Comparative Evaluation of Earthquake Forecasts: Application to Italy

Jonas Brehmer

Joint work with Kristof Kraus, Tilmann Gneiting, Marcus Herrmann, and Warner Marzocchi

(My) background

- T. Gneiting's HITS group working on forecast evaluation
- Evaluation based on scoring rules/functions is common in meteorology and economics
- (How) can we use them in a point process/spatial setting?

Consistent scoring functions

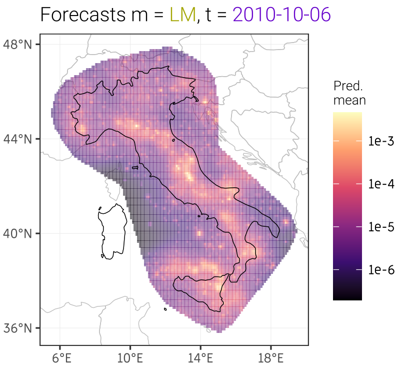
- Assign score/loss $\mathsf{S}(x, y)$, where x is the forecast and y is the outcome
- Consistency ensures that the true value minimizes the score/loss on average
- Example: S is consistent for the mean if

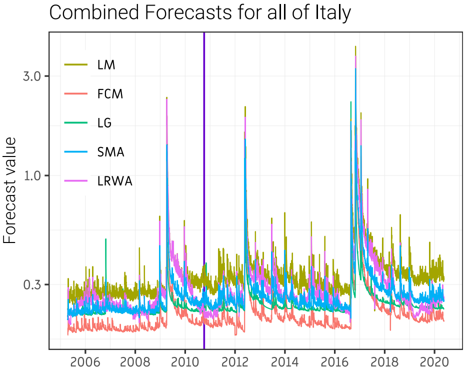
$$\mathbb{E}_{Y \sim F}\, \mathsf{S}(\text{mean}(F), Y) \leq \mathbb{E}_{Y \sim F}\, \mathsf{S}(x, Y) \quad \text{for all } x$$

- $\mathsf{S}(x, y) = (x - y)^2$ is consistent for the mean
- $\mathsf{S}(x, y) = |x - y|$ is not consistent for the mean (but for the median)

M4+ earthquakes in Italy

We get $\begin{cases} \text{predicted means of model LM, FCM, LG, SMA, LRWA} \\ \text{observed counts } (0, 1, 2, \ldots) \text{ of M4+ earthquakes} \end{cases}$





⇝ Which model issues the best predictions?

⇝ How well do the predictions represent reality?

Model Scores

Quadratic score

$$S(x, y) = (x - y)^2$$

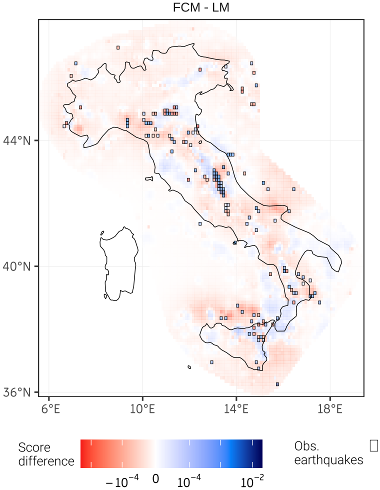
Poisson score (preferred)

$$S(x, y) = x - y \log x$$

Overall mean scores $\overline{S}$

Model	Poisson	Quadratic
LM	2.71	**0.841**
FCM	2.80	0.846
LG	3.02	0.847
SMA	2.73	0.844
LRWA	**2.69**	0.842

Poisson Score Difference



The Poisson Score $S(x, y) = x - y \log x$

- Commonly used in CSEP methodology as *Poisson log-likelihood*
- Consistent for the mean. Connection to Poisson distribution is purely formal
  - Forecasts x don't need to come from a Poisson model
  - Observations y don't have to be Poisson
- Evolve T-test (Rhoades et al. 2011) into Diebold-Mariano test of equal predictive performance (measured in terms of Poisson score)

Are forecasts and outcomes compatible?

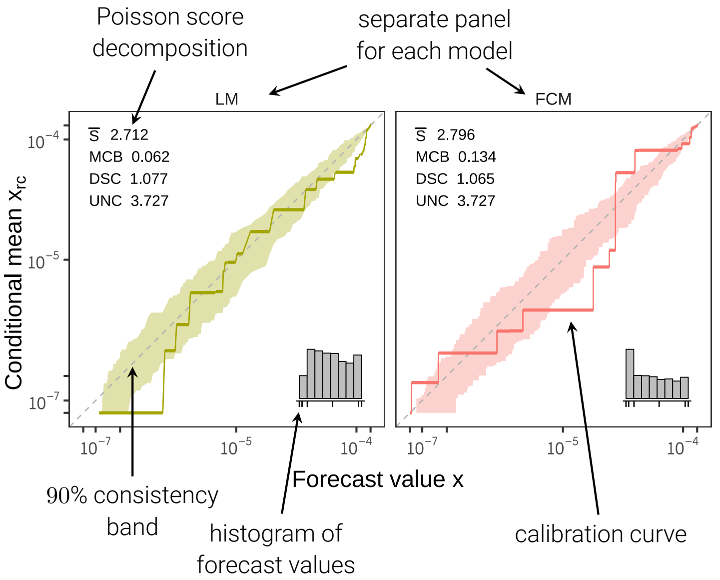
- Reliability quantifies how good forecasts represent outcomes
- The forecast X is mean-calibrated for the outcome Y if

  conditional mean of Y given X equals X

$$\mathbb{E}[Y|X] = X$$

- Estimate conditional mean X_{rc} and plot X_{rc} vs. X in a reliability diagram

Reliability Diagrams



High level overview

	CSEP Framework	vs.	New Proposal
Comparative	Poisson metric + heuristic statistical test	$\approx$	consistent scoring function + theory-based statistical test
Absolute	statistical tests	$\neq$	diagnostic tools reliability diagram
	often many assumptions		non-parametric weak assumptions

Summary

- Evaluate earthquake forecasts in the form of predicted mean counts
- The Poisson scoring function is convenient, no distributional assumptions are needed
- For forecasts in other formats (e.g. quantile forecasts), ‘simply’ change the scoring function

References

In review. Brehmer, J. R., Kraus K., Gneiting, T., Herrmann, M., and Marzocchi, W. (2024). Comparative evaluation of earthquake forecasting models: An application to Italy. Preprint, https://arxiv.org/abs/2405.10712.

Mathematical background (1). Brehmer, J. R., Gneiting, T., Herrmann, M., Marzocchi, W., Schlather, M., and Strokorb, K. (2024). Using scoring functions to evaluate point process forecasts. *Annals of the Institute of Statistical Mathematics*, 76, 47–71, https://doi.org/10.1007/s10463-023-00875-5.

Mathematical background (2). Gneiting, T. and Resin, J. (2023). Regression diagnostics meets forecast evaluation: Conditional calibration, reliability diagrams, and coefficient of determination. *Electronic Journal of Statistics*, 17, 3226–3286, https://doi.org/10.1214/23-EJS2180.