Vienna, Austria | 3-8 May 2026
NH4.8
Machine Learning and Statistical Models Applied to Earthquake Occurrence

Sharing is encouraged

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

L. Caravella[1] and S. Gentili[1]
[1]National Institute of Oceanography and Applied Geophysics - OGS

Method

Context:

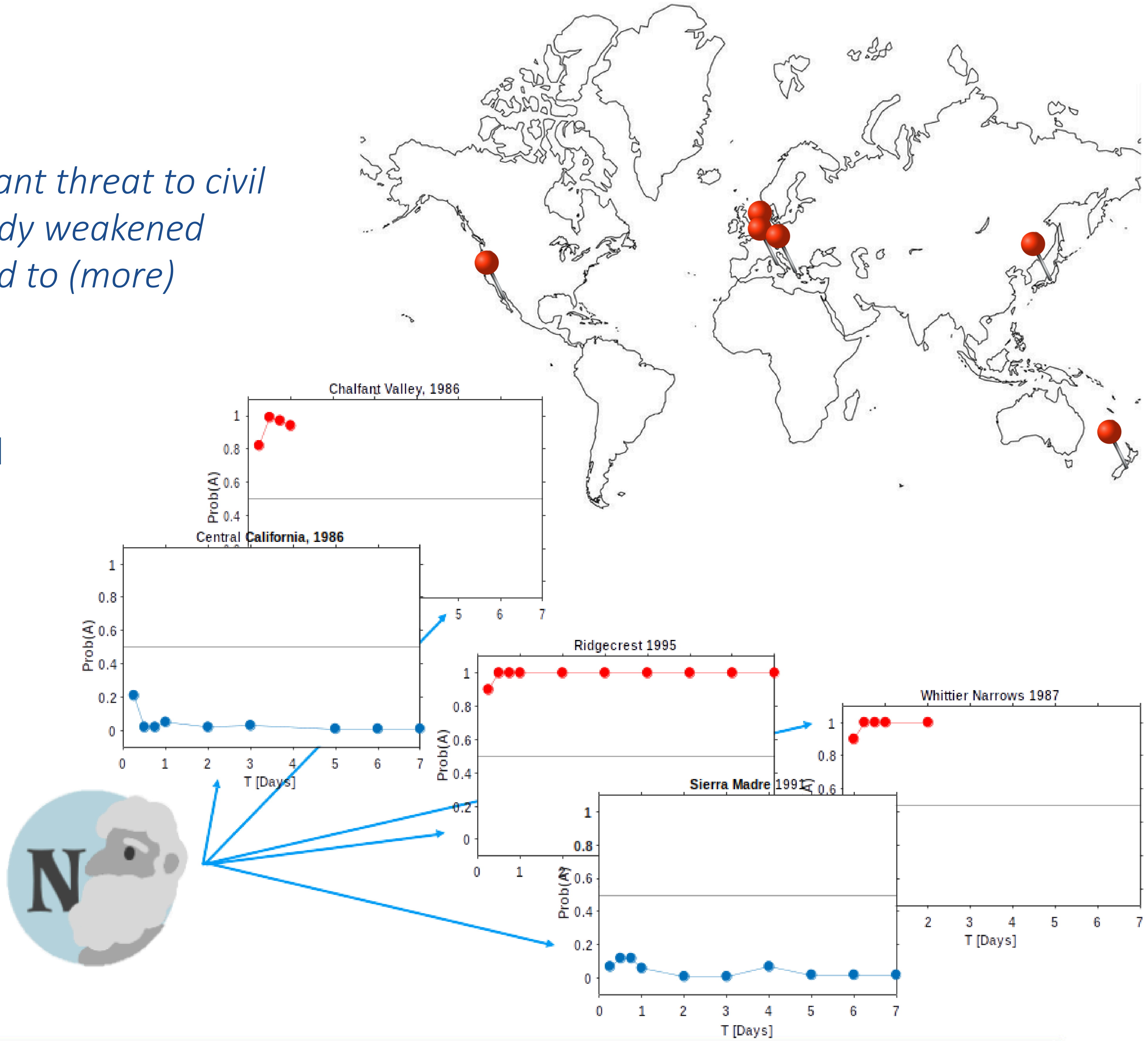
Large earthquakes following a strong event pose a significant threat to civil protection because they cause additional damage to already weakened buildings, raising the risk of building collapse, and may lead to (more) fatalities.

NESTORE – (NExt STrOng Related Earthquake) is an algorithm based on machine learning approach developed by Gentili et. al (2023), available on GitHub (v1.0), ***soon to be updated: see EGU PICO Gentili et al. 2026***

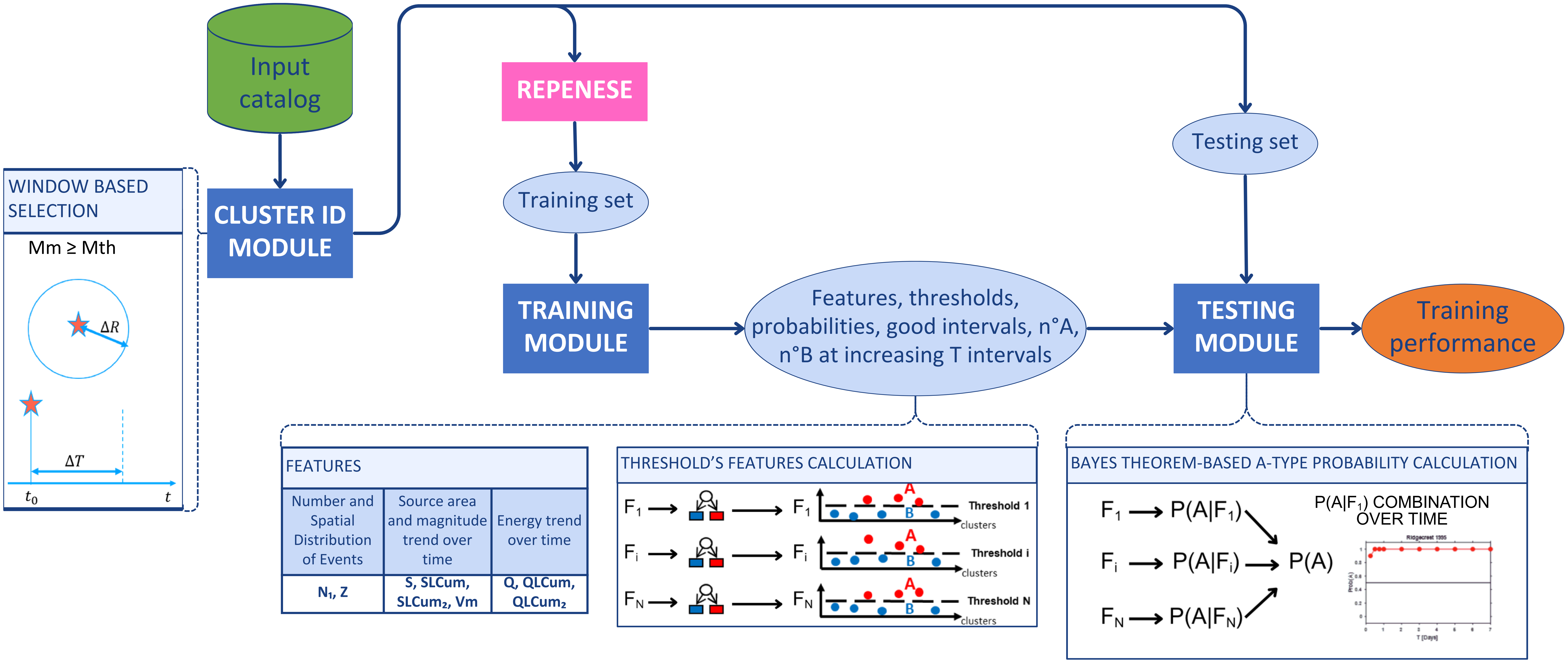
Two classes based on the difference in magnitude Dm between the **o-mainshock** (the first large shock e.g. M>4) and the **strongest following earthquake** (**Type A: Dm≤1**; **Type B: Dm>1**)

