

Summary factsheet