typically $M_0 = 5.0$).

Candidate models are RFs [25, 26] and GB [25, 27]. Tree-based ensembles are selected because the available dataset is relatively small and tabular, and probability outputs that are reasonably well calibrated and interpretable are required. RF and GB are widely used benchmark models in catastrophe risk applications under such conditions. More complex approaches, such as deep neural networks, stacked ensembles, or fully Bayesian models, are promising directions but are beyond the scope of this first exploratory study and would require more extensive data and tuning. Models are evaluated via walk-forward cross-validation, training on historical intervals and testing on subsequent, non-overlapping intervals to avoid temporal leakage [28]. Unless stated otherwise, Area Under the Receiver Operating Characteristic Curve (AUROC) is reported to assess discrimination, and the Brier score is reported to evaluate probability accuracy; reliability diagrams are provided to assess calibration [29].

Implementation details: (i) features include (a, b, λ) and geographic context (latitude, longitude; optional depth summary), (ii) class imbalance is handled via class-weighting, (iii) horizons t and thresholds M_0 are kept fixed within each experiment, and (iv) a physics baseline (GR–Poisson) is reported for comparison. Full hyperparameter settings and split definitions are documented for reproducibility.

3.4 Parametric trigger design and premium calculation

Earthquake intensities predicted from the hazard model are mapped into payouts using a piecewise-linear trigger function defined by an attachment point (A) and an exhaustion point (E). In this study, the values are fixed at $A = 5.0$ and $E = 7.5$, meaning that payouts start once a magnitude $M \geq 5.0$ is exceeded, and reach full exhaustion at $M \geq 7.5$. The values $A = 5.0$ and $E = 7.5$ are chosen as illustrative thresholds rather than calibrated design points. In the Indonesian context, $M \approx 5.0$ is often regarded as a lower bound for potentially damaging shallow events, whereas $M \approx 7.5$ represents high-severity shaking where a full payout is intended. These values are not based on direct stakeholder input or explicit loss data; they are used here to demonstrate how a simple parametric rule can be combined with hazard and risk metrics. A systematic sensitivity analysis of A and E , and co-design of trigger levels with insurers and public stakeholders, is left for future work.

The analysis adopts a piecewise-linear payout mapping $g(M)$:

- if $M < A \rightarrow \text{payout} = 0$
- if $A \leq M < E \rightarrow \text{payout} = (M - A) / (E - A)$
- if $M \geq E \rightarrow \text{payout} = 1$

The expected premium indication (Π) is computed as: $\Pi = E[\text{Payout}] + \text{Expense Load} + \text{Risk Load}$, where, Π : premium indication, $E[\text{Payout}]$: expected payout (burning cost), Expense Load = allowance for operational/administrative expenses, Risk Load = additional margin to account for model and tail uncertainty.

The risk load is calibrated using TVaR at $\alpha = 0.95$, ensuring sensitivity to catastrophic outcomes beyond the value-at-risk (VaR) threshold [9, 17]. Specifically, if simulated payouts L_i are generated via Monte Carlo sampling conditional on hazard exceedance probability $P(N \geq 1|T)$, then:

$$\text{AAL} = \frac{1}{N} \sum_{i=1}^N L_i, \text{TVaR}_\alpha = E[L | L \geq \text{VaR}_\alpha(L)]$$

and the premium is adjusted as: $\Pi = \text{AAL} + (\text{TVaR}_\alpha -$

$\text{AAL}) + \text{Expense Load}$.

The yearly payout is defined as $L = g(M_{\max})$, where M_{\max} denotes the maximum magnitude observed within horizon t for a given grid. This construction ensures that the burning cost (AAL) reflects average loss, while the TVaR risk margin captures tail dependence under rare but severe earthquake scenarios [9, 17]. Monte Carlo simulation (10,000 synthetic years per grid) is used to estimate the payout distribution, from which AAL, TVaR, and Premiums are derived. These values are then aggregated at the grid or provincial level to inform potential insurance contract pricing (see Section 4.4).

In the present implementation, the Monte Carlo simulation for premiums is driven by the long-term GR–Poisson rate λ estimated for each grid cell from the 1925–2025 catalog. These rates are used to simulate annual event counts and corresponding losses, from which AAL, TVaR, and indicative premiums are obtained. The machine-learning models provide calibrated exceedance probabilities for 5-year horizons in parallel, but their predictions are not yet directly used as inputs to the pricing engine.

