

Socio-economic value of data-driven eruption forecasts to balance false alarms against catastrophic loss

Received: 10 January 2026

Accepted: 13 August 2026

Published online: 26 August 2026

 Check for updatesAlberto Ardid¹✉, David Dempsey¹, Garry McDonald² & Shane Cronin³

The adoption of new forecasting methods for volcanic eruptions typically emphasizes predictive accuracy, often overlooking their potential to reduce overall life and economic losses. We introduce a Potential Economic Value (PEV) framework that evaluates forecast utility by balancing precautionary actions against avoidable catastrophic losses (such as mass casualties). Using machine-learning forecasts from continuous seismic data at five volcanoes, we show that non-forecasted eruptions (missed) have disproportionate consequences, compared to false alarms, which generate recurring and manageable disruption. Retrospective analyses of the 2019 Whakaari (New Zealand) and 2014 Ontake (Japan) eruptions, along with three additional volcanoes, indicate that losses could have been reduced by 30–90%, despite generating numerous false alarms. Across all case studies, effective forecasting prioritizes reducing missed eruptions over maximizing accuracy. This supports the use of more precautionary warning thresholds, provided that the resulting increase in alert frequency is managed in ways that maintain public compliance and trust.

Volcanic eruptions are among the most destructive natural hazards on Earth, capable of causing widespread disruption, loss of life, and long-term socioeconomic impacts. Globally, 50–70 eruptions occur each year, with a minority causing disproportionate losses¹. Recent events span a wide spectrum of local to global consequences: the 2010 Merapi eruption in Indonesia killed more than 350 people and forced the evacuation of 400,000, many of whom were permanently displaced². Other large eruptions such as Eyjafjallajökull in 2010, Puyehue–Cordón Caulle in 2011, Calbuco in 2015, La Palma in 2021, and Hunga Tonga–Hunga Ha’apai in 2022, primarily produced widespread infrastructure damage, aviation disruption, and economic losses, with the latter also causing seven fatalities and many injuries^{3–9}. Smaller eruptions can also produce severe human impacts despite limited physical damage. The Ontake 2014 eruption (63 fatalities¹⁰) and the 2019 Whakaari eruption (22 fatalities and dozens of severe injuries¹¹) exemplify how short-warning, exposure-driven events can be disproportionately deadly, raising the question of where the true value of

forecasting systems lies. If most infrastructure damage, aviation disruption, and economic losses are largely unavoidable, does the primary value of forecasting lie in preventing loss of life?

Despite extended advances in monitoring over the last decades, anticipating eruptions remains challenging. Seismic, geodetic, and gas precursors often herald unrest^{12–14}, yet forecasts remain uncertain because genuine precursor signals vary across volcanoes and are difficult to identify and evaluate robustly across eruptive and non-eruptive periods, particularly under out-of-sample conditions^{15–17}. Also, some eruptions occur with subtle or absent warnings (Ontake 2014¹⁰; Whakaari 2016, 2019^{11,18}), and most volcanoes lack sufficient recorded events to identify reliable patterns¹⁹.

Recent data-driven approaches have improved the detection of subtle patterns in monitoring data, producing likelihood of eruption forecasts that can be evaluated under retrospective conditions^{20–24}. While some models show encouraging skill, most studies focus on pre-eruptive intervals without comparing against non-eruptive

¹University of Canterbury, Christchurch, New Zealand. ²Market Economics, Auckland, New Zealand. ³University of Auckland, Auckland, New Zealand.

✉ e-mail: alberto.ardid@canterbury.ac.nz

repose^{17,25,26}. Plus, objective metrics that account for missed eruptions and false alarms over long periods, such as ROC curves and AUC scores, are necessary to evaluate forecast reliability^{19,27–29}.

Alongside accuracy, it is important to consider the socio-economic costs and benefits of forecasting systems. False alarms may trigger unnecessary evacuations, closures of businesses and transport networks, legal disputes, and erosion of public trust^{30,31}. Equally, missed eruptions for which evacuation would have been feasible incur financial costs associated with loss of life, injury, and recovery. Hence, forecast value can be framed more generally than predictive skill alone, extended to include its financial performance through mitigated socioeconomic costs. Here, we address this gap by introducing a quantitative framework that explicitly balances the short-term costs of precautionary action against the avoided losses from catastrophic events. We restrict our attention to costs arising from short-term economic disruption and associated loss-modeling components only, excluding longer-term equilibrium effects^{32–35}.

Evaluating a forecast's socioeconomic value requires estimating costs that are inherently uncertain, either because they are difficult to measure or they vary considerably with context. International benchmarks for human life valuation vary widely (e.g., New Zealand: USD 2.7M³⁶; United States: USD 12.5 M per fatality³⁷), and communities differ in acceptable disruption levels and risk tolerance^{15,38,39}. Both false alarms and missed eruptions carry psychological costs. Both can result in a loss of institutional trust, while repeated alerts may erode compliance^{40,41} and missed events cause trauma^{31,41}. However, the consequences can be asymmetric: false alarms generate recurring but manageable costs, whereas missed eruptions can lead to substantial loss of life^{42,43}. Despite the clear need for eruption forecast evaluation that incorporates socio-economic consequences, no standardized framework exists for comparing forecast value across volcanoes with differing hazard profiles and exposure contexts.

Here, we introduce the Potential Economic Value (PEV) framework to volcanology as a generalizable and interpretable metric, where generalizability refers to the ability to apply the same dimensionless cost–loss formulation across different volcanoes. Originally developed for meteorology and aviation^{44,45} and now applied in flood, drought, and wildfire risk management^{46–49}, PEV balances the short-term costs of precautionary actions (false alarms) against the avoided losses when forecasts correctly anticipate eruptions, yielding a single dimensionless measure of forecast utility. We demonstrate the PEV framework on multi-year eruption forecasts derived from a machine-learning model¹⁹ at five volcanoes that collectively span a spectrum of eruption

frequencies, monitoring histories, and socio-economic settings. Two cases, Whakaari (New Zealand) and Ontake (Japan), provide documented human and economic losses, enabling retrospective evaluation of the framework. Three additional applications (Ruapehu (New Zealand), Copahue (Chile/Argentina), Bezmyanny (Russia)) serve as stress tests, illustrating PEV's generalizability to hypothetical high-impact scenarios and data-limited systems.

Results

Framework evaluation: retrospective cases at Whakaari and Ontake

We begin with two cases where documented casualties and economic losses enable retrospective evaluation of the PEV framework.

For the forecast model applied to Whakaari¹⁹, forecast value (PEV) is maximized at intermediate decision thresholds that balance false alarms against missed eruptions (Fig. 1). At low thresholds ($\tau < 0.5$), frequent false alarms dominate expected costs, whereas at high thresholds ($\tau > 0.9$), missed eruptions increasingly drive losses, causing PEV to decline. This trade-off produces a characteristic hump-shaped PEV curve (Fig. 1a), a well-established feature of cost–loss decision systems⁴⁵. Monte Carlo ensemble sampling suggests that while absolute PEV values shift under different cost assumptions, the location of the optimum remains relatively stable near $\tau \approx 0.85–0.90$. At this threshold, forecasts avoid up to ~USD 220 M in losses over the forecast period (Fig. 1b).

In operational terms, the decision threshold τ is the forecast eruption probability above which protective action is initiated: when the model output exceeds τ , the framework recommends precautionary measures such as access restrictions, evacuation of exposed populations, or heightened monitoring posture. A lower τ produces more alerts (including more false alarms), while a higher τ produces fewer alerts but risks missing eruptions preceded by weaker precursory signals.

A retrospective analysis of a 12-year monitoring period at Whakaari shows that the particular forecast evaluated here could have reduced cumulative costs by >30% relative to a no-forecasting baseline in which protective actions are never taken (Fig. 2a). This result is obtained using a decision threshold of $\tau = 0.6$, which is lower than the value-maximizing threshold identified in the preceding PEV analysis ($\tau \approx 0.85–0.90$), and therefore represents a deliberately precautionary choice that favors reducing missed eruptions at the expense of more frequent false alarms. As a result, the reported cost reduction should

