



Forecasting strong aftershocks in New Zealand with the machine-learning NESTORE algorithm: two different testing approaches

Letizia Caravella¹ and Stefania Gentili²

¹National Institute of Oceanography and Applied Geophysics, Udine, Italy (lcaravella@ogs.it)

²National Institute of Oceanography and Applied Geophysics, Udine, Italy (sgentili@ogs.it)

We applied the machine-learning-based probabilistic forecasting algorithm NESTORE (NEXt STRong Related Earthquake) to the seismicity of New Zealand. NESTORE analyses nine features describing aftershock occurrence, source area evolution, and temporal trends in magnitude and radiated energy, computed over progressively increasing time windows following the mainshock. These features enable the algorithm to estimate the probability that a mainshock of magnitude M_m will be followed by a subsequent event of magnitude $\geq M_m - 1$ within the space-time domain of the associated seismic cluster. Clusters in which such a strong aftershock occurs are classified as "Type A," indicating higher potential hazard, while others are classified as "Type B." For each cluster, the algorithm outputs the corresponding probability of belonging to Type A.

New Zealand's position along the boundary between the Australian and Pacific plates results in widespread, complex deformation and a relevant seismic activity, including major events up to magnitude 7.8. This setting makes the region an ideal testing ground for operational, data-driven forecasting tools such as NESTORE. Understanding and forecasting seismic activity is critical for rapid hazard assessment and mitigation efforts.

To evaluate NESTORE's performance, we employed two testing strategies. The first was a chronological approach, in which the algorithm was trained using seismic clusters occurring before a chosen cutoff time and then used to retrospectively forecast cluster behaviour after that time. The second approach employed stratified k-fold cross-validation, allowing us to assess model generalization across multiple randomized data partitions. To further enhance training quality, we applied the outlier-detection procedure REPENESE (RElevant features, PErcentage class weighting, NEighborhood detection and SElection).

Our results show that the k-fold validation approach provides a more robust and stable performance evaluation than the chronological approach, although changes in the catalogue may make the more recent clusters a more reliable test set. NESTORE correctly classified 88% of seismic clusters 18 hours after the mainshock, including 77% of Type A clusters and 92% of Type B clusters. Notably, the Canterbury/Christchurch 2010–2011 sequence, a critical and highly destructive Type A cluster, was correctly classified by the algorithm.

Overall, the results of this work underscore the potential for use of NESTORE for short-term

aftershock forecasting in New Zealand.