3.5 Model evaluation

For reproducibility, Monte Carlo uses $N = 10,000$ synthetic years per grid with fixed random seeds; walk-forward evaluation specifies the number and boundaries of temporal splits, and all code/configuration files are available upon request.

The framework is evaluated using both probabilistic and financial metrics:

- a. Probabilistic metrics: Brier Score, AUROC, and reliability diagrams (calibration curves).
- b. Financial metrics: AAL and TVaR.

To ensure robustness, two validation schemes are employed:

- Walk-forward backtesting: models are trained on earlier historical intervals and tested on subsequent, non-overlapping intervals to avoid temporal leakage.
- Spatial holdout validation: certain regions are excluded during training to assess geographic generalization.

4. RESULTS AND DISCUSSION

This section presents the results of the proposed framework and their interpretation. The findings are structured into six parts. First, the outcomes of data cleaning and spatial aggregation are described (Section 4.1). Second, seismic hazard quantification based on GR and Poisson models is summarized (Section 4.2). Third, the performance of ML models in predicting exceedance probabilities is reported (Section 4.3). Fourth, parametric trigger simulations and premium estimations are provided (Section 4.4). Fifth, probabilistic and financial evaluations are integrated to assess model robustness (Section 4.5). Finally, broader implications, limitations, and policy relevance are discussed (Section 4.6). Together, these results demonstrate the feasibility of linking seismological modeling, ML, and actuarial methods into a unified framework for earthquake risk financing in Indonesia.

4.1 Data preprocessing and spatial aggregation

The cleaned and standardized USGS earthquake catalog (1925–2025) contained 71,340 events, of which 14,466 were

$M \geq 5.0$. After deduplication, imputation, and consistency checks, the data were aggregated into $0.5^\circ \times 0.5^\circ$ spatial grids. A total of 2,193 active grids were identified, with 711 grids exhibiting non-zero annual exceedance rates (λ). These statistics are summarized in Table 1, which also highlights the average b-value and exceedance rate. This preprocessing step ensures a consistent dataset for hazard quantification and model training.

Table 1. Earthquake statistics per $0.5^\circ \times 0.5^\circ$ grid in Indonesia (1925–2025), including number of events, active grid cells, average b-value, and mean annual exceedance rate (λ)

Statistic	Value
Total events	71,340
Events with $M \geq 5.0$	14,466
Active grid cells	2,193
Grids with $\lambda > 0$	711
Average b-value	1.15
Average λ (per grid, $M \geq 5.0$)	0.058 events/year

4.2 Seismic hazard quantification

Applying the GR relation per grid yielded average b-values ≈ 1.15 , consistent with the global mean for tectonically active regions. The mean annual exceedance rate was $\lambda = 0.058$ events/year per grid for $M \geq 5.0$. The exceedance probabilities computed for three horizons are summarized in Figure 1:

- 10-year horizon: exceedance probabilities reached $P \geq 0.5$ along the western coast of Sumatra, Maluku, and Sulawesi.
- 25-year horizon: probabilities increased, with most of Sumatra showing $P \geq 0.7$.
- 50-year horizon: exceedance probabilities approached 0.9–0.95 along the Sunda Megathrust, highlighting extreme long-term hazard.

These findings confirm the high seismic productivity of Sumatra and eastern Indonesian arcs, consistent with BMKG hazard maps and prior PSHA studies [8–10].

4.3 Machine learning risk modelling

A 5-year walk-forward validation was implemented to assess the short-term predictive capability of the GR parameters (a , b , λ). As shown in Table 2, the RF model achieved an AUROC of 0.604 and a Brier score of 0.228, indicating fair discrimination and well-calibrated probability estimates. The GB model obtained a slightly higher discrimination (AUROC = 0.629) but exhibited mild overconfidence in high-probability bins (Brier = 0.272). Both ML models significantly outperformed the GR–Poisson baseline (AUROC = 0.553; Brier = 0.345), confirming that the λ parameter—computed from historical 5-year windows—remains informative yet not over-dominant.

These results demonstrate that combining physical features with ML improves reliability in short-term seismic risk prediction and maintains consistency with long-term exceedance probability maps (10-, 25-, and 50-year horizons).