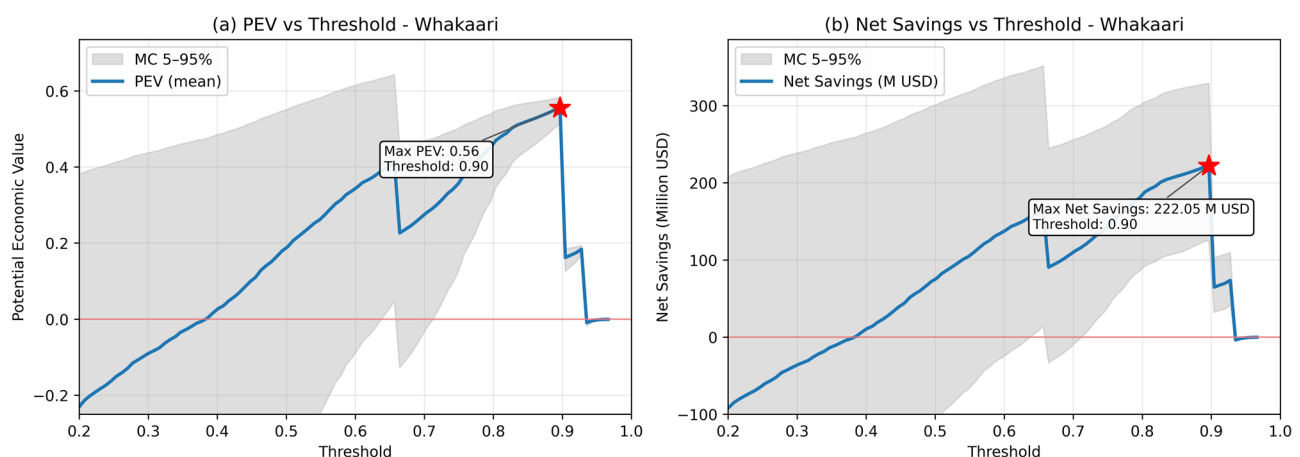


Fig. 1 | Threshold optimization for Whakaari volcano eruption forecasting using tailored machine learning models. a Potential Economic Value (PEV) versus forecast threshold, showing the trade-off between detection sensitivity and economic efficiency. The gray shaded area represents Monte Carlo uncertainty (5th–

95th percentile) around the mean PEV. **b** Net economic savings versus threshold, quantifying the monetary benefits of forecasting in millions of USD. Both metrics peak at threshold $\tau = 0.89$, indicating optimal forecast sensitivity that balances early warning benefits against false alarm costs.

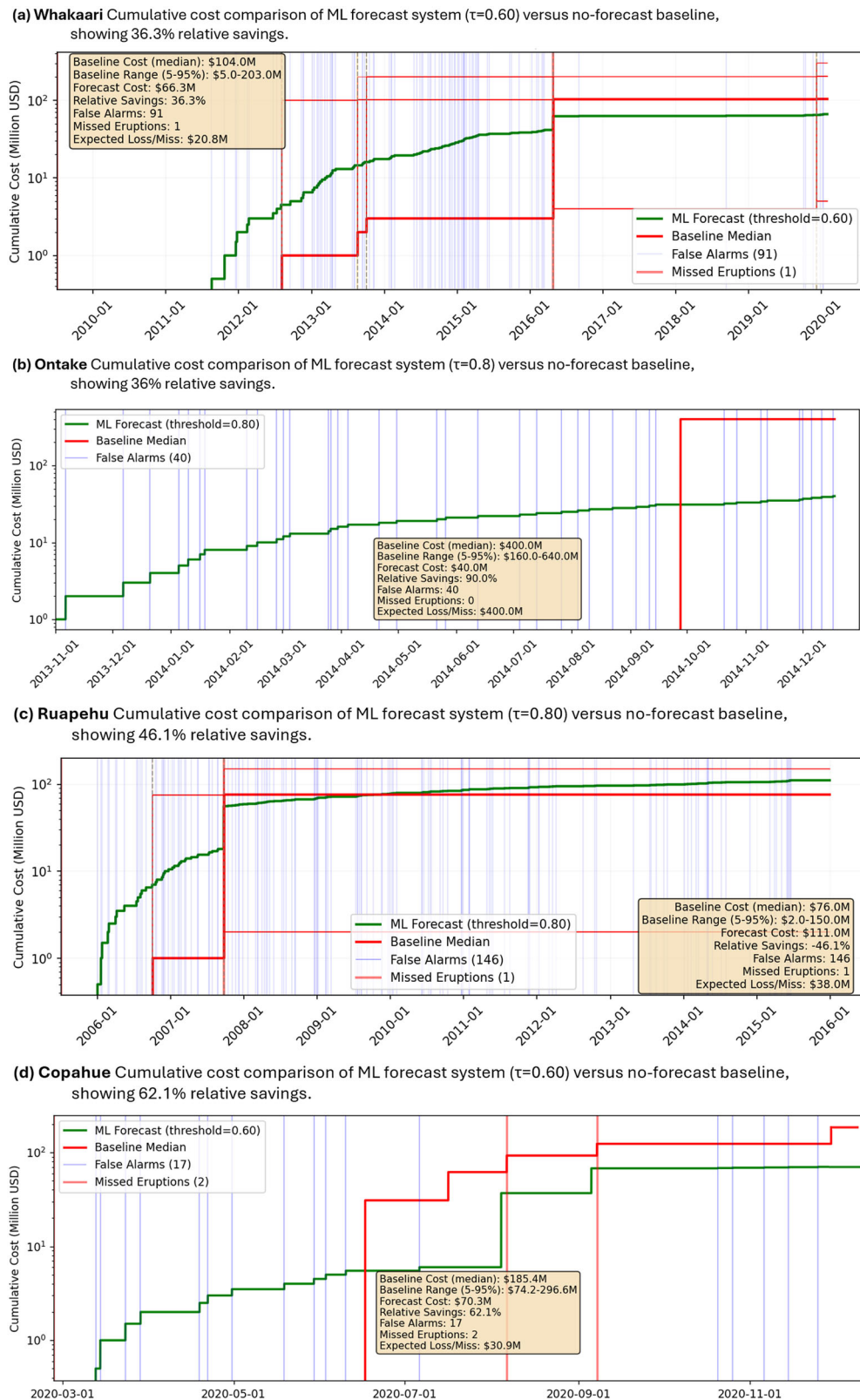


Fig. 2 | Cumulative cost comparison of ML forecast systems versus no-forecast baselines for Whakaari, Ontake, Ruapehu and Copahue volcanoes. Green lines show cumulative costs with ML forecasts at optimal thresholds; red lines show baseline median costs (no forecast). Vertical markers indicate false alarms (blue) and missed eruptions (red dashed). Summary statistics in each panel quantify economic performance: **a** Whakaari (2010–2020, $\tau = 0.60$) achieved 36.3% relative

savings (tested in 5 recorded eruption); **b** Ontake (2013–2014, $\tau = 0.80$) achieved 90.0% relative savings with zero missed eruptions (only tested in one recorded eruption); **c** Ruapehu (2006–2016, $\tau = 0.80$) resulted in 46.1% relative savings due to frequent false alarms (tested in two recorded eruption). **d** Copahue (during 2020, $\tau = 0.60$) resulted in 40.1% relative savings due to frequent false alarms (tested in five recorded eruption).

