26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

When will it happen? On the evaluation of time-to-event, survival time and first-time passage forecasts

Robert J. Taggart [a], **Nicholas Loveday** [b], **Simon Louis** [c]

[a] *Research, Bureau of Meteorology, Sydney, Australia,* [b] *Research, Bureau of Meteorology, Melbourne, Australia,* [c] *Research, Bureau of Meteorology, Hobart, Australia*
Email: robert.taggart@bom.gov.au

Abstract: To improve planning and investment decisions, humans require some knowledge of whether a significant event will occur and if so, when. Time-to-event forecasts are useful in a variety of fields, such as emergency management (predicting the timing of flood peak or of river height exceedance over a critical threshold), engineering (predicting failure times of components), medicine (predicting patient life expectancy) and business (predicting timing of loan defaults).

Improving decision making involves evaluating competing time-to-event forecast procedures and assessing which procedure makes superior predictions. One challenge with time-to-event forecasts is that, within the evaluation period, the event may not yet have occurred, so it is not immediately clear how to evaluate the corresponding forecast. Excluding such instances can lead to biased evaluations and incorrect inferences regarding which forecast procedure is better.

This issue can be dealt with in a methodologically sound way by using so-called "threshold-weighted proper scoring rules" and "threshold-weighted consistent scoring functions". In the last few years, the threshold-weighted continuous ranked probability score (Gneiting & Ranjan 2013) has been used by two authors to assess time-to-event forecasts that are in the form of predictive distributions. We show that common forms of single-valued time-to-event forecasts, such as the median, mean and 90th percentile, can be soundly evaluated using the threshold-weighted scores of Taggart (2022). As an extension, we also show how time-to-event prediction intervals (e.g. the interquartile range) can be evaluated using "threshold-weighted interval scores".

The calibration of time-to-event forecasts is also challenging for similar reasons. We show that quantile forecasts can be readily recalibrated using isotonic regression applied to training data sets, even in the presence of forecast cases where the event has not yet occurred.

These methods are illustrated using forecasts for wind speed, tropical cyclogenesis and stochastic processes.

REFERENCES

Gneiting T and Ranjan R (2011) Comparing density forecasts using threshold-and quantile-weighted scoring rules. Journal of Business & Economic Statistics, 29(3), 411-422.

Taggart R (2022) Evaluation of point forecasts for extreme events using consistent scoring functions. Quarterly Journal of the Royal Meteorological Society, 148(742), 306-320.

Keywords: *Decision theory, forecast evaluation, scoring function, survival forecast, time-to-event forecast, forecast recalibration*