For the GR–Poisson baseline in Table 2, the long-term GR parameters (a , b) and the corresponding annual exceedance rate λ for $M \geq 5.0$ are estimated from the full 1925–2025 catalog for each grid cell, and the 5-year exceedance probability used for comparison with the ML models is computed as $P(N \geq 1 | 5 \text{ years}) = 1 - \exp(-5\lambda)$.

These AUROC values of about 0.6 indicate only modest discrimination between grids that do and do not experience at least one $M \geq 5.0$ event in the next five years. This level of performance is, however, consistent with the inherent difficulty of the task, which involves rare events and a relatively coarse $0.5^\circ \times 0.5^\circ$ spatial grid. In this context, the main advantage of the machine-learning models lies in their improved probability calibration rather than in their classification power: they achieve lower Brier scores than the GR–Poisson baseline and produce reliability curves that are closer to the diagonal, indicating that predicted probabilities better match observed frequencies. Given the strong class imbalance inherent in rare-event seismic prediction, AUROC is interpreted cautiously and treated as a secondary metric, while primary reliance is placed on Brier scores and reliability diagrams to assess model quality; precision–recall analysis and log-loss are acknowledged as useful complementary diagnostics for future extensions of this work.

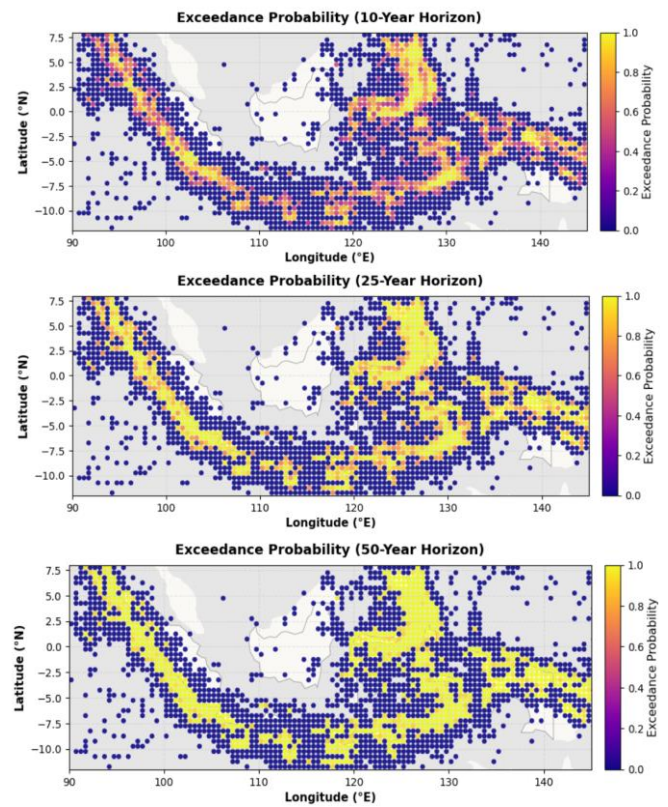


Figure 1. Exceedance probability maps for $M \geq 5.0$ earthquakes

Table 2. Machine learning model performance under 5-year walk-forward validation

Model	Area Under the Receiver Operating Characteristic Curve (AUROC)	Brier Score	Calibration
Random Forests (RFs)	0.604	0.228	Well calibrated
Gradient Boosting (GB)	0.629	0.272	Slight overconfidence
Gutenberg–Richter (baseline)	0.553	0.345	Under-dispersed

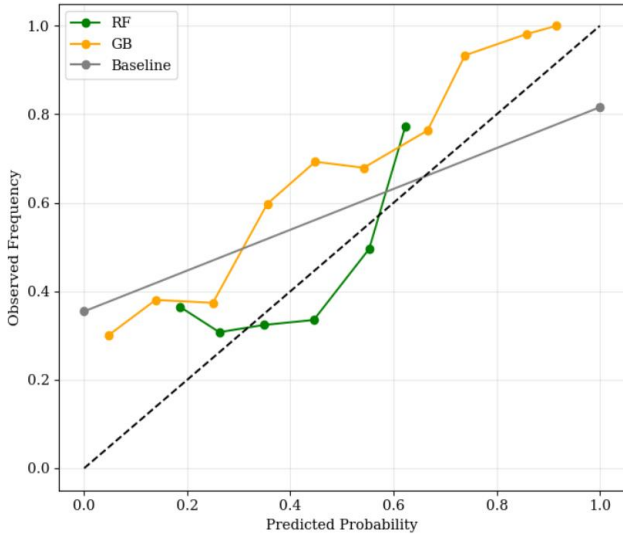


Figure 2. Reliability calibration curves of machine-learning models (5-year horizon)