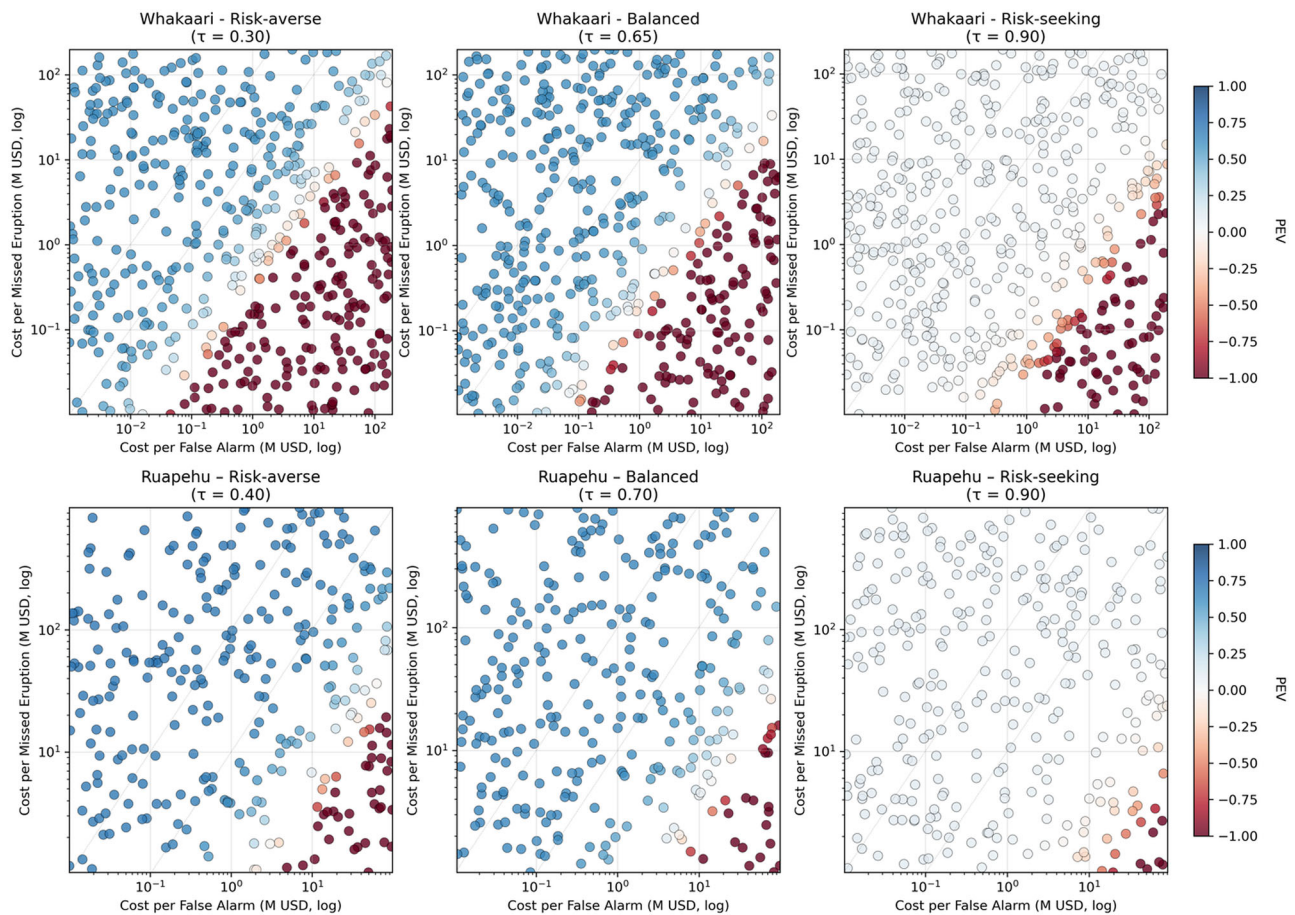


Fig. 3 | Sensitivity of potential economic value (PEV) to cost parameters for Whakaari (top row) and Ruapehu (bottom row). Each panel shows how PEV varies across false-alarm costs (C) and missed-eruption losses (L) for three forecast thresholds—risk-seeking ($\tau = 0.30/0.40$), balanced ($\tau = 0.65/0.70$), and risk-averse ($\tau = 0.90$). Each point represents a unique C – L combination, colored by PEV (blue = economic loss; red = economic benefit). The diagonal $PEV = 0$ boundary illustrates

the trade-off between false-alarm and missed-eruption costs: lower thresholds favor detection (positive PEV when $L \gg C$), while higher thresholds require very low false-alarm costs to remain beneficial. The geometry of these plots highlights that balanced thresholds provide the broadest region of robust, positive value under uncertainty, particularly in speculative scenarios such as Ruapehu.

be interpreted as conservative; greater savings would have been achievable had the optimal threshold been known in advance.

For the selected cost-modeling assumptions (20% of missed eruptions cause catastrophic losses of USD 20 M; each false alarm costs USD 0.5 M; see Methods and Supplementary Text S5), total costs accumulate in small increments from repeated false alarms, whereas the no-forecast baseline exhibits large jumps whenever an eruption is missed. In this case, frequent minor costs are less damaging than rare catastrophic outcomes. Short-term forecast costs can occasionally exceed the no-forecast baseline, because only a subset of missed eruptions are modeled to incur catastrophic losses; consequently, accepting a high false-alarm rate is economically rational only over sufficiently long-time horizons.

We examined PEV robustness by varying false-alarm costs (C) and missed-eruption losses (L) across several orders of magnitude (Fig. 3). Across plausible cost ratios, PEV remains positive, indicating that forecasts reduce expected losses even when valuations are uncertain. Threshold selection shapes this robustness: risk-seeking thresholds ($\tau = 0.30$) yield positive value only when $L \gg C$, whereas balanced thresholds ($\tau \approx 0.65$) expand the positive-PEV region substantially. Risk-averse thresholds ($\tau \approx 0.90$) erode value rapidly, as even small numbers of missed eruptions collapse economic benefit.

Ontake provides a complementary test of the PEV framework to Whakaari, representing a high-exposure, land-based volcanic setting with public access, minimal structural damage, and dominant life-

safety consequences. Using documented casualties and exposure at Ontake, a retrospective scenario analysis suggests that, under assumptions of timely precautionary action, a short-term forecast could have substantially reduced exposure-related losses while requiring repeated precautionary alerts over the analysis period (Fig. 2b).

If there was a forecasting method available at the time, and it was acted upon with evacuation planning, retrospective analysis shows that a forecast at Ontake could have reduced human and exposure-related losses by ~90%, avoiding short-term impacts despite 40 false alarms (Fig. 2b). This is consistent with the findings for Whakaari: when catastrophic human losses dominate the cost structure, even imperfect forecasts with high false-alarm rates provide substantial value over the long term.

Unlike Whakaari, where forecast actions primarily affect commercial operators and tourism access, Ontake illustrates a setting where precautionary actions are limited to temporary closure of public trails and summit access. As a result, false alarm costs are relatively modest compared to the potential consequences of a missed eruption, which are dominated by life-safety losses. This contrast shows how PEV responds not only to forecast performance, but also to differences in exposure structure and institutional control.

Ontake represents a challenging test case for forecast-based decision-making, as the retrospective analysis involves a large number of precautionary alerts relative to a single catastrophic outcome. It is

Table 1 | Performance comparison of tailored, generalized, and RSAM forecasting models for Whakaari, Bezymianny, and Copahue volcanoes

Volcano	Model	N Eruptions	Threshold	AUC	PEV	Savings (M USD)
Whakaari	tailored	5	0.89	0.88	0.55	219.98
Whakaari	generalized	5	0.62	0.96	0.75	300.43
Whakaari	rsam	5	94	0.85	0.47	186.93
Bezymianny	tailored	5	0.86	0.76	0.35	265.45
Bezymianny	generalized	5	0.88	0.69	0.38	284.74
Bezymianny	rsam	5	94	0.7	0.38	283.69
Copahue	tailored	6	0.59	0.78	0.39	162.14
Copahue	generalized	6	0.62	0.88	0.45	189.42
Copahue	rsam	6	87	0.76	0.37	154.75

Generalized models demonstrate consistent cross-volcano value, achieving the highest economic savings (189–300 M USD) and AUC values (predictive accuracy and economic performance) across all three volcanoes despite being trained on data from other volcanic systems. Values represent net cumulative savings over the monitoring periods analyzed (Whakaari: 2010–2020; Bezymianny: 2007–2009; Copahue: 2019–2020) at optimal thresholds ($\tau = 0.7\text{--}0.9$). This highlights the transferability and economic benefits of generalized eruption forecasting approaches.

important to note, however, that such high false-alarm rates are not an inherent feature of data-driven eruption forecasting. At volcanoes with more persistent and coherent precursory signals, short-term forecasting models can generate substantially fewer alerts while still providing clear advance warning. For example, out-of-sample application of a generalized forecasting model at Mount St. Helens showed elevated eruption likelihoods confined to a narrow window immediately preceding eruption, with limited spurious excursions during non-eruptive periods¹⁹. This contrast suggests that Ontake is a conservative stress test of the PEV framework, rather than a typical operational scenario.

Framework generalizability: hypothetical high-impact scenarios
Having examined retrospective cases, we tested the generalizability of PEV to volcanoes lacking recent major loss events but facing credible high-impact scenarios.

In high-exposure alpine settings like Mount Ruapehu, where lahars threaten people working in and visiting ski fields, accommodation areas, and hiking trails, short-term forecasts could reduce expected losses by enabling temporary access restrictions and evacuation of people from exposed areas (Fig. 2c).

At Ruapehu, false alarm costs accumulate steadily, but in scenario-based analysis the avoidance of rare catastrophic impacts dominates net expected value. Although only two minor eruptions occurred during the 14-year observation period, and neither resulted in casualties, this outcome reflects favorable timing and limited exposure rather than the absence of credible life-safety risk. Under plausible assumptions about the probability that a missed eruption coincides with high exposure (e.g., during ski season or peak visitation), ensemble modeling indicates that precautionary access restrictions could reduce expected losses by >40%. Toward the end of the observation period, forecast costs exceed the no-action baseline, reflecting an extended period without a high-impact event; however, this behavior is an expected feature of systems in which impact is realized intermittently, and highlights that apparent short-term ‘success’ without precautionary action can represent an extended period of risk accumulation rather than safety.

Sensitivity analysis (Fig. 3) reveals that balanced thresholds ($\tau \approx 0.70$) maintain positive PEV across wide cost ranges, whereas risk-averse thresholds ($\tau = 0.90$) eliminate value in the absence of recent catastrophic losses. This illustrates a practical challenge: in the absence of recent damaging eruptions to calibrate perceived risk, high alert thresholds may seem cautious, yet they increase expected losses by allowing rare but catastrophic events to be missed.

Applications to Copahue and Bezymianny (Table 1) further demonstrate framework generalizability across volcanic systems with

contrasting dominant avoidable loss pathways. At Bezymianny, short-term forecast value is driven primarily by aviation risk mitigation, as explosive eruptions frequently generate ash clouds that intersect major trans-Pacific air routes, while ground-based human exposure is minimal due to the volcano’s remote location and restricted access⁵⁰. In this context, precautionary actions consist mainly of airspace management and flight rerouting rather than evacuation of exposed populations.

In contrast, Copahue represents a system where forecast value is dominated by evacuation and access-control logistics. Eruptive activity necessitates the evacuation of nearby communities and large-scale movement of people, livestock, and infrastructure across the Chile–Argentina border, generating short-term economic and operational costs that are sensitive to alert timing⁵¹. While Copahue also produces ash affecting aviation, its principal avoidable losses arise from ground-based exposure and evacuation management.

Accordingly, at Copahue we explored the sensitivity of forecast value to alert threshold choice by quantifying how false alarms, missed eruptions, and economic outcomes vary across τ under fixed cost–loss assumptions. This analysis shows that positive forecast value is not confined to a single optimal threshold, but arises across a range of intermediate thresholds where reducing the likelihood of rare, high-impact misses outweighs the cumulative cost of precautionary actions (Supplementary Text S1; Figs. S1–S2).