Goal: **Type A clusters probabilistic forecasting** based on features extracted from seismic catalogues in the first hours/days after the o-mainshock

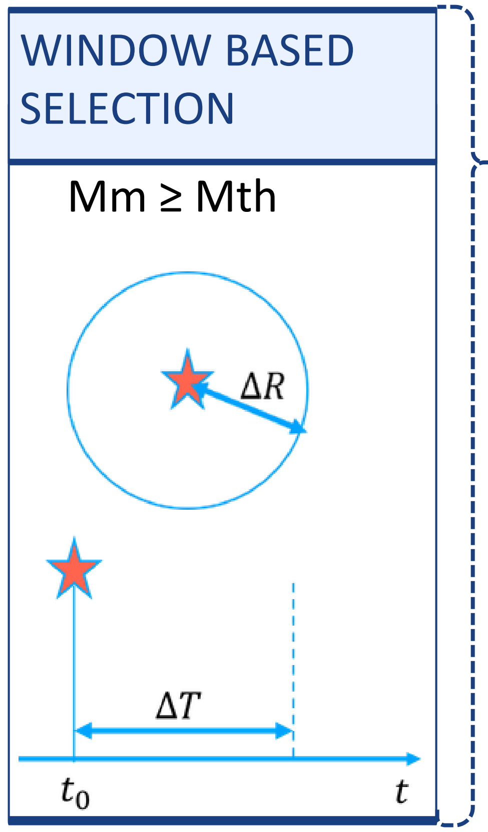
Already applied for seismicity of: Italy, NE Italy, Greece, California, Japan.

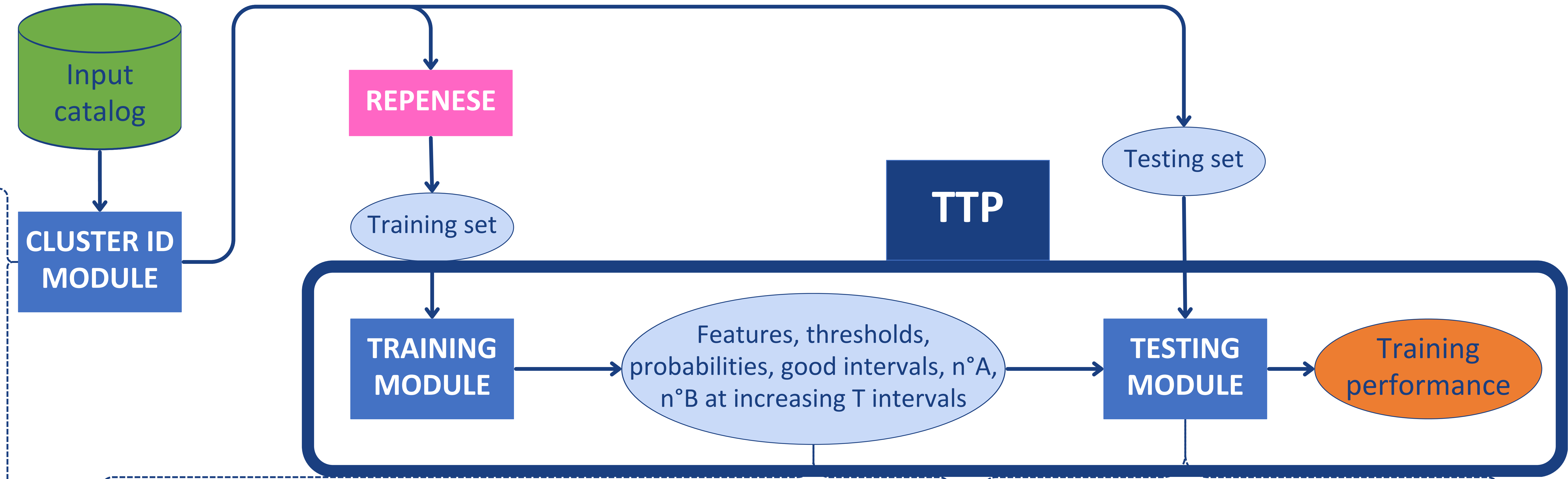


Method — Chronological cluster selection



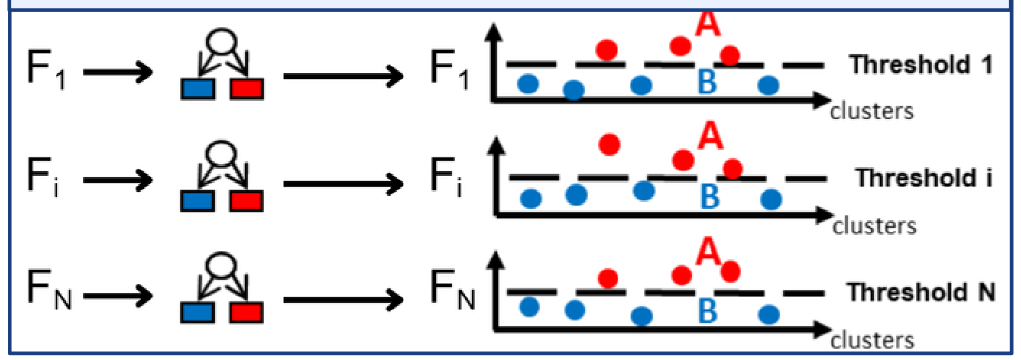
Method — Chronological cluster selection



