JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

Abstract

Creative Society Design Research Area

Graduate School of Advanced Science and Technology

Knowledge Science

A Framework for Forecasting High-Impact, Low-Probability Events: A Case Study of Volcanic Tsunami

By Elmo Juanara

High-Impact, Low-Probability (HILP) disasters are infrequent but extreme events with the capacity to cause substantial harm, requiring proactive measures to mitigate their unpredictable and devastating effects. The impacts of HILP events frequently cascade across sectors, resulting in secondary and tertiary effects. Recent events illustrate the profound challenges posed by HILP disasters. The 2018 Anak Krakatau tsunami struck without warning, highlighting the limitations of current early warning systems in mitigating sudden impacts. On a global scale, the 2020 COVID-19 pandemic showcased how a low-probability event could lead to catastrophic societal and economic disruptions. These examples highlight the devastating impact of HILP events, which is often exacerbated by a lack of preparedness compared to similar-strength events occurring in high-probability zones or predictable timings. Consequently, research on HILP events forecasting is both urgent and essential for advancing disaster preparedness and prevention. The aim of this study is to enhance disaster preparedness and prevention strategies for HILP events. Objectively, proposing a novel framework for forecasting methods and strengthening risk management models to minimize HILP events impact. The significance lies in advancing disaster research by providing the forecasting methods to mitigate the impacts of such unpredictable and devastating disasters.

The methodology for developing the forecasting framework is structured into three phases: Conceptualization, Simulation and Database Generation, and Analysis and Implementation.

Conceptualization. The conceptualization phase establishes the theoretical basis for the framework by critically examining the nature of HILP events. A comprehensive review of significant historical HILP disasters is conducted to identify their characteristics, impacts, and gaps in current preparedness and prevention strategies.

Simulation and Database Generation. Accurate prediction of HILP events necessitates a comprehensive dataset encompassing a wide range of possible conditions. However, due to the infrequent nature of such events, data collection through field surveys or historical records is often inadequate. To address this limitation, simulations are conducted to systematically generate diverse event scenarios by considering various parameters. The resulting simulation outputs are validated against empirical data from historical or real events to ensure their

reliability and accuracy. Once validated, these results are organized into a precomputed database, providing a robust foundation for further phase of analysis and implementation.

Analysis and Implementation. The analysis and implementation phase leverages the precomputed database to develop predictive and decision-making capabilities through following stages. Scenario categorization: Data mining and clustering techniques are applied to group the simulated scenarios based on shared characteristics, enabling efficient pattern recognition and rapid event matching. Predictive modeling: Advanced machine learning techniques are employed to predict complete hazards using limited input from observations. This approach enables the framework to generate full forecasts in real-time, significantly enhancing early warning capabilities. Classification of warning levels: Rule-based reasoning is applied to classify the warning level HILP. These thresholds are calibrated to align with local standards and practical evacuation applications. Performance evaluation: Established metrics from the machine learning domain were used to evaluate the performance of prediction capability, including validation loss, root mean square error (RMSE), F1 score, accuracy, and confusion matrices.

The proposed framework was implemented and evaluated through a case study of the 2018 Anak Krakatau volcanic tsunami, a quintessential example HILP event. Volcanic tsunamis are selected as the case study, owing to their significant threat to rapidly growing coastal populations, their sudden onset without seismic precursors, and the inherent challenges in issuing timely warnings. The designed framework was used to address the forecasting gaps by integrating scenario analysis, predictive modeling, and classification techniques. A total of 1,000 hypothetical tsunami scenarios were simulated by varying collapse parameters, such as collapse volume, dip angle, and direction on coastal and synthetic observation stations. These simulations formed the foundation of a precomputed database. Validation against historical data ensured the simulated scenarios accurately reflected real-world dynamics. The framework integrated the static nature of the pre-computed database with the dynamic demands of real-time disaster response. When new event data, such as the first 9 minutes of tsunami synthetic observations, were received, the system rapidly matched the data to the most similar scenario in the database. To provide actionable insights on coastal observation stations into predefined warning levels: “No Warning,” “Minor Tsunami,” “Tsunami,” and “Major Tsunami,” based on wave height thresholds.

The case study provides an evaluation of the proposed framework's performance. Various machine learning algorithms were applied to predict the maximum amplitude waves at coastal observation stations. Random Forest proved to be the most reliable model, performing best at three out of four coastal stations. Gradient Boosting also showed strong performance at one station but faced challenges in real-time applications due to its computational demands. Neural Networks, while flexible, were less suitable for real-time use because of their high resource requirements and frequent large errors. For waveform forecasting, the LSTM Test model, using full observation data, achieved the lowest error metrics (MSE: 0.0153, MAE: 0.0965), outperforming its more complex variant. However, with shorter input durations (9 minutes), all models experienced significant performance declines, underscoring the impact of data constraints. Additionally, a decision tree model trained on simulated scenarios successfully classified warning levels with more than 90% accuracy, effectively categorizing events based on wave height thresholds.

The results highlight the potential of integrating precomputed databases with machine learning techniques to enhance forecasting and decision-making for HILP events. The proposed framework contributes to a structured methodology for managing tsunami risks, enabling accurate predictions, effective classification, and timely warning dissemination to mitigate impacts on coastal communities. By generating full forecasts from limited initial observations, the framework demonstrates adaptability to generate early warning with diverse scenarios and conditions.

The method has limitations in its current implementation, as it is specifically tailored to water-related HILP hazards, relying on specific types of data structures. Future work may focus on expanding the framework's scope to include multi-hazard scenarios by integrating diverse data sources. Further efforts could involve adapting the framework to address non-water-related HILP disasters, such as pandemics or nuclear events, which require different dynamics and data structures.