Comparison of PEV across forecasting models

To assess how PEV depends on forecast model, we evaluated the metric on three different models: (1) a machine-learning model tailored to a specific volcano^{21,52}, (2) a model that uses transfer-learning to pool precursor information from multiple volcanoes¹⁹, and (3) a benchmark forecast based on the unmodified seismic data (RSAM) that these machine-learning models learn from. As tailored models can only be developed for a small subset of volcanoes with sufficient recorded eruptions, we have restricted this analysis to Whakaari, Bezymianny and Copahue.

For Whakaari, the generalized model yielded highest PEV (0.38–0.75) and greatest cost savings, with the tailored model performing slightly worse (Table 1). RSAM produced lower but still positive value, demonstrating that even unmodified amplitude metrics have forecasting and decision-making utility. The same comparison for Copahue and Bezymianny also indicates that generalized models achieve the highest positive economic value (189–300 M USD savings), despite training on data from other volcanic systems (Table 1). This suggests that models incorporating generalized multi-volcano precursory signals¹⁹ offer operational advantages for data-limited settings where local eruption histories are insufficient.

An important operational consideration is the trade-off between forecast performance and interpretability. RSAM, as a simple seismic amplitude metric, is straightforward to implement, explain to stakeholders, and audit, and our analysis confirms that it provides positive economic value at all tested volcanoes (Table 1). The ML models provide additional value (USD 30–115 M in marginal savings over RSAM across the tested volcanoes) by detecting subtler precursory patterns, and are partially interpretable through feature importances and physically meaningful input features^{19,21}. In operational settings, RSAM and ML models need not be competing alternatives: RSAM can serve as a transparent baseline that agencies and the public can readily understand, while ML outputs provide supplementary decision support with quantifiable additional value⁵³. The choice of which to deploy depends on local institutional capacity, data infrastructure, and the communication context in which forecasts are used.

The finding that generalized models outperform tailored models (Table 1) is consistent with a well-established regularity in multi-task machine learning applied to data-scarce settings. In hydrological prediction⁵⁴, it has been shown that a single model trained across 531 catchments outperformed individually calibrated models at each site, because the pooled network captures process invariances that small per-site samples cannot resolve. With 5–6 eruptions per volcano (Table 1), our dataset is in the regime where pooled training has been shown to outperform per-site calibration in other domains. This result also supports the interpretation that seismic precursory features are more universal across volcanic systems than previously assumed, consistent with the transfer-learning findings of Ardid and Dempsey et al., 2025¹⁹.

Discussion

Advantages of the PEV framework in volcanology

Application of the PEV framework to Whakaari and Ontake suggests that a substantial fraction (30–90%) of preventable losses in these retrospective case studies could, in principle, be reduced through timely precautionary action. A key insight from these analyses is that forecast utility derives primarily from reducing the probability of rare, high-consequence outcomes rather than from maximizing conventional accuracy metrics alone. This behavior is consistent with established cost-loss theory in meteorology⁴⁵, but has not been routinely incorporated into eruption forecasting evaluation. By focusing on avoidable losses driven by exposure rather than eruptive magnitude, the PEV framework aligns with evidence that vulnerability dominates eruption impact across diverse volcanic settings (see Supplementary Text S2).

An additional strength of PEV is that it provides a common, dimensionless structure for evaluating forecast value across volcanic systems with differing eruption frequencies, exposure patterns, and cost structures. By focusing on relative trade-offs between false alarms and missed events, rather than absolute costs, the framework remains informative even when economic valuations are uncertain by orders of magnitude.

Operational challenges

Although the PEV analysis identifies a value-maximizing decision threshold in hindsight, this threshold is not known a priori in operational settings. In practice, agencies must choose alert thresholds under uncertainty, informed by imperfect knowledge of future eruption behaviour, exposure, and societal tolerance for false alarms versus missed events. Importantly, our sensitivity analyses indicate that while absolute economic value varies with cost assumptions, the location of the PEV optimum remains relatively stable across a broad range of plausible cost ratios. This stability suggests that threshold selection can be robust to substantial uncertainty in economic valuation, supporting the use of PEV as a decision-support tool rather than a prescriptive optimization rule.

Applying the PEV framework in operational settings presents important challenges. Most notably, the decision threshold that maximizes expected value can only be identified retrospectively and is therefore not known with foresight. In practice, agencies must select thresholds under uncertainty, informed by imperfect knowledge of future eruption behaviour, exposure patterns, and their tolerance for false alarms relative to missed events. As a result, PEV should be interpreted as a decision-support tool for exploring trade-offs, rather than as a prescriptive rule for optimal action.

A further limitation arises at volcanoes with sparse eruptive records, where prospective estimation of false-alarm rates and missed-eruption probabilities is inherently uncertain. In such settings, forward simulation and sensitivity analysis are essential for examining plausible long-term cost-loss trajectories and for avoiding overconfidence based on limited historical experience.

In practice, we recommend that agencies adopt a robust initial threshold, one that maintains positive PEV across a broad range of plausible cost assumptions, rather than attempting to identify a value-maximizing threshold a priori. Our sensitivity analyses (Fig. 3) indicate that balanced thresholds ($\tau \approx 0.65$ – 0.70) sustain positive PEV across wide cost-ratio ranges, making them robust initial choices. This approach parallels earthquake early warning practice, where alert thresholds are set below the target damage threshold to bound miss probability at a policy-specified ceiling, accepting a measurable increase in false alerts^{55,56}. As forecast-outcome data accumulate, thresholds can be refined through Bayesian updating of the posterior expected loss, following established methods in flood and weather forecasting^{57,58}. This sequential approach protects against the catastrophic-regret scenario in which a too-permissive threshold is chosen and a major eruption is missed.

Finally, the application of PEV requires assigning monetary values to human life and other impacts, a methodologically necessary but ethically constrained practice in risk analysis⁵⁹. Values such as the Value of Statistical Life (VSL) vary substantially across jurisdictions, reflecting differences in regulatory practice, income levels, and policy conventions rather than any normative assessment of individual lives. Moreover, economic valuation cannot fully capture non-monetary consequences such as psychological trauma, long-term displacement, or loss of cultural and social assets^{31,41}. For this reason, PEV should not be interpreted as a comprehensive measure of societal impact. Rather, it provides a structured means of comparing alternative warning strategies under uncertainty. Importantly, sensitivity analyses show that the qualitative conclusions of this study are robust across order-of-magnitude variation in cost assumptions, indicating that the potential value of forecasting does not hinge on any single valuation choice. Decisions about acceptable risk and warning practice must therefore be informed by, but not determined solely through, economic optimization, and should be situated within broader deliberative and governance processes^{15,38}.

The false-alarm cost C is primarily composed of fixed operational expenditures (evacuation logistics, lost revenue, emergency service mobilization) that are largely independent of VSL, whereas the missed-eruption loss L scales with the chosen VSL benchmark through its fatality component. Increasing VSL therefore decreases the cost-loss ratio C/L ⁶⁰ and shifts the optimal threshold toward more precautionary values, consistent with the ethical intuition that greater weight on human life should lower the bar for protective action. The sensitivity analyses in Fig. 3 show that PEV remains positive across a wide range of C/L values, and the broad, flat optimum means that shifts in optimal threshold location across plausible VSL ranges are small. The framework's policy-relevant conclusions, namely which thresholds provide positive value and that balanced thresholds are preferred, are therefore robust to the choice of VSL benchmark.