Figure 2 presents the reliability calibration curves of the ML models developed in this study. These curves compare the predicted probabilities of earthquake occurrence with the observed frequencies across spatial grids. By examining the deviation of each model’s calibration line from the diagonal “perfect reliability” reference, the consistency between predicted and actual seismic probabilities can be assessed. This analysis complements the quantitative performance metrics summarized in Table 2.

Reliability diagrams comparing predicted and observed frequencies for RF, GB, and GR–Poisson (baseline) models

under 5-year walk-forward validation. The RF curve is closest to the diagonal, indicating the best probability calibration. The GB model tends to slightly overestimate high-probability bins, whereas the GR–Poisson baseline remains under-dispersed. As shown in Figure 2, the RF model demonstrates relatively stable and well-calibrated probabilities across the 5-year horizon, confirming its ability to capture nonlinear relationships between GR parameters (a , b , λ) and spatial coordinates. The GB model yields marginally higher discrimination (AUROC = 0.629) but displays a tendency toward overconfidence for higher-probability bins. The GR–Poisson baseline underestimates variability, resulting in under-dispersed probabilities. Together, these results confirm that integrating physical parameters with machine-learning methods enhances short-term earthquake risk estimation and aligns with the exceedance probability trends observed in long-term (10-, 25-, and 50-year) hazard maps [14, 15].

4.4 Parametric trigger simulation and premium estimation

Using a piecewise-linear trigger (Attachment A = 5.0, Exhaustion E = 7.5), a Monte Carlo simulation [9] of 10,000 synthetic years per grid was conducted to estimate the distribution of earthquake-induced losses and corresponding parametric insurance premiums. The simulation covered 711 spatial grids across Indonesia, using the GR-derived parameters a , b , and λ as inputs for stochastic generation of maximum yearly magnitudes M_{max} .

Each simulated year produced a payout $L = g(M_{max})$ according to the trigger function, from which the AAL, TVaR₉₅, and premium were derived. Statistical summaries are presented in Table 3.

Table 3. Financial risk metrics from Monte Carlo simulation (10,000 synthetic years per grid)

Metric	Range	Mean	25th Percentile	50th Percentile (Median)	75th Percentile	Provincial Hotspots
Average Annual Loss (AAL)	0.00–0.23	0.03	0.012	0.019	0.035	Aceh, West Sumatra, Sulawesi, Maluku
TVaR ₉₅	0.00–0.84	0.32	0.210	0.314	0.437	Sumatra subduction corridor Aceh, West Sumatra, Bengkulu, Maluku
Premium	0.00–0.84	0.32	0.210	0.314	0.437	

- AAL values were generally low (mean ≈ 0.03), consistent with the low annual probability of high-magnitude exceedance, but reached > 0.20 in Aceh, West Sumatra, Sulawesi, and Maluku—indicating elevated background hazard rates.
- TVaR₉₅ exhibited a markedly heavier tail, extending up to 0.84 in high-risk grids along the Sumatra subduction corridor, confirming that extreme but infrequent events dominate the loss potential.
- Premiums followed a similar right-skewed distribution, with typical values between 0.20 and 0.40 (interquartile range), but rising above 0.80 in the highest-risk provinces such as Aceh, West Sumatra, and Bengkulu.

It is important to note that Table 3 summarizes variability across grid cells (ranges and quartiles) rather than Monte Carlo sampling uncertainty for individual grids. For a given grid g , the standard error of the AAL estimate \widehat{AAL}_g scales as $SE(\widehat{AAL}_g) \approx s_g/\sqrt{N}$, where s_g is the sample standard deviation of yearly payouts and $N = 10,000$ synthetic years. In our experiments, these simulation errors are typically small

relative to the spatial heterogeneity summarized in Table 3, therefore, point estimates and cross-grid quantiles are reported. A more detailed treatment of parameter uncertainty—e.g., through bootstrap confidence intervals or Bayesian credible intervals for AAL, TVaR₉₅, and premiums is left to future work.