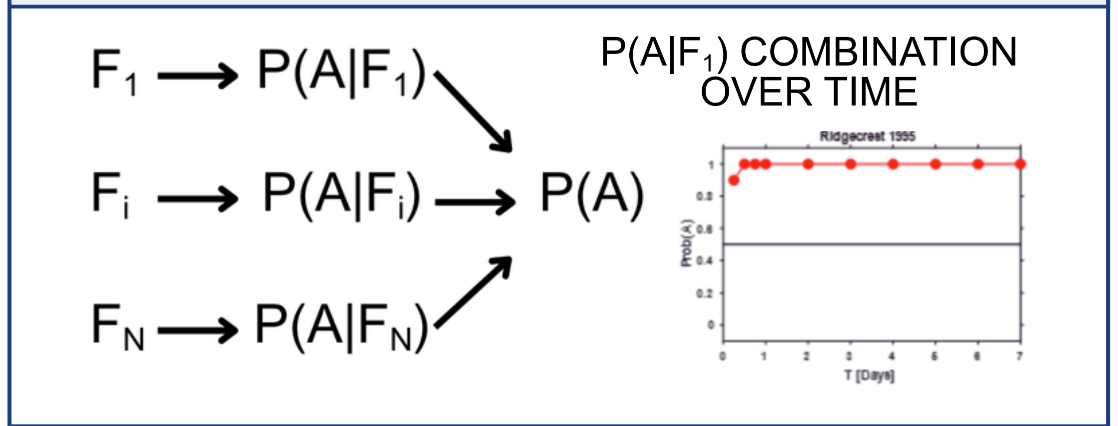


FEATURES		
Number and Spatial Distribution of Events	Source area and magnitude trend over time	Energy trend over time
N_1, Z	**S, SLCum, $SLCum_2$, Vm**	**Q, QLCum, $QLCum_2$**