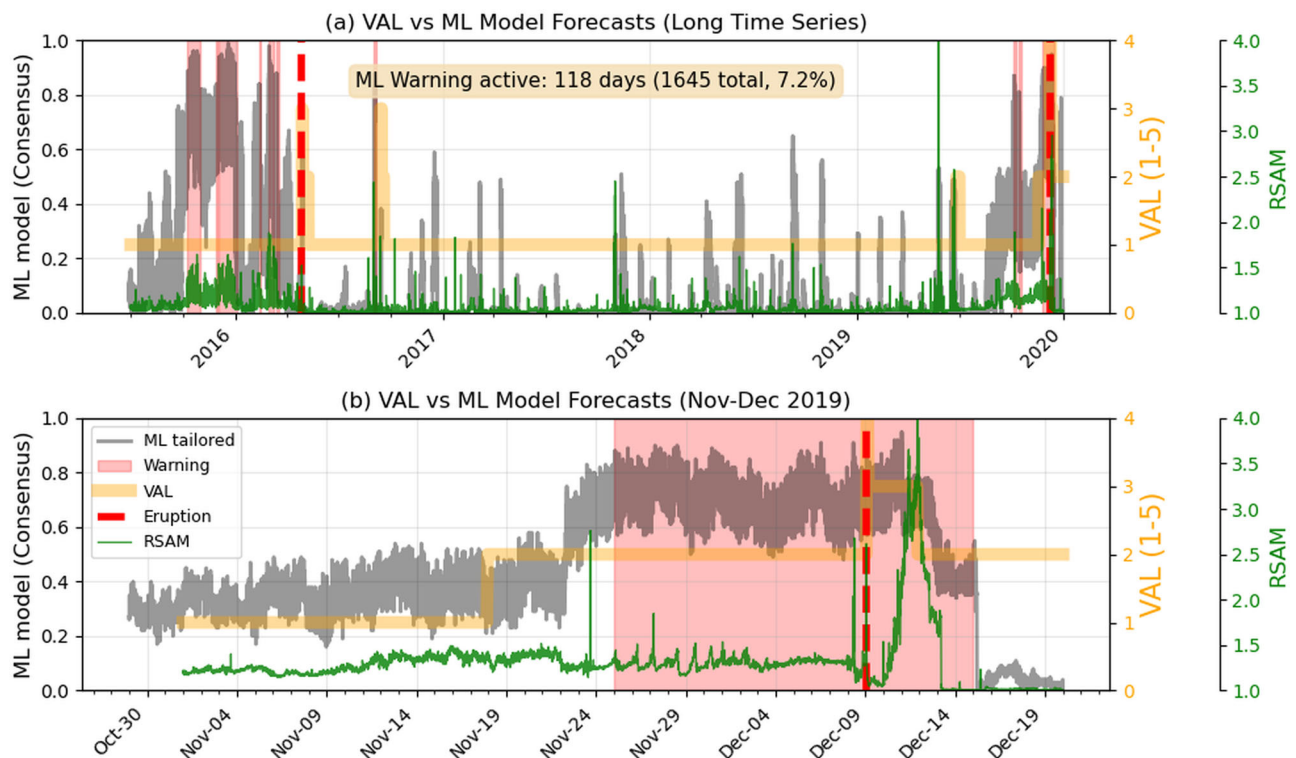


Fig. 4 | Comparison of ML eruption probability forecasts with operational alert levels at Whakaari (2015–2020). **a** Long-term context showing ML forecast probabilities (gray, left axis), RSAM seismic amplitude (green, right axis), and as reference official Volcanic Alert Level is plotted (VAL; orange, right axis). Red shaded regions indicate periods when the ML warning system would have been active using a threshold-and-hold scheme ($\tau = 0.7$, 5-day hold; see text), totaling

118 days (7.2% of the monitoring period). Red dashed lines mark eruptions.

b Detailed view of November–December 2019 showing ML forecast behavior before and during the fatal 9 December eruption. The threshold-and-hold post-processing converts high-frequency ML probability fluctuations into discrete warning periods (red shading), providing actionable alerts while limiting warning-active time.

Distribution of forecast-related costs and societal acceptance

Forecast-informed decision-making redistributes the short-term costs and consequences associated with volcanic risk, and these are not borne uniformly across affected groups. For example, at Whakaari, precautionary actions triggered by elevated forecasts primarily impose costs on tourism operators and visitors through disrupted access, lost revenue, and travel interruptions. In contrast, the consequences of missed eruptions are dominated by life-safety impacts borne by exposed individuals (including visitors, workers, and operators) as well as by public agencies responsible for emergency response, recovery, and legal accountability (for an hypothetical analysis on this asymmetry in cost incidence under false-alarm and missed-eruption outcomes, see Supplementary Text S3 and Figure S3).

Within the PEV framework, forecast value arises when frequent, relatively low-impact disruptions are exchanged for a reduced probability of rare, high-consequence outcomes. This exchange highlights a central challenge for operational forecasting: the groups that experience the immediate costs of precautionary actions are not necessarily those most directly affected by avoided catastrophic losses. As a result, societal acceptance of frequent precautionary alerts depends not only on forecast performance, but also on how these uneven cost burdens are perceived, communicated, and managed over time.

We emphasize that PEV is intended as a decision-support tool for facilitating transparent stakeholder deliberation, not as a prescriptive metric that determines policy. A stakeholder-partitioned analysis, decomposing aggregate PEV into private-sector costs (e.g., tourism revenue losses from closures) and public-safety gains (e.g., avoided fatalities and emergency response costs), can inform how the costs and benefits of precautionary action are distributed across affected groups. Supplementary Text S3 provides an illustrative decomposition

for Whakaari. Formal methods for such distributional analysis are well-established in environmental regulatory analysis⁶¹, and represent a natural extension of the PEV framework for site-specific policy application.

Operational integration and behavioral considerations

Successful implementation of PEV-informed forecasting requires addressing both technical and behavioral dimensions. Research on warning compliance shows that repeated false alarms can erode public response by reducing perceived credibility and willingness to act, a phenomenon commonly referred to as the cry-wolf effect^{40,62}, while missed eruptions can cause severe loss of institutional trust and long-lasting societal impacts⁶³. Historical volcanic crises illustrate this asymmetry: the 1976–77 La Soufrière evacuation⁶⁴ (Saint Vincent, Caribbean), which was not followed by eruption, was retrospectively judged justified, whereas failures to warn at 1975 Nevado del Ruiz eruption⁶⁵ (Colombia), Ontake (2014), and Whakaari (2019), generated enduring criticism. Importantly, evidence from risk-communication research suggests that transparent communication of uncertainty can sustain credibility even when forecasts are imperfect⁶⁶.

A central behavioral challenge is therefore managing societal response to frequent precautionary warnings. Compliance erodes most strongly when alerts are perceived as binary predictions that repeatedly “fail” to materialize. Addressing this could require, for instance, reframing alerts as indicators of elevated risk rather than definitive eruption statements. Analogous approaches are widely used in other hazard domains: fire-danger rating systems, for example, routinely signal heightened risk and recommend precautionary behaviour without implying that a wildfire is imminent^{49,67}. A similar

emphasis on relative risk and proportional action may help improve societal acceptance of precautionary volcanic warnings.

Operationally, this behavioral framing must be supported by alert designs that translate high-frequency forecast outputs into warnings that are actionable but not overwhelming. Short-term forecasting tools (e.g., RSAM and machine-learning models) can fluctuate on much shorter timescales than established alert systems such as the Volcanic Alert Level (VAL⁶⁸), reflecting sensitivity to transient seismic patterns (Fig. 4a). To bridge this mismatch, we apply a threshold-and-hold post-processing scheme¹¹, in which alerts are triggered only when forecasts exceed a chosen threshold and remain active until elevated risk subsides for sustained periods (e.g., several consecutive days below threshold before deactivation).

In the Whakaari example, this approach reduces the proportion of time spent under active warning to 7.2% of days (118 of 1,645 days over 2015–2020; Fig. 4a), while preserving critical early-warning capability prior to the 2019 eruption (Fig. 4b). Such smoothing illustrates how false-alarm fatigue can be mitigated through alert design rather than by suppressing forecast sensitivity altogether. These parameters are illustrative; operational deployment requires calibration with agencies to balance competing objectives, including cost efficiency (maximizing PEV), behavioral response (maintaining compliance), and workflow compatibility (avoiding cognitive overload; in other words, operational deployment must remain simple enough to be useful for non-expert operators and should not add unnecessary layers of complexity). In this sense, PEV-informed forecasting functions as a decision-support and communication tool rather than a purely technical optimization.

Future extensions of the PEV framework could incorporate a dynamic trust variable that decays after false alarms and recovers after correct alerts, following the socio-hydrological warning model⁶⁹. The ‘crying wolf’ hypothesis⁷⁰ predicts that repeated false alarms erode public compliance, and some studies support this for tornado warnings⁷¹. However, others do not replicate the finding for protective behaviour^{72,73}, and tsunami warning systems routinely operate with high false-alarm ratios while maintaining effective evacuation compliance, suggesting that warning credibility depends on risk communication practices and institutional trust rather than alarm frequency alone. Our threshold-and-hold post-processing scheme (Fig. 4) and the recommended reframing of alerts as risk indicators rather than binary predictions provide practical mechanisms for managing alert fatigue without requiring explicit trust modeling.

We acknowledge that this framework assumes that forecasts, once issued, lead to protective action. In practice, the translation of forecast to action depends on institutional decision-making, effective communication, and political acceptance of precautionary costs, particularly where tourism operators exert lobbying pressure (e.g., Whakaari, Copahue). The PEV framework supports this translation by providing quantitative evidence that forecast-based action yields net economic benefit, but successful implementation requires engagement with communities and civil authorities to build understanding of forecast uncertainty and acceptance that some alerts will not be followed by eruption.

Model assumptions and data constraints

Our analysis focuses on direct, short-term economic impacts that can be reduced through early warning, including fatalities, injuries, emergency response, business interruption, and operational closures. We intentionally exclude longer-term consequences (e.g., infrastructure reconstruction, agricultural recovery, and national productivity decline), as these are largely unavoidable even under perfect forecasting. As a result, our estimates are conservative and likely underestimate the total benefits of effective forecasting, while ensuring that PEV captures only losses that are genuinely preventable through warning.

The cost model is necessarily simplified using binary outcomes (eruption/no eruption) and time-invariant parameters, whereas real losses vary with exposure, season, and behaviour. For example, exposure at Ruapehu peaks during the winter ski season (~10,000 people per day), while Whakaari hosted ~70 people per day on average during its operation period.