The mean premium ≈ 0.32 implies that on average, the expected risk cost—including the tail-risk adjustment (TVaR–AAL)—represents about one-third of the normalized maximum payout. This pattern reflects the amplification of tail risk in subduction and transform fault zones where event clustering and high-magnitude recurrence dominate the stochastic hazard structure.

4.5 Integrated evaluation

The framework was evaluated under both probabilistic and financial criteria to ensure robust performance across predictive and actuarial dimensions. As shown in Figure 3, the histogram in panel (a) reveals that most grids have premiums concentrated between 0.2 and 0.4, while a few high-risk

regions exceed 0.8, particularly along the Sumatra and Sulawesi subduction zones. The boxplot in panel (b) confirms a right-skewed distribution with several extreme outliers, emphasizing the non-uniformity of risk exposure across Indonesia. Panel (c) displays a strong positive correlation between AAL and TVaR₉₅, where increases in mean losses are accompanied by disproportionately higher tail losses, a phenomenon known as tail-risk amplification. This pattern underscores that while the AAL remains small for most areas, catastrophic exceedance potential drives significant premium escalation in high-risk provinces such as Aceh, West Sumatra, Bengkulu, and Maluku.

- Probabilistic: The RFs and GB models achieved AUROC values between 0.60 and 0.63 and Brier scores between 0.23 and 0.27, indicating moderate discrimination and well-calibrated probability estimation relative to the GR–Poisson baseline (AUROC = 0.55; Brier = 0.35). The reliability curves

confirmed that RF was well calibrated, while GB showed slight overconfidence at higher predicted probabilities.

- Financial: The distribution of premiums derived from 10,000-year Monte Carlo simulations (Figure 3) exhibits heavy-tailed behaviour, with extreme values concentrated in high-risk provinces such as Aceh, West Sumatra, Bengkulu, Sulawesi, and Maluku. The scatter plot of AAL versus TVaR₉₅ further shows that regions with higher mean AAL also experience disproportionately larger tail losses, highlighting the need for tail-risk adjustments in actuarial pricing.

The temporal (walk-forward) and spatial (holdout) validation together indicate that the framework generalizes well to unseen time intervals and geographic regions. Such robustness is crucial for actuarial applications, ensuring that parametric pricing remains stable and not overly sensitive to catalog artefacts or regional peculiarities.

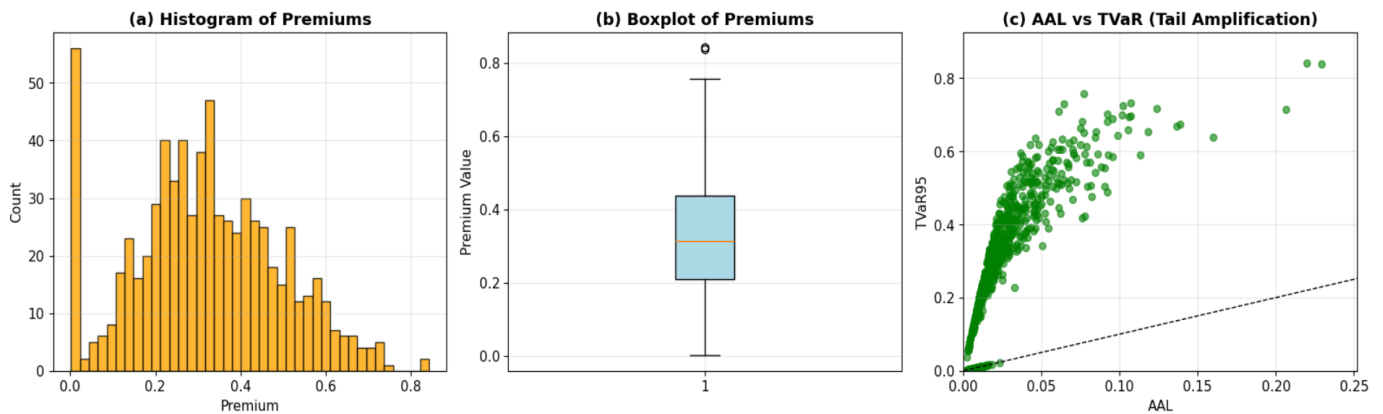


Figure 3. Distribution of premiums under the parametric trigger ($A = 5.0$, $E = 7.5$): (a) Histogram, (b) Boxplot of premiums across all grids confirms heavy-tailed distribution and skewness toward high-risk regions, (c) The AAL–TVaR₉₅ scatter plot illustrates non-linear tail amplification: Even moderate AAL levels can correspond to extremely high TVaR values in subduction provinces