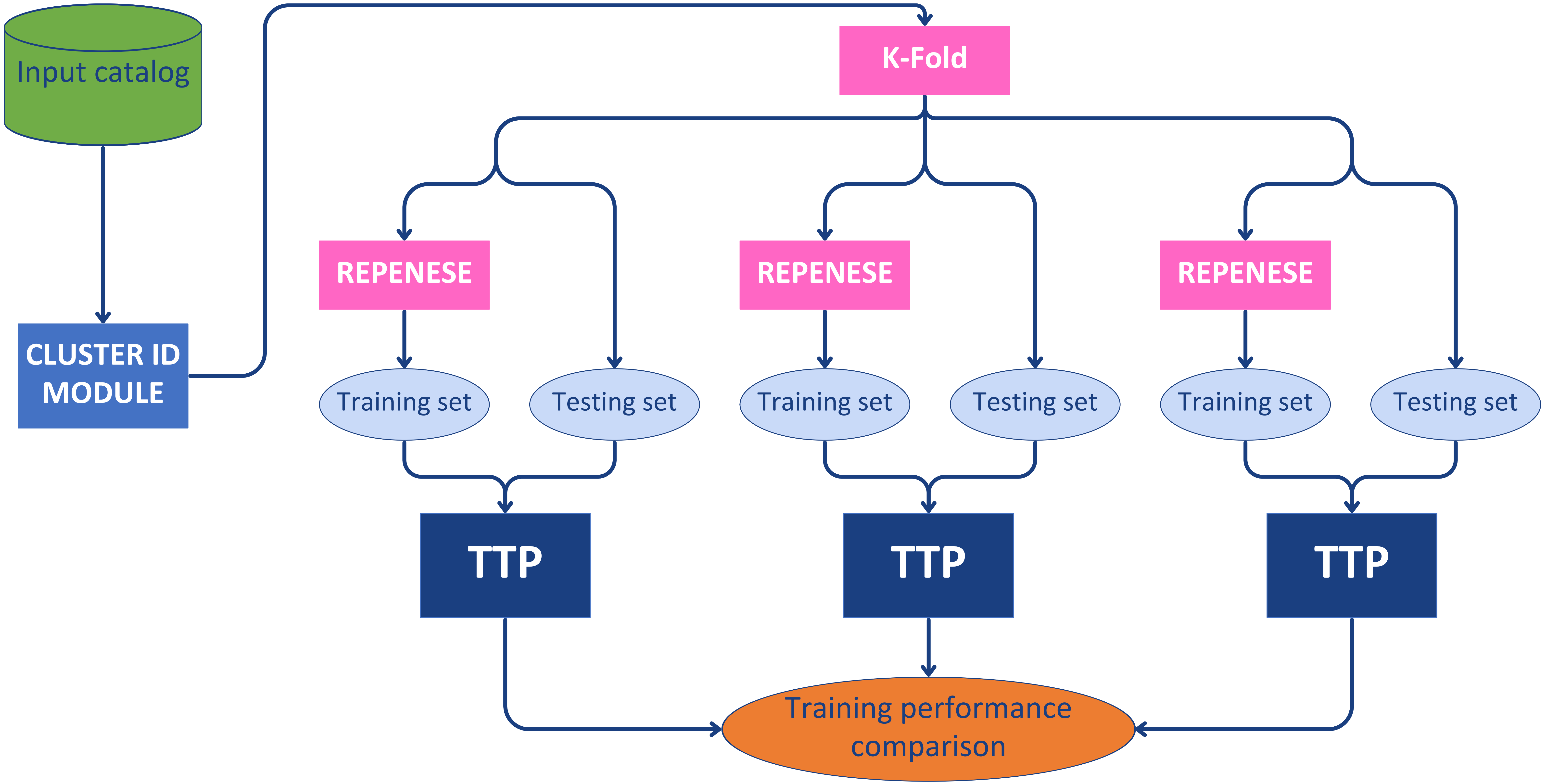
THRESHOLD'S FEATURES CALCULATION



BAYES THEOREM-BASED A-TYPE PROBABILITY CALCULATION



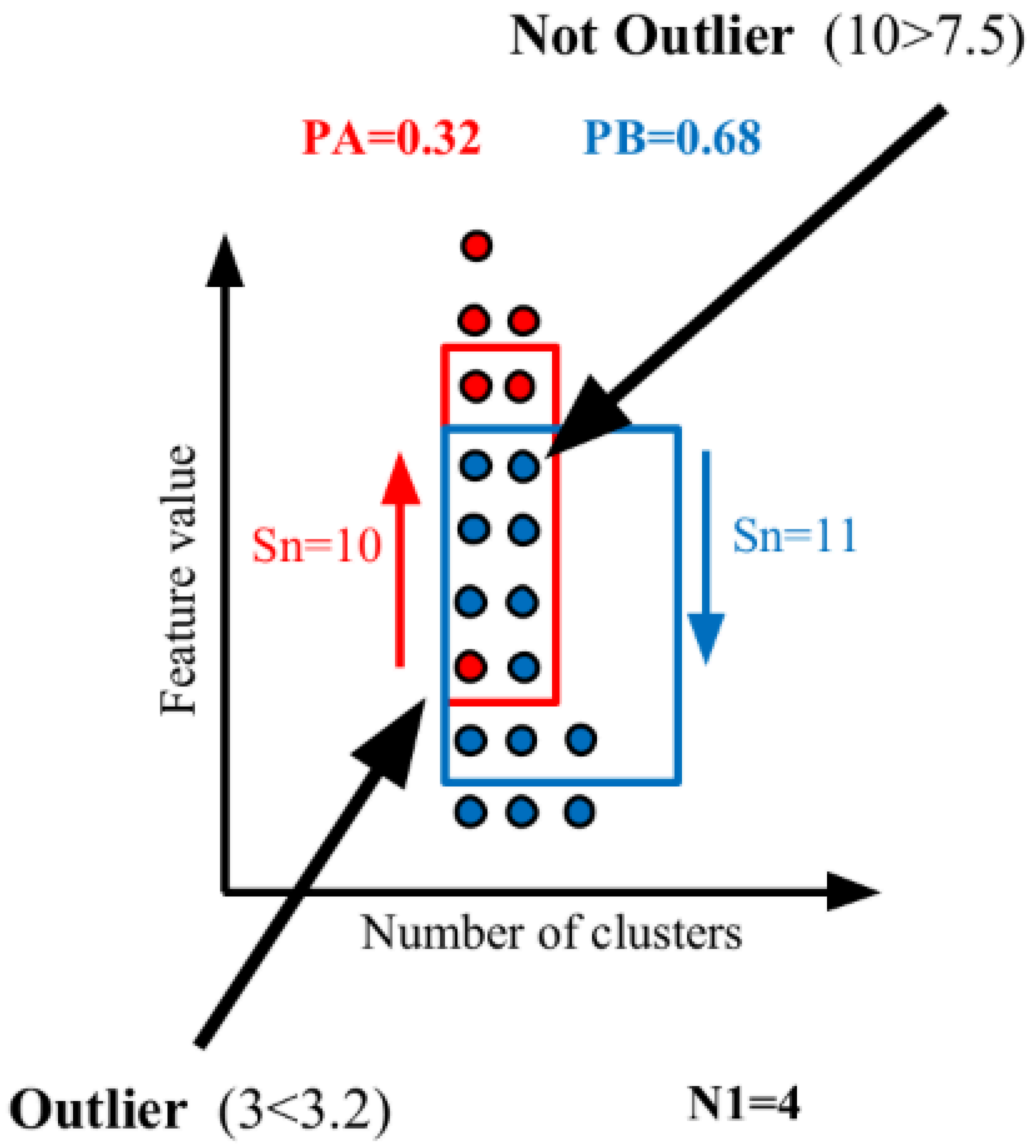
Method — k-fold cluster selection



Method — REPENESE

Algorithm for **outlier removal** (Gentili et al. 2025).
Needed to account for **class imbalance** e **skewed distribution** of class A.

- **RE** = **RElevant features** selection based on TPR, FPR, Precision (>0.5, <0.5, >0.5)
- **PE** = **class imbalance PErcentage**. Percentage of A and B type samples PA and PB for A and B classes
- **NE** = **Neighborhood detection**. For each sample of type A, the first N1 independent larger feature values are determined, the neighborhood for the features is the set of all samples that have these feature values. For type B samples, the first N1 smallest independent feature values.
- **SE = SElection**. A sample is considered a possible outlier if N≤P·Sn where P = PA or PB, depending on the sample class. Outliers are only those samples that are common for all relevant features and for all different Ti.



Sn=# of samples in the neighbor
N = # of samples of the same class in the neighborhood

Method — Stratified k-fold cross validation

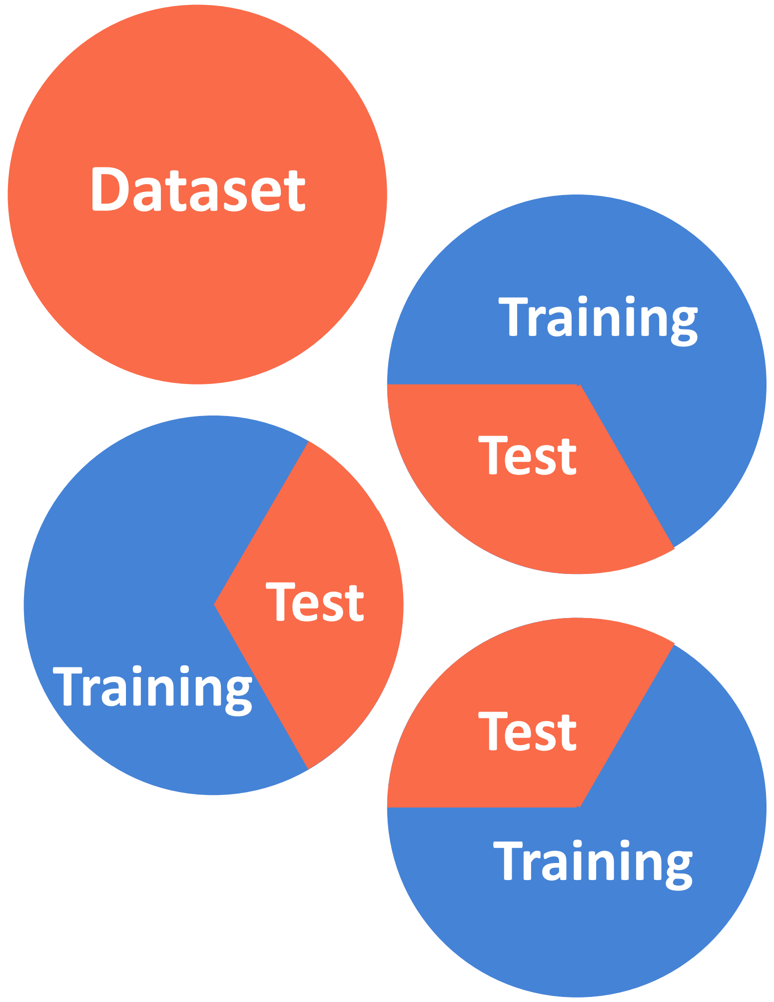
Goal
Evaluate model performance reliability using all available clusters.