A natural extension of the framework is to replace the scalar missed-eruption loss L with a time-varying function $L(t)$ that reflects seasonal or real-time exposure, following established practice in flood-damage assessment⁷⁴ and hurricane warning decision models⁷⁵. At Whakaari, for example, $L(t)$ could be indexed to daily visitor manifests; at Ruapehu, to ski-field attendance. Because our fixed- L approach does not weight misses by exposure, the reported PEV values are conservative: dynamic weighting would increase estimated forecast value during peak-exposure periods when the consequences of a miss are greatest.

We do not address longer-term questions such as optimal monitoring network design, multi-hazard interactions, or integration with non-seismic data streams, nor do we explicitly model real-time exposure dynamics or graduated response strategies. Despite these limitations, consistent results across diverse volcanic settings demonstrate framework robustness, including stable optimal thresholds ($\tau \approx 0.65$ – 0.90) and positive PEV under order-of-magnitude cost uncertainty.

Methods

The modeling methodology used here links eruption forecast outputs with short-term economic impact assessment using the Potential Economic Value (PEV) framework. By assigning costs to false alarms and missed eruptions and incorporating probabilistic severity distributions, we are able to compute PEV metrics for different forecast decision thresholds, evaluate how costs accumulate over time and are distributed across stakeholders.

Potential economic value (PEV) Framework

Forecast utility is quantified by comparing expected costs under three strategies for taking precautionary measures, termed “acting.” Precautionary measures are a binary choice to evacuate (or not) and it is assumed that evacuation reduces loss-of-life exposure to zero. The three strategies evaluated are (1) acting imperfectly on the instruction of a forecast, (2) never acting, which is the same as having no forecast information on which to act, and (3) acting perfectly, which assumes a forecast that is 100% accurate.

We consider forecasting systems that output a continuous numerical “risk of eruption” variable $y(t) \in [0, y_{\max}]$. A binary precautionary action is parameterized by a decision threshold τ , with an action (e.g., evacuation/closure) taken whenever $y(t) \geq \tau$.

Forecast performance at threshold τ is summarized using contingency statistics computed over fixed decision windows. We define, the miss fraction:

$$m(\tau) = 1 - \text{TPR}(\tau), \quad (1)$$

where $\text{TPR}(\tau)$ (True Positive Rate) is the fraction of pre-eruptive decision windows correctly flagged. The false-alarm burden over the evaluation horizon

$$b(\tau) = \text{FPR}(\tau)T, \quad (2)$$

where $\text{FPR}(\tau)$ (False Positive Rate) is the fraction of non-event windows that exceed τ , and T is the number of non-event windows (or equivalently the duration of evaluated non-event time expressed in the same window units).

Under these definitions, the expected cost over the evaluation horizon when acting on the forecast at threshold τ is

$$E_{\text{forecast}}(\tau) = Cb(\tau) + Lm(\tau) \quad (3)$$

where C is the cost per false-alarm action (e.g., per day of closure) and L is the loss per missed eruption (expressed as a cost).

We compare this to two baselines: a no-forecast strategy that never acts, $E_{\text{no}} = L$ (i.e., all eruptions are missed in the binary-action sense), and a perfect forecast, $E_{\text{perfect}} = 0$. Then the Potential Economic Value is then defined as

$$\text{PEV}(\tau) = \frac{E_{\text{no}} - E_{\text{forecast}}(\tau)}{E_{\text{no}} - E_{\text{perfect}}} = 1 - \frac{Cb(\tau) + Lm(\tau)}{L} \quad (4)$$

PEV ranges from $-\infty$ to 1, where:

- PEV = 1: perfect forecast (no false alarms, no missed eruptions)
- PEV = 0: forecast performs equivalently to no-forecast strategy
- PEV < 0: forecast causes more harm than doing nothing
- PEV > 0: forecast provides net benefit; magnitude indicates fraction of avoidable losses prevented

This formulation follows the standard cost–loss framework used to compute relative economic value in forecast verification^{44,45}. Importantly, because PEV is a ratio of costs with consistent normalization, it is invariant to whether costs are expressed per decision window, per unit time, or over a fixed horizon, provided C , L , and the forecast performance terms are defined on the same basis.

As a framework, the PEV is dependent on the decision threshold τ , and therefore allows evaluation of different choices of τ . Furthermore, PEV is principally sensitive to a cost-ratio, C/L , rather than absolute magnitudes of individual cost contributions. As a dimensionless quantity, it is theoretically comparable across different locations and hazard types. Finally, it is directly evaluative of forecasts for decision-making contexts.

We applied PEV to a series of pseudo-prospective eruption forecasts that were obtained from a machine-learning model trained on seismic data from multiple volcanoes¹⁹. We tested the framework on five case study volcanoes excluded from the training set to compare the relative quality of these forecasts under different decision thresholds and cost assumptions.

The first two volcanoes were Whakaari, in New Zealand, and Mt Ontake, in Japan. For these examples, cost parameters are derived from documented casualties, economic impacts, and recovery expenses from recent fatal eruptions. Mt Ruapehu and Copahue, has not experienced a mass-casualty eruption. However, for PEV to be useful in justifying the introduction of an operating forecast system at a particular volcano, such hypothetical applications are necessary. In these situations, cost parameters are estimated based on historical precedents, regional economic data, and risk assessments, and then subject to sensitivity analysis.

This approach distinguishes evaluation cases, where PEV results can be compared to actual outcomes, from demonstration where PEV reveals forecast value under plausible but unconstrained future scenarios. PEV measures only the avoidable portion of eruption-related losses (those that can be mitigated through timely warning and protective actions) and therefore does not include physical damage that would occur even under perfect forecasting.

Forecast data and model outputs

We evaluate a forecasting model that estimates a likelihood proxy of an eruption occurring within the next 48 h based on the preceding 96 h of seismic activity. These forecasts are derived from the model of Ardid & Dempsey et al. (2025)¹⁹, which extracts time series features from seismic data from a 24-volcano training dataset and outputs a

continuous time series (“consensus”) sampled every 10 min between 0 and 1.

PEV analysis was applied to these forecast time series across extended eruptive and non-eruptive periods for each volcano. Forecasts were evaluated under out-of-sample conditions (pseudo-prospective forecasting¹⁹).

To translate continuous forecast outputs into operational decisions, a decision threshold (τ) was applied to the forecast likelihood time series. Values exceeding τ were interpreted as eruption alarms, while values below τ indicated no alarm. Alarms were evaluated relative to eruption catalogs using a fixed pre-eruptive window of 48 h prior to eruption onset. Alarms occurring outside these windows were classified as false alarms, whereas failures to raise an alarm before an eruption were classified as missed eruptions. By sweeping τ across the $[0, 1]$ range, we quantified how false-alarm rates, missed-eruption rates, and cumulative costs vary with threshold choice, forming the basis of the PEV analysis (Details of model architecture, alarm construction, and pseudo-prospective evaluation are discussed in Ardid & Dempsey et al. (2025)¹⁹ and also provided in Supplementary Text S4).

Cost assumptions

We applied the PEV framework to five volcanoes spanning diverse eruption frequencies, hazard styles, exposure patterns, and socio-economic contexts: Whakaari, Ruapehu, Ontake, Copahue, and Bezymianny. Costs represent short-term, forecast-sensitive impacts, including economic disruption from precautionary actions (false alarms; C) and representative human and economic losses from eruptions occurring without mitigation (missed eruptions; L).

Cost parameters were chosen using documented impacts from historical eruptions, official loss estimates, and established valuations of human life, supplemented by conservative assumptions where empirical data are incomplete. Across all cases, false-alarm costs are likely overestimated, while missed-eruption losses are likely underestimated (particularly for rare, high-consequence events with cascading effects) yielding a cautious lower bound on forecast value. This conservative bias ensures that positive PEV values do not depend on optimistic cost assumptions.

Uncertainty in cost estimates was explored using ± 50 –60% Monte Carlo variation, and volcano-specific justifications, references, and sensitivity analyses are provided in Supplementary Text S5 and S6.

For each volcano we specify central values of costs around which sensitivity is tested:

Whakaari (New Zealand).

- **False-alarm cost:** NZD 0.5M per day (short-term disruption to commercial tourism operations, marine transport, and associated services during precautionary closures).
- **Missed-eruption cost:** USD 20 M (risk-weighted representation of human and economic losses from a Whakaari-style catastrophic event, distributed across multiple forecast cycles; see Supplementary Text S5.1).

Ontake (Japan).

- **False-alarm cost:** USD 1M per day (temporary closure of summit trails, ropeways, and mountain facilities during peak hiking seasons).
- **Missed-eruption cost:** USD 400 M (human-capital losses from ~63 fatalities combined with prolonged post-eruption tourism and access restrictions).

Ruapehu (New Zealand).

- **False-alarm cost:** USD 0.5M per day (ski-field closures and associated regional tourism disruption during winter operations).

- **Missed-eruption cost:** USD 75 M (conservative estimate informed by the 1953 Tangiwai lahar disaster and the 1995–96 eruption sequence).

Copahue (Chile/Argentina).

- **False-alarm cost:** USD 0.5 M per day (logistical and economic costs associated with mandatory evacuations and access restrictions, based on the 2013 crisis).
- **Missed-eruption cost:** USD 30 M (risk-weighted average combining plausible minor phreatic events and lower-probability catastrophic scenarios affecting nearby communities).