4.6 Discussion

From a probabilistic standpoint, the machine-learning component of the framework provided added value primarily through improved calibration rather than strong discrimination. The AUROC values around 0.6 reflect the inherent difficulty of rare-event forecasting at a coarse $0.5^\circ \times 0.5^\circ$ resolution, where event sparsity and spatial–temporal heterogeneity limit separability between “event” and “no-event” grids. For disaster-risk financing, however, well-calibrated probabilities are often more critical than sharp classification: risk measures such as expected loss and tail metrics depend directly on the accuracy of the probability scale. The fact that the tree-based models yield lower Brier scores and more reliable probability estimates than the GR–Poisson baseline suggests that they can serve as calibrated, physics-informed predictors of exceedance risk, even when their AUROC remains modest.

From a financial perspective, the integration of probabilistic hazard modelling and Monte Carlo-based premium estimation highlights how tail-risk amplification and spatial concentration of risk translate into premium differentiation. The heavy-tailed premium distribution and the strong AAL–TVaR₉₅ relationship imply that small increases in mean loss can be associated with disproportionately large tail losses in subduction corridors such as Aceh, West Sumatra, and

Bengkulu. This pattern is consistent with the principles of catastrophe insurance and risk layering: relatively modest base-layer coverage can be complemented by higher-layer or regional risk-transfer instruments to address extreme, low-probability events. In this sense, the parametric trigger used here is best interpreted as a stylised illustration of how hazard-based indices can inform the pricing of such layers, rather than as a final product design.

At this stage, the actuarial module remains driven by long-term GR–Poisson rates, and the ML probabilities are analysed alongside rather than fully embedded in the premium calculations. This creates a partial disconnect between Sections 4.3 and 4.4, which is regarded as a deliberate first step: the current framework demonstrates that seismic models, ML forecasts, and actuarial simulation can be combined in a consistent workflow, while the full integration of ML-adjusted probabilities into the pricing engine is left for future work. Possible extensions include scenario-dependent adjustments of λ , blended GR–ML exceedance probabilities that preserve physical interpretability, and hierarchical models that propagate predictive uncertainty into premiums.

A further limitation is the assumption of independent Poisson occurrence per grid cell, with no explicit representation of spatially correlated ground motion or fault-based clustering, which is particularly important in Indonesia’s segmented subduction and strike-slip fault

systems. The coarse spatial resolution ($0.5^\circ \times 0.5^\circ$) also limits the ability to resolve near-fault effects and basin amplification. Nonetheless, the alignment between AAL and TVaR patterns, and the consistency of the ML probabilities with long-term exceedance maps, indicate that the proposed framework captures key features of Indonesia's seismic risk while remaining computationally tractable. Future work should explore fault-based or spatially correlated formulations, richer exposure models, and co-designed trigger levels with stakeholders to move from illustrative to implementable parametric products.

5. CONCLUSION

This study proposed an integrated framework that combines probabilistic seismic hazard analysis, simple machine-learning-based probability forecasting, and Monte Carlo simulation of parametric payouts to study earthquake risk and disaster financing in Indonesia. In its present form, the actuarial module remains driven by long-term GR–Poisson rates, while the machine-learning probabilities are analysed in parallel as a calibrated short- to medium-term view of exceedance risk rather than being used directly inside the pricing engine. In out-of-sample experiments, the RFs and GB models achieved only modest discrimination (AUROC \approx 0.6) but provided better-calibrated exceedance probabilities (Brier scores \approx 0.23–0.27) than the GR–Poisson baseline. Thus, within this framework, their main contribution lies in improved probability calibration rather than strong classification power.

From a financial perspective, the parametric trigger simulation ($A = 5.0$, $E = 7.5$) captured the heavy-tailed nature of earthquake loss, with AAL values typically below 0.1 but TVaR₉₅ reaching very high levels (up to about 0.84) in high-risk provinces such as Aceh, West Sumatra, and Bengkulu. These patterns underscore that catastrophic tail risk, rather than average loss, drives premium escalation in a small number of subduction-zone regions. The results also show how a simple, transparent parametric rule, when combined with hazard-based indices, can generate spatially differentiated indicative premiums as a basis for earthquake disaster-risk financing in Indonesia.