Method:

- Split the valid dataset into **K equally sized folds** (here, K = 3)
- **Balanced class distribution** in each fold (stratified split)
- Each fold is used once as the test set and K – 1 times as part of the training set: **no overlap between test sets**
- matlab cvpartition(..., 'KFold', K)

Outcome:

- Every cluster is tested exactly once
- Training uses the complementary data for that fold
- Provides a **more robust estimate of performance** than a single random split



Dataset — about New Zealand

1988-2025, how and why:

- **ML** from GNS Science (2022), up to 2020
- following years from online database (GNS Science, 1970), ML from orthogonal regression where missing

- **Shallow seismicity**: max 50 km depth

Clustering: **M-dependent, window based**:

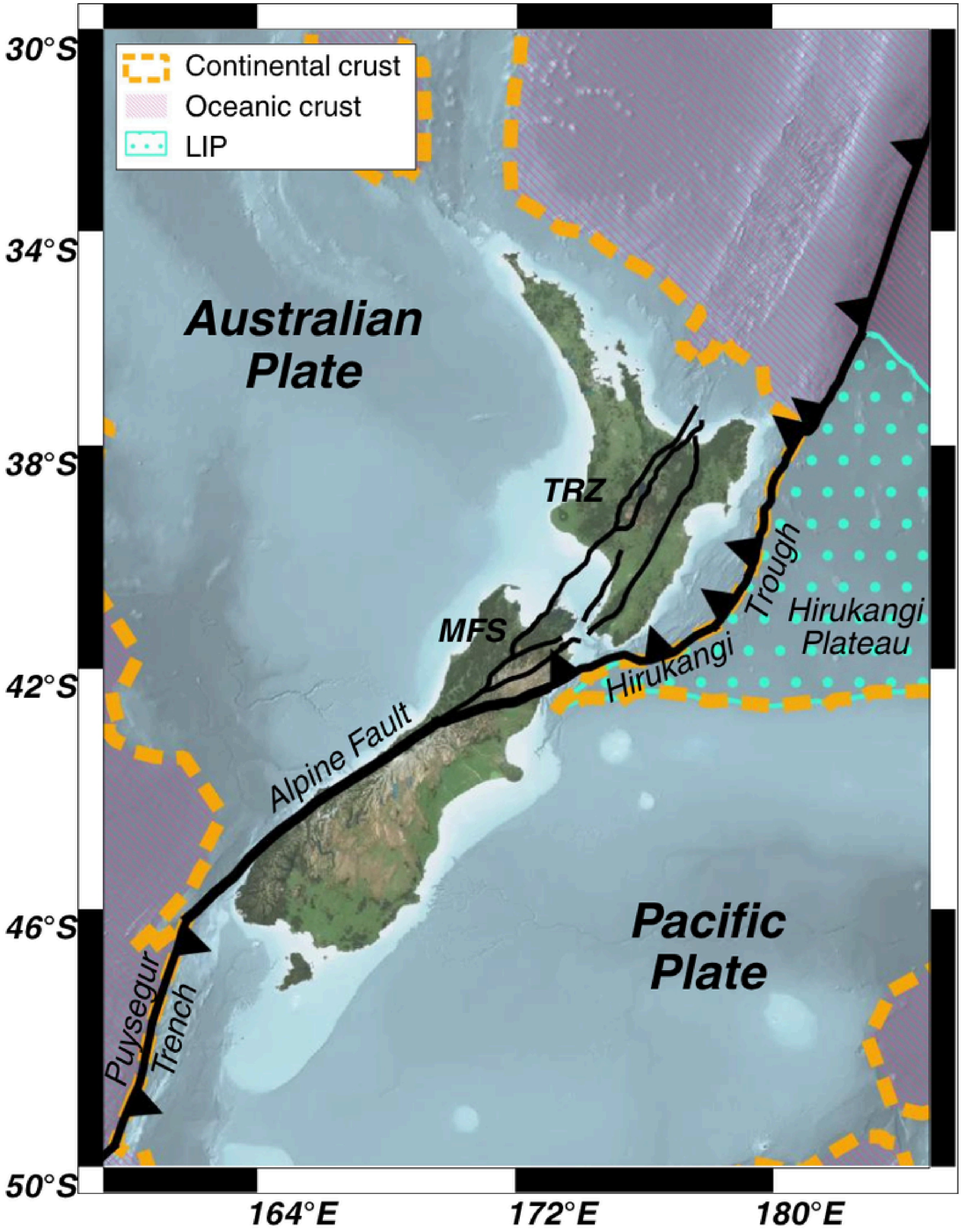
- time by Gardner and Knopoff (1974)
- space by Christophersen (2006)

52 clusters, 14 A, 38 B

Chronological division training-testing:

- Training: 1988-2013
- Testing: 2014-2025 (retrospective forecasting)

k-fold: 3 folds



Method — k-fold cluster selection

1988-2025, how and why:
- **ML** from GNS Science (2022), up to 2020
- following years from online database (GNS Science, 1970), ML from orthogonal regression where missing