Bezymianny (Russia).

- **False-alarm cost:** USD 1 M per day (operational costs of precautionary airspace management and trans-Pacific flight rerouting).
- **Missed-eruption cost:** USD 250 M (aviation disruption and infrastructure losses associated with large ash-producing eruptions, informed by historical aircraft–ash encounters).

Data availability

We provide the ML forecasting model output (consensus time series) for Whakaari over ~10 years, from which the PEV is computed. The data are provided in the inputs/ folder and include: (1) ML model consensus predictions in pickle format for both tailored (volcano-specific) and generalized (multi-volcano) models, containing eruption probability predictions at 10-minute intervals; (2) eruption times (5 eruptions from 2012–2019) in text format used for forecast evaluation; and (3) pre-processed seismic amplitude datasets (RSAM) sampled at 10-minute intervals with pre-computed ROC curve data. Raw seismic waveform data for New Zealand volcanoes are publicly available from GeoNet and can be accessed via standard FDSN web services. The datasets used in this study are accessible as FDSN clients through the International Federation of Digital Seismograph Networks (FDSN) web service infrastructure (<https://www.fdsn.org/networks/>). Source data are provided with this paper.

Code availability

Code used to generate the ML forecasts are given in previous work (ref. 19). Here we provide the code to evaluate Potential Economic Value (PEV) from the ML forecasting model outputs. This includes scripts for threshold evaluation and cost-loss analysis. The main script (pev_model.py) contains functions for: (1) threshold evaluation (roc_counts()) that computes True Positive Rate (TPR) and False Positive Rate (FPR) across probability thresholds using pre-eruption windows; (2) cost-loss analysis (pev_curve()) that calculates PEV and net savings by comparing forecast costs with losses from missed eruptions; and (3) Monte Carlo uncertainty quantification (monte_carlo_pev()) for cost/loss parameter uncertainty. The code also includes visualization functions to generate ROC curves, PEV curves, and summary tables. Supporting files include example configuration templates, usage examples, documentation, and Python package dependencies. The code is designed to work with the provided Whakaari data and can be extended to other volcanoes by modifying the configuration.

References

- Auker, M. R., Sparks, R. S. J., Siebert, L., Crowther, H. S. & Ewert, J. A statistical analysis of the global historical volcanic fatalities record. *J. Appl. Volcanol.* **2**, 2 (2013).
- Mei, E. T. W. et al. Lessons learned from the 2010 evacuations at Merapi volcano. *J. Volcanol. Geotherm. Res.* **261**, 348–365 (2013).
- Ulfarsson, G. F. & Unger, E. A. Impacts and responses of Icelandic aviation to the 2010 Eyjafjallajökull volcanic eruption: case study. *Transp. Res. Rec.* **2214**, 144–151 (2011).
- Doocy, S., Daniels, A., Dooling, S. & Gorokhovich, Y. The human impact of volcanoes: a historical review of events 1900–2009 and systematic literature review. *PLoS Curr.* **5**, ecurrents.di-s.841859091a706efebf8a30f4ed7a1901 (2013).
- Craig, H. et al. Impacts to agriculture and critical infrastructure in Argentina after ashfall from the 2011 eruption of the Cordón Caulle volcanic complex: an assessment of published damage and function thresholds. *J. Appl. Volcanol.* **5**, 7 (2016).
- Hayes, J. L. et al. Tephra clean-up after the 2015 eruption of Calbuco volcano, Chile: a quantitative geospatial assessment in four communities. *J. Appl. Volcanol.* **8**, 7 (2019).
- Troll, V. R. et al. The 2021 La Palma eruption: social dilemmas resulting from life close to an active volcano. *Geol. Today* **40**, 96–111 (2024).
- Amigo, A. Volcano monitoring and hazard assessments in Chile. *Volcanica* **4**, 1–20 (2021).
- Borrero, J. C. et al. Tsunami runup and inundation in Tonga from the January 2022 eruption of Hunga Volcano. *Pure Appl. Geophys.* **180**, 1–22 (2023).
- Yamaoka, K., Geshi, N., Hashimoto, T., Ingebritsen, S. E. & Oikawa, T. Special issue “The phreatic eruption of Mt. Ontake volcano in 2014. *Earth Planets Space* **68**, 175 (2016).
- Dempsey, D. E. et al. Evaluation of short-term probabilistic eruption forecasting at Whakaari, New Zealand. *Bull. Volcanol.* **84**, 91 (2022).
- Chouet, B. A. Long-period volcano seismicity: its source and use in eruption forecasting. *Nature* **380**, 309–316 (1996).
- Biggs, J. et al. Global link between deformation and volcanic eruption quantified by satellite imagery. *Nat. Commun.* **5**, 3471 (2014).
- de Moor, J. M. et al. Short-period volcanic gas precursors to phreatic eruptions: insights from Poás Volcano, Costa Rica. *Earth Planet. Sci. Lett.* **442**, 218–227 (2016).
- Marzocchi, W. & Woo, G. Probabilistic eruption forecasting and the call for an evacuation. *Geophys. Res. Lett.* **34**, L22310 (2007).
- Marzocchi, W. & Bebbington, M. S. Probabilistic eruption forecasting at short and long time scales. *Bull. Volcanol.* **74**, 1777–1805 (2012).
- Ardid, A., Dempsey, D., Caudron, C. & Cronin, S. Seismic precursors to the Whakaari 2019 phreatic eruption are transferable to other eruptions and volcanoes. *Nat. Commun.* **13**, 2002 (2022).
- Walsh, B. et al. Geophysical examination of the 2016 Whakaari eruption. *Earth Planets Space* **71**, 25 (2019).
- Ardid, A. et al. Ergodic seismic precursors and transfer learning for short term eruption forecasting at data scarce volcanoes. *Nat. Commun.* **16**, 1758. <https://doi.org/10.1038/s41467-025-56689-x> (2025).
- Manley, G. F. et al. Understanding the timing of eruption end using a machine learning approach to classification of seismic time series. *J. Volcanol. Geotherm. Res.* **401**, 106917 (2020).
- Ardid, A., Barth, A., Dempsey, D., Manga, M. & Cronin, S. J. Forecasting eruptions at Steamboat Geyser: time scales, differentiability, and detectability of seismic precursors through data-driven methods. *J. Geophys. Res. Mach. Learn. Comput.* **2**, e2024JH000569 (2025).
- Wang, T., Bebbington, M., Cronin, S. & Carman, J. Forecasting eruptions at poorly known volcanoes using analogs and multivariate renewal processes. *Geophys. Res. Lett.* **49**, e2021GL096715 (2022).
- Zali, Z., Mousavi, S. M., Ohrnberger, M., Eibl, E. P. S. & Cotton, F. Tremor clustering reveals pre-eruptive signals and evolution of the 2021 Geldingadalir eruption of the Fagradalsfjall Fires, Iceland. *Commun. Earth Environ.* **5**, 1 (2024).
- Rey-Devesa, P. et al. Characterization of volcanic stages using seismic features: case of Tajogaite (2021) and Colima (2013–2022). *J. Volcanol. Geotherm. Res.* **468**, 108454 (2025).