Several extensions are needed to move from this proof-of-concept to operational applications. First, ML-adjusted probabilities or blended GR–ML exceedance rates should be embedded directly into the pricing engine, and their impact on AAL, TVaR, and premiums should be quantified. Second, the independent Poisson assumption at the grid level should be relaxed by incorporating fault-based or spatially correlated hazard models that better capture multi-grid event dependence in Indonesia's complex tectonic regime. Third, trigger levels and payout structures should be co-designed with stakeholders using exposure and loss data, with an eye toward concrete instruments such as national contingency funds, provincial parametric covers, or risk-pooling schemes. These steps would help translate the proposed framework into practical tools for earthquake disaster-risk financing in Indonesia and similar high-risk settings.

ACKNOWLEDGMENT

The authors acknowledge the valuable contributions of

colleagues and supervisors whose constructive feedback greatly improved the methodology and interpretation of the research findings.

REFERENCES

- [1] Wu, M.H., Wang, J.P., Sung, C.Y. (2023). Earthquake magnitude probability and gamma distribution. *International Journal of Geosciences*, 14(7): 585-602. <https://doi.org/10.4236/ijg.2023.147031>
- [2] Natawidjaja, D.H., Triyoso, W. (2007). The sumatran fault zone — From source to hazard. *Journal of Earthquake and Tsunami*, 1(1): 21-47. <https://doi.org/10.1142/S1793431107000031>
- [3] Prastowo, T., Ayudia, G.P., Risanti, H. (2022). Magnitude-rupture area scaling derived from global earthquakes of moderate to large sizes: Implications for seismic hazards in Indonesia. *Bulletin of the Geological Society of Malaysia*, 73(1): 1-12. <https://doi.org/10.7186/bgsm73202201>
- [4] Bilek, S.L., Lay, T. (2018). Subduction zone megathrust earthquakes. *Geosphere*, 14(4): 1468-1500. <https://doi.org/10.1130/GES01608.1>
- [5] Motaghed, S., Khazae, M., Eftekhari, N., Mohammadi, M. (2023). A non-extensive approach to probabilistic seismic hazard analysis. *Natural Hazards and Earth System Sciences*, 23(3): 1117-1124. <https://doi.org/10.5194/nhess-23-1117-2023>
- [6] Shrestha, N. (2019). Estimating the probability of earthquake occurrence and return period using generalized linear models. *Journal of Geoscience and Environment Protection*, 7(9): 11-24. <https://doi.org/10.4236/gep.2019.79002>
- [7] Gerstenberger, M.C., Marzocchi, W., Allen, T., Pagani, M., et al. (2020). Probabilistic seismic hazard analysis at regional and national scales: State of the art and future challenges. *Reviews of Geophysics*, 58(2): 1-49. <https://doi.org/10.1029/2019RG000653>
- [8] GEM. (2023). Global seismic hazard map. <https://www.globalquakemodel.org/product/global-seismic-hazard-map>.
- [9] Mistry, H.K., Lombardi, D. (2023). A stochastic exposure model for seismic risk assessment and pricing of catastrophe bonds. *Natural Hazards*, 117: 803-829. <https://doi.org/10.1007/s11069-023-05884-4>
- [10] McGuire, R.K. (2004). Seismic hazard and risk analysis engineering monographs on miscellaneous earthquake engineering topics. Earthquake Engineering Research Institute, University of California.
- [11] Mousavi, S.M., Ellsworth, W.L., Zhu, W.Q., Chuang, L.Y., Beroza, G.C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11: 3952. <https://doi.org/10.1038/s41467-020-17591-w>
- [12] Cornell, C.A. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, 58(5): 1583-1606. <https://doi.org/10.1785/BSSA0580051583>
- [13] Gvishiani, A.D., Dzeranov, B.V., Skorkina, A.A., Dzeboev, B.A. (2024). World seismic networks and earthquake catalogs. *Russian Journal of Earth Sciences*, 24(1). <https://doi.org/10.2205/2024ES000901>
- [14] Grigoratos, I., Poggi, V., Danciu, L., Monteiro, R.

- (2023). Homogenizing instrumental earthquake catalogs – A case study around the Dead Sea Transform Fault Zone. *Seismica*, 2(2). <https://doi.org/10.26443/seismica.v2i2.402>
- [15] DeVries, P.M.R., Viégas, F., Wattenberg, M., Meade, B.J. (2018). Deep learning of aftershock patterns following large earthquakes. *Nature*, 560: 632-634. <https://doi.org/10.1038/s41586-018-0438-y>
- [16] Duro, E. (2023). A multi-criteria methodology for the integration of risk assessment into spatial planning as a basis for territorial resilience. The case of seismic risk. *FORUM A+P Interdisciplinary Journal of Architecture and Built Environment*, (26): 54-60. <https://doi.org/10.37199/f40002607>
- [17] OECD. (2015). *Disaster Risk Financing: A Global Survey of Practices and Challenges*. OECD Publishing. <https://doi.org/10.1787/9789264234246-en>
- [18] Akinci, A. (2010). Hazgridx: Earthquake forecasting model for $M_L \geq 5.0$ earthquakes in Italy based on spatially smoothed seismicity. *Annals of Geophysics*, 53(3): 51-61. <https://doi.org/10.4401/ag-4811>
- [19] Hutchings, S.J., Mooney, W.D. (2021). The seismicity of Indonesia and tectonic implications. *Geochemistry, Geophysics, Geosystems*, 22(9): 1-42. <https://doi.org/10.1029/2021GC009812>
- [20] Cardona, O.D., Ordaz, M.G., Yamín, L.E., Arámbula, S., Marulanda, M.C., Barbat, A.H. (2008). Probabilistic seismic risk assessment for comprehensive risk management: Modeling for innovative risk transfer and loss financing mechanisms. In *The 14th World Conference on Earthquake Engineering*, Beijing, China.
- [21] Ali, J.R. (2013). The SE Asian gateway: History and tectonics of the Australia–Asia collision. *Marine Geophysical Research*, 34: 147-149. <https://doi.org/10.1007/s11001-013-9178-4>
- [22] Petersen, M.D., Mueller, C.S., Frankel, A.D., Zeng, Y. (2008). Spatial seismicity rates and maximum magnitudes for background earthquakes. USGS Open-File Report. <https://doi.org/10.3133/ofr20071437J>
- [23] Iwata, D., Nanjo, K.Z. (2024). Adaptive estimation of the Gutenberg–Richter b value using a state space model and particle filtering. *Scientific Reports*, 14: 4630. <https://doi.org/10.1038/s41598-024-54576-x>
- [24] Hartati, H., Effendie, A.R., Susyanto, N., Suryanto, W. (2026). Probabilistic earthquake hazard assessment in Indonesia using Poisson model and spatial grid analysis. *Science and Technology Indonesia*, 11(1): 207-216. <https://doi.org/10.26554/sti.2026.11.1.207-216>
- [25] Ajai, V., Usha, S.G.A., Suntosh, B.D.S., Muthukumar, M., Raj, K.M., Suriyanarayanan, V. (2024). Machine learning-based seismic activity prediction. In *Utilizing AI and Machine Learning for Natural Disaster Management*, pp. 293-306. <https://doi.org/10.4018/979-8-3693-3362-4.ch017>
- [26] Ng'ombe, J.N., Addai, K.N., Mzyece, A., Han, J., Temoso, O. (2023). Uncovering the factors that affect earthquake insurance uptake using supervised machine learning. *Scientific Reports*, 13: 21314. <https://doi.org/10.1038/s41598-023-48568-6>
- [27] Usman, F., Nanda., Sumantyo, J.T.S. (2022). Prediction of ground surface deformation induced by earthquake on urban area using machine learning. *Science and Technology Indonesia*, 7(4): 435-442. <https://doi.org/10.26554/sti.2022.7.4.435-442>
- [28] Karakulath, S. (2024). Comparing time series forecasting models and validation methods for non-transparent markets (wood waste). Master Thesis of Science in Data Science. Constructor (Jacobs) University Bremen. <https://doi.org/10.13140/RG.2.2.20867.11047>
- [29] Siebert, S. (2016). Simplifying and generalising Murphy's Brier score decomposition. *Quarterly Journal of the Royal Meteorological Society*, 143(703): 1178-1183. <https://doi.org/10.1002/qj.2985>