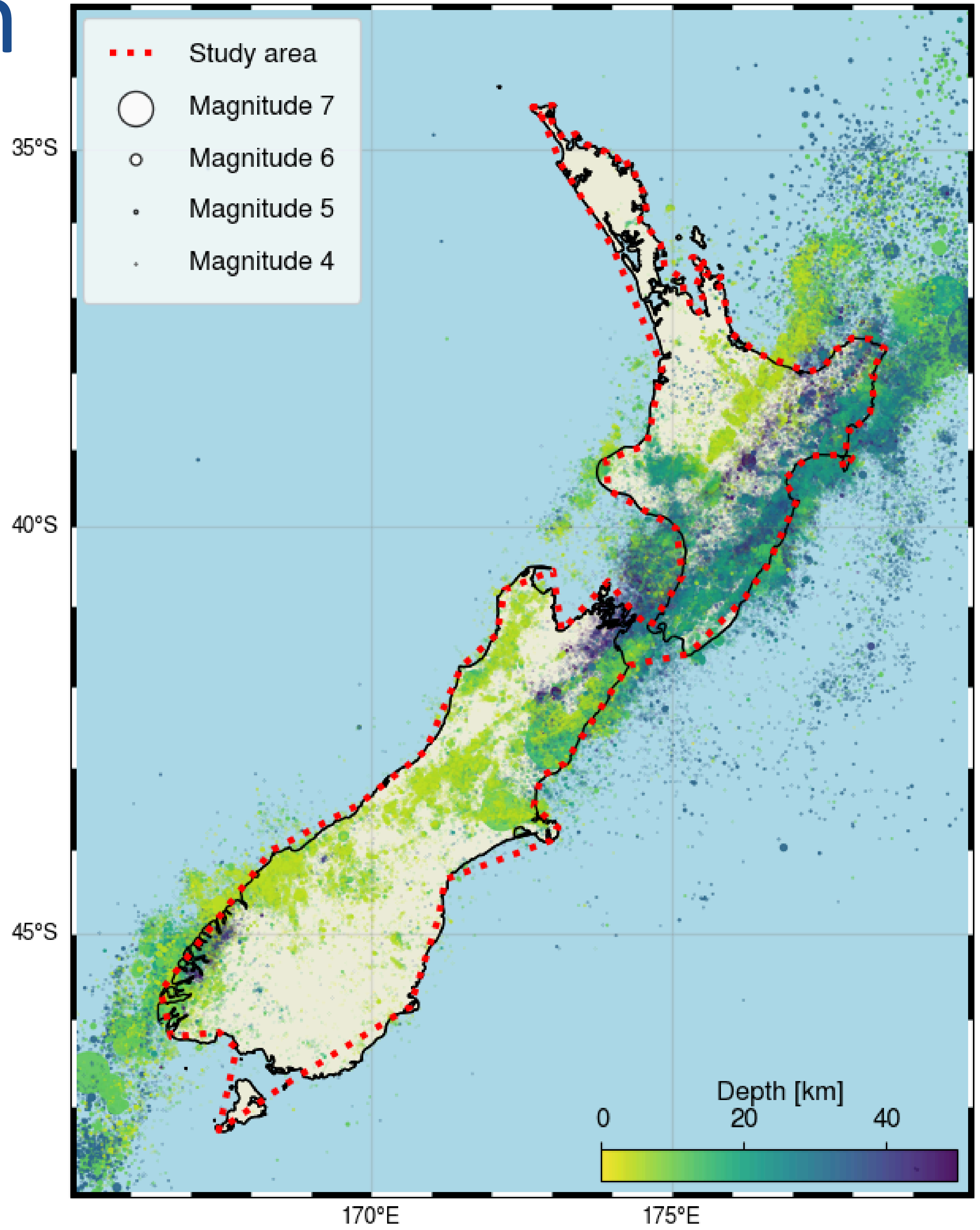
- **Shallow seismicity**: max 50 km depth

Clustering: **M-dependent, window based**:
- time by Gardner and Knopoff (1974)
- space by Christophersen (2006)

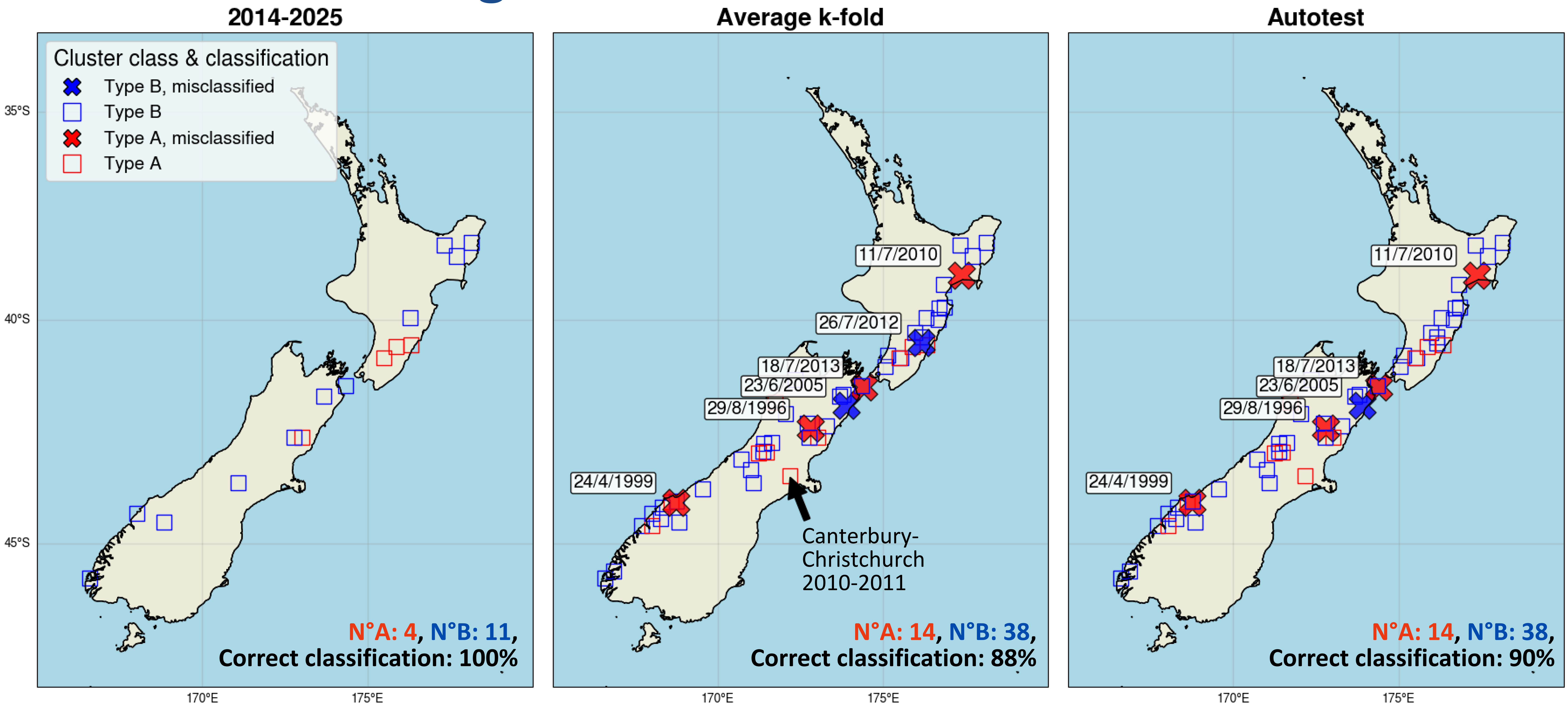
52 clusters, 14 A, 38 B

Chronological division training-testing:
- Training: 1988-2013
- Testing: 2014-2025 (retrospective forecasting)