25. Chastin, S. F. & Main, I. G. Statistical analysis of daily seismic event rate as a precursor to volcanic eruptions. *Geophys. Res. Lett.* **30**, 1671 (2003).
26. Pesicek, J. D., Wellik, J. J., Prejean, S. G. & Ogburn, S. E. Prevalence of seismic rate anomalies preceding volcanic eruptions in Alaska. *Front. Earth Sci.* **6**, 100 (2018).
27. Behr, Y., Christophersen, A. & Miller, C. Probabilistic, multi-sensor eruption forecasting. *Geophys. Res. Lett.* **52**, e2024GL112029 (2025).
28. Ardid, A. et al. Using template matching to detect hidden fluid release episodes beneath crater lakes in Ruapehu, Copahue, and Kawah Ijen volcanoes. *J. Geophys. Res. Solid Earth* **128**, e2023JB026729 (2023).
29. Ardid, A. et al. Multitimescale template matching: discovering eruption precursors across diverse volcanic settings. *Seismol. Res. Lett.* **95**, 2611–2621 (2024).
30. Haynes, K., Barclay, J. & Pidgeon, N. The issue of trust and its influence on risk communication during a volcanic crisis. *Bull. Volcanol.* **70**, 605–621 (2008).
31. Andreastuti, S. D., Paripurno, E. T., Subandriyo, S., Syahbana, D. K. & Prayoga, A. S. Volcano disaster risk management during crisis: implementation of risk communication in Indonesia. *J. Appl. Volcanol.* **12**, 3 (2023).
32. Rose, A. Economic resilience to disasters: toward a consistent and comprehensive formulation. in *Disaster Resilience: An Integrated Approach* 275–303 (Charles C Thomas, 2006).
33. McDonald, G., Smith, N., Kim, J.-H. & Cronin, S. The spatial and temporal ‘cost’ of volcanic eruptions: assessing the economic impact, business inoperability, and spatial distribution of risk in the Auckland region, New Zealand. *Bull. Volcanol.* **79**, 48 (2017).
34. McDonald, G. et al. Computable general equilibrium modelling of economic impacts from volcanic event scenarios at regional and national scale. *Mt. Taranaki, N.Z. Bull. Volcanol.* **79**, 87 (2017).
35. Choumert-Nkolo, J., Lamour, A. & Phelinas, P. The economics of volcanoes. *Econ. Disasters Clim. Change* **5**, 277–299 (2021).
36. Ministry of Transport. Social Cost of Road Crashes and Injuries 2020 Update. https://www.transport.govt.nz/assets/Uploads/Social-Cost-of-Road-Crashes-and-Injuries-2020_final.pdf (2020).
37. U.S. Department of Transportation. Departmental Guidance on Valuation of a Statistical Life in Economic Analysis (2023 Update). <https://www.transportation.gov/resources/value-of-a-statistical-life-guidance> (2023).
38. Woo, G. Probabilistic criteria for volcano evacuation decisions. *Nat. Hazards* **45**, 87–97 (2008).
39. Logan, T. M., Aven, T., Guikema, S. & Flage, R. The role of time in risk and risk analysis: implications for resilience, sustainability, and management. *Risk Anal.* **41**, 1959–1970 (2021).
40. Roulston, M. S. & Smith, L. A. The boy who cried wolf revisited: the impact of false alarm intolerance on cost-loss scenarios. *Weather Forecast* **19**, 391–397 (2004).
41. Johnston, D. M., Houghton, B. F., Neall, V. E., Ronan, K. R. & Paton, D. Impacts of the 1945 and 1995–1996 Ruapehu eruptions, New Zealand: an example of increasing societal vulnerability. *Geol. Soc. Am. Bull.* **112**, 720–726 (2000).
42. Selva, J., Orsi, G., Di Vito, M. A., Marzocchi, W. & Sandri, L. Probability hazard map for future vent opening at the Campi Flegrei caldera, Italy. *Bull. Volcanol.* **74**, 497–510 (2012).
43. Kilgour, G. et al. Whakaari/White Island: a review of New Zealand’s most active volcano. *N. Z. J. Geol. Geophys.* **64**, 273–295 (2021).
44. Murphy, A. H. The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Mon. Weather Rev.* **105**, 803–816 (1977).
45. Richardson, D. S. Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **126**, 649–667 (2000).
46. Ramos, M. H., Van Andel, S. J. & Pappenberger, F. Do probabilistic forecasts lead to better decisions?. *Hydrol. Earth Syst. Sci.* **17**, 2219–2232 (2013).
47. Coughlan de Perez, E. et al. Forecast-based financing. *Nat. Hazards Earth Syst. Sci.* **15**, 895–904 (2015).
48. Busker, T., de Moel, H., van den Hurk, B. & Aerts, J. C. Impact-based seasonal rainfall forecasting to trigger early action for droughts. *Sci. Total Environ.* **898**, 165506 (2023).
49. Ardid, A. et al. Sub-hourly forecasting of fire potential using machine learning on time series of surface weather variables. *Int. J. Wildland Fire* **34**, WF24113 (2025).
50. Miller, T. P. & Casadevall, T. J. Volcanic ash hazards to aviation. in *Encyclopedia of Volcanoes* (ed. Sigurdsson, H.) 915–930 (Academic Press, 2000).
51. Caselli, A. T., Liccioli, C. & Tassi, F. Risk assessment and mitigation at Copahue volcano. in *Copahue Volcano* (eds Tassi, F., Vaselli, O. & Caselli, A.) 239–254 (Springer, 2016).
52. Cabrera, L. et al. Eruption forecasting model for Copahue volcano (southern Andes) using seismic data and machine learning: a joint interpretation with geodetic data (GNSS and InSAR). *Seismol. Res. Lett.* **95**, 2595–2610 (2024).
53. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019).
54. Kratzert, F. et al. Toward improved predictions in ungauged basins: exploiting the power of machine learning. *Water Resour. Res.* **55**, 11344–11354 (2019).
55. Saunders, J. K., Minson, S. E. & Baltay, A. S. How low should we alert? Quantifying intensity threshold alerting strategies for earthquake early warning in the United States. *Earths Future* **10**, e2021EF002515 (2022).
56. Minson, S. E. et al. The limits of earthquake early warning accuracy and best alerting strategy. *Sci. Rep.* **9**, 2478 (2019).
57. Martina, M. L. V., Todini, E. & Libralon, A. A Bayesian decision approach to rainfall thresholds based flood warning. *Hydrol. Earth Syst. Sci.* **10**, 413–426 (2006).
58. Katz, R. W. & Ehrendorfer, M. Bayesian approach to decision making using ensemble weather forecasts. *Weather Forecast* **21**, 220–231 (2006).
59. Viscusi, W. K. & Masterman, C. Anchoring biases in international estimates of the value of a statistical life. *J. Risk Uncertain.* **54**, 103–128 (2017).
60. Richardson, D. S. Economic value and skill. in *Forecast Verification: A Practitioner’s Guide in Atmospheric Science* 2nd edn (eds Jolliffe, I. T. & Stephenson, D. B.) 167–184 (Wiley, 2012).
61. Robinson, L. A., Hammitt, J. K. & Zeckhauser, R. J. Attention to distribution in U.S. regulatory analyses. *Rev. Environ. Econ. Policy* **10**, 308–328 (2016).
62. Mileti, D. S. & Sorensen, J. H. Communication of Emergency Public Warnings: A Social Science Perspective and State-of-the-art Assessment. Report No. ORNL-6609 (Oak Ridge National Laboratory, 1990).
63. Brown, S. K., Jenkins, S. F., Sparks, R. S. J., Odbert, H. & Auker, M. R. Volcanic fatalities database: analysis of volcanic threat with distance and victim classification. *J. Appl. Volcanol.* **6**, 15 (2017).
64. Feuillard, M. et al. The 1975–1977 crisis of la Soufrière de Guadeloupe (FWI): a still-born magmatic eruption. *J. Volcanol. Geotherm. Res.* **16**, 317–334 (1983).
65. Voight, B. The 1985 Nevado del Ruiz volcano catastrophe: anatomy and retrospection. *J. Volcanol. Geotherm. Res.* **44**, 349–386 (1990).
66. Doyle, E. E., McClure, J., Paton, D. & Johnston, D. M. Uncertainty and decision making: volcanic crisis scenarios. *Int. J. Disaster Risk Reduct.* **10**, 75–101 (2014).

67. Sharples, J. J., McRae, R. H., Weber, R. O. & Gill, A. M. A simple index for assessing fire danger rating. *Environ. Model. Softw.* **24**, 764–774 (2009).
68. Potter, S. H., Jolly, G. E., Neall, V. E., Johnston, D. M. & Scott, B. J. Communicating the status of volcanic activity: revising New Zealand's volcanic alert level system. *J. Appl. Volcanol.* **3**, 13 (2014).
69. Sawada, Y., Kanai, R. & Kotani, H. Impact of cry wolf effects on social preparedness and the efficiency of flood early warning systems. *Hydrol. Earth Syst. Sci.* **26**, 4265–4278 (2022).
70. Dow, K. & Cutter, S. L. Crying wolf: repeat responses to hurricane evacuation orders. *Coast. Manag.* **26**, 237–252 (1998).
71. Simmons, K. M. & Sutter, D. False alarms, tornado warnings, and tornado casualties. *Weather Clim. Soc.* **1**, 38–53 (2009).
72. Trainor, J. E., Nagele, D., Philips, B. & Scott, B. Tornadoes, social science, and the false alarm effect. *Weather Clim. Soc.* **7**, 333–352 (2015).
73. Lim, J. R., Liu, B. F. & Egnoto, M. Cry wolf effect? Evaluating the impact of false alarms on public responses to tornado alerts in the southeastern United States. *Weather Clim. Soc.* **11**, 549–563 (2019).
74. Lazzarin, T., Viero, D. P., Molinari, D., Ballio, F. & Defina, A. A new framework for flood damage assessment considering the within-event time evolution of hazard, exposure, and vulnerability. *J. Hydrol.* **615**, 128687 (2022).
75. Regnier, E. & Harr, P. A. A dynamic decision model applied to hurricane landfall. *Weather Forecast* **21**, 764–780 (2006).

Acknowledgments

We gratefully acknowledge GeoNet (New Zealand) for providing access to monitoring data. We also thank our Japanese collaborators and the staff of the Ontake Volcano Observatory for providing data and supporting this research.

Author contributions

A.A. developed the implementation of the Potential Economic Value (PEV) framework, performed the analyses, and led the writing of the manuscript. A.A. and D.D. developed the machine learning pipeline¹⁹, which includes the design and programming, and writing of the manuscript. G.M., and S.C. contributed to the analysis, interpretation and contributed to manuscript development. All authors contributed to the manuscript.

Funding

Alberto Ardid, David Dempsey, and Shane Cronin acknowledge support through the MBIE Endeavor Research Programmes E7370 – Transitioning Taranaki to a Volcanic Future and E7774 – Adapting to Climate Change

Through Stronger Geothermal Enterprises, both led by SC in the University of Auckland.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-77242-4>.

Correspondence and requests for materials should be addressed to Alberto Ardid.

Peer review information *Nature Communications* thanks Albert Lamberte and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.