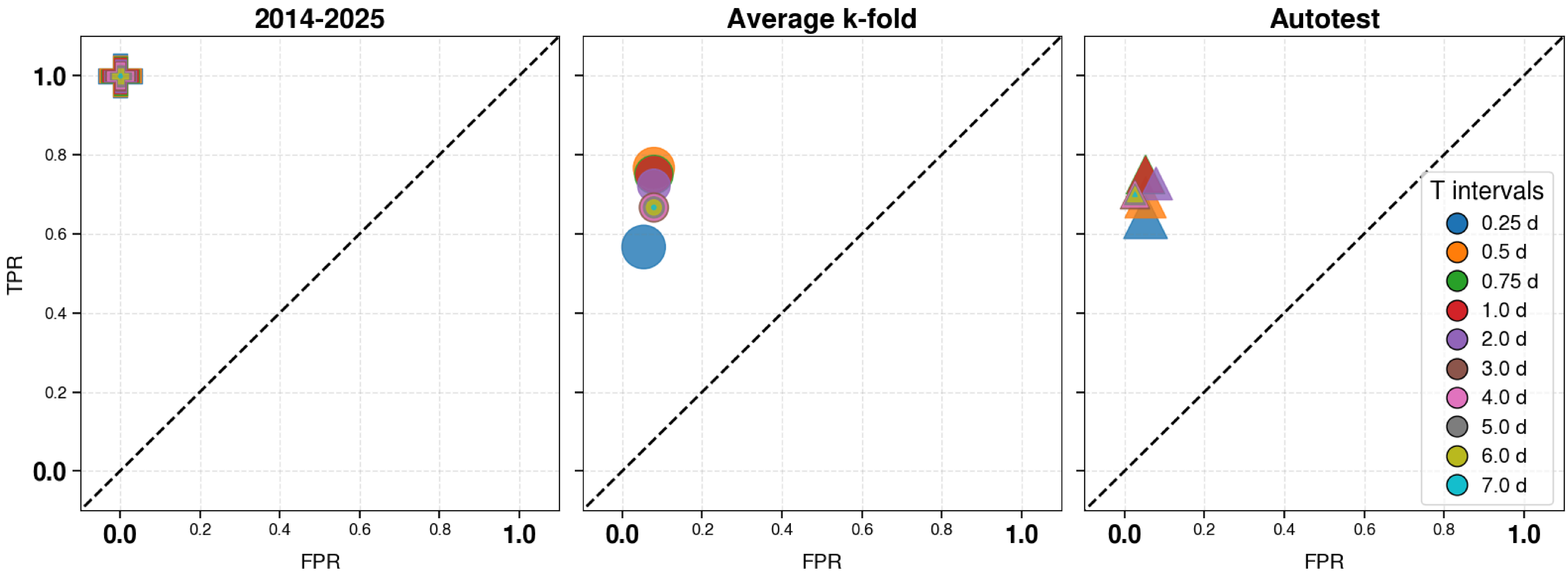
k-fold: 3 folds



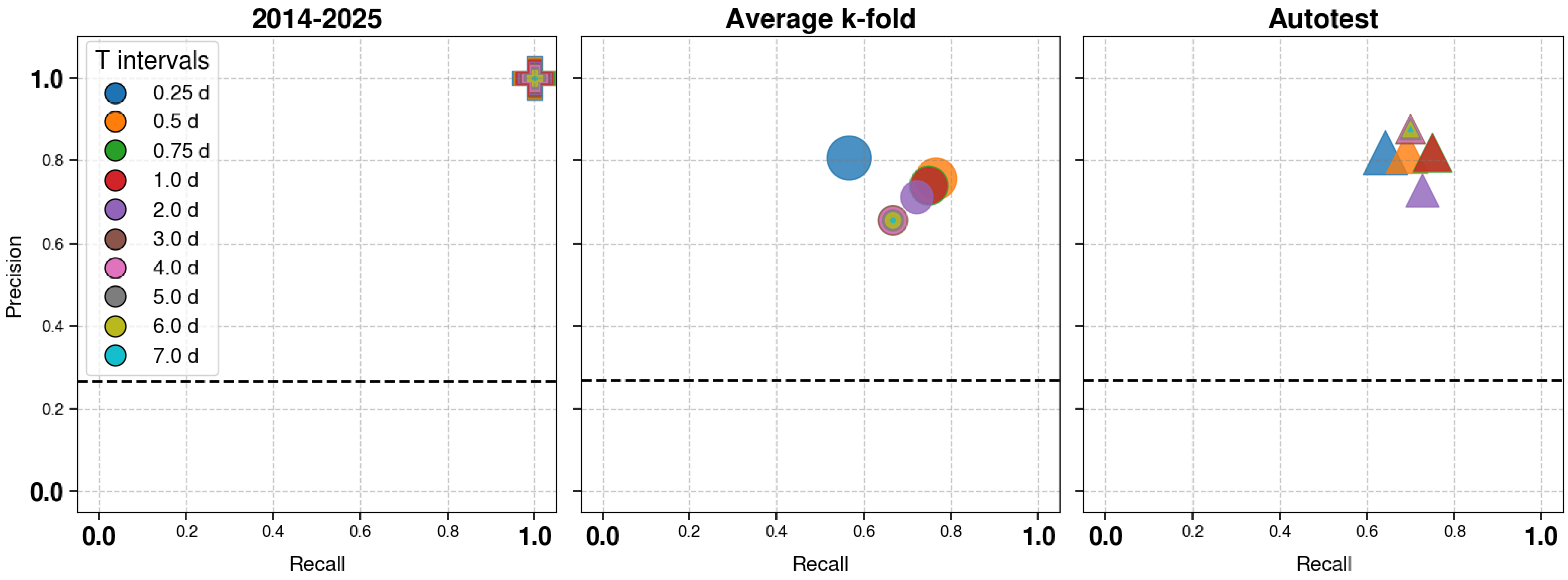
Results — Testing



Results — ROC diagrams



Results — Precision-Recall diagrams



Conclusions

Application of NESTORE to NZ seismicity to forecast type A clusters

- Dataset issues:
  - small dataset
  - classes imbalance
  - skewed distribution
- NESTORE solutions:
  - one-node decision tree statistically selected
  - new approach for outlier removal: REPENESE

- Chronological training-testing division (2014-2025): 100% correct classification
- k-fold cross validation (average) at 12h:
  - 88% correct classification
  - TPR=0.77, FPR=0.08, PRE=0.76
  - performances influenced by large number of outliers and outliers distribution within each fold

- Canterbury-Christchurch 2010-2011 sequence correctly classified as Type A

Thank you for your attention!

lcaravella@ogs.it

Caravella, L., & Gentili, S. (2026). Machine Learning Forecasting of Strong Subsequent Events in New Zealand Using the NESTORE Algorithm. *Forecasting, 8(1)*, 16. https://doi.org/10.3390/forecast8010016

References

Christophersen, A. (2006). T*owards a New Zealand model for short-term earthquake probability: Aftershock productivity and parameters from global catalogue analysis*. Natural Hazards Commission Tōka Tū Ake.

Gardner, J. K., & Knopoff, L. (1974). Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian? Bulletin of the Seismological Society of America, 64(5), 1363–1367.

Gentili, S., Brondi, P., & Di Giovambattista, R. (2023). NESTOREv1.0: A MATLAB package for strong forthcoming earthquake forecasting. *Seismological Research Letters, 94*(4), 2003–2013.

Gentili, S., Caravella, L., & Chiappetta, G. D. (2026). An enhanced version of the NESTORE software for strong aftershock forecasting. *EGU General Assembly 2026*, Vienna, Austria. https://doi.org/10.5194/egusphere-egu26-7901

Gentili, S., Chiappetta, G. D., Petrillo, G., Brondi, P., & Zhuang, J. (2025). Forecasting strong subsequent earthquakes in Japan using an improved version of NESTORE machine learning algorithm. *Geoscience Frontiers, 16*(3), Article 102016.

GNS Science. (1970). *New Zealand earthquake catalogue* [Data set]. GNS Science, GeoNet. https://doi.org/10.21420/0S8P-TZ38

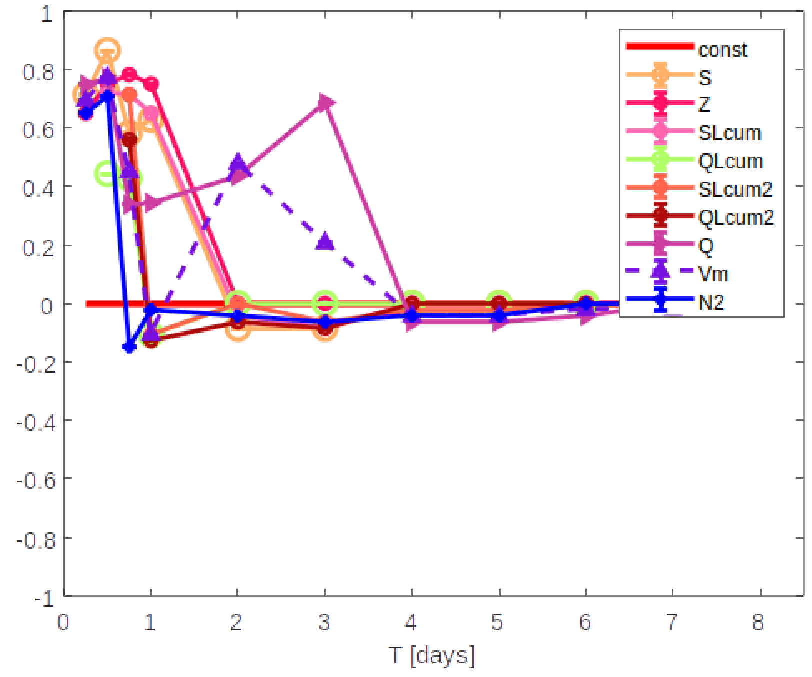
Training module — Classifier(feature) quality

<u>How “good” is the feature?</u>

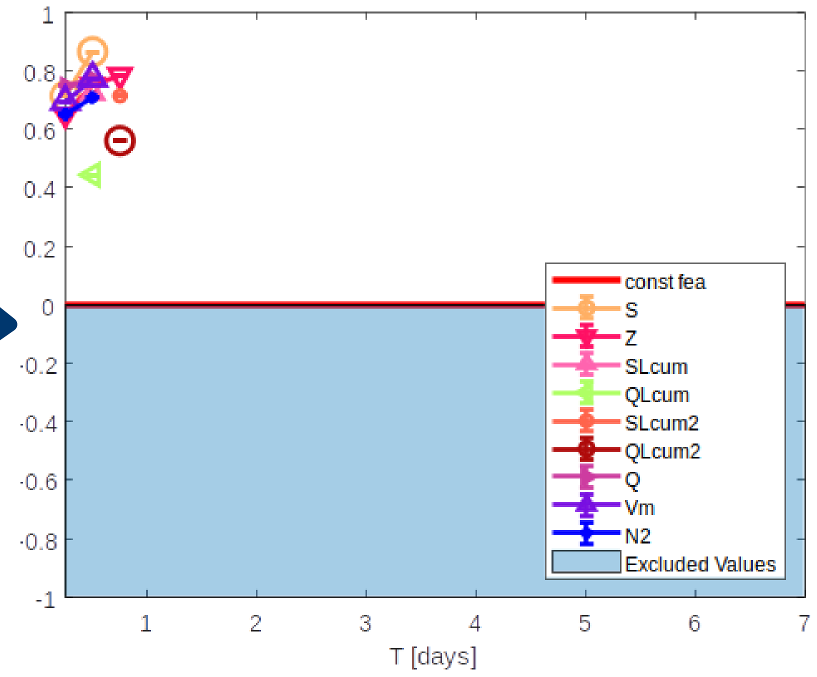
At each Ti, NESTORE estimates the performances by a **LOO method** and **selects “good intervals”** in which:

- Accuracy, Precision, and Recall > 0.5;
- Accuracy >= the one obtained by the constant response B;
- Informedness > 0.

Informedness **before** Ti selection



Informedness **after** Ti selection



<u>How “good” is the threshold?</u>

The **Gini gain** helps in understanding how much the threshold affects the classification of the positive class (type A), i.e. how much the classes are mixed (**Gini impurity**) before and after the threshold application.

NB: This parameter at the moment is not used to select the features but only for quality evaluation.

About the results — Outliers distribution

- Anomalous clusters are generally well separated in both space and time
- most occurred **before 2012**
- possible interpretation: outliers associated with catalogue inaccuracies between 1988–2011

We assume anomalous clusters

1. occur randomly and independently,
2. have a constant probability over time
3. do not influence one another.

The occurrence of anomalous clusters can be described by a Poisson distribution

- null hypothesis H0 (outliers timing due to chance) vs (hypothesis H1 independent causes)

1988-2011: 24y, 4 outliers
2012-2025: 13.37y, 1 outlier

annual mean: 0.17 outliers/year
μ = 2.23 outliers in 13.39

Probability of having *x=1* outlier: $P(X = x) = \frac{\mu^x e^{-\mu}}{x!} = e^{-\mu} = 0.24$

hypothesis H0 can’t be rejected at 5% level
i.e. occurrence of outliers statistically compatible with the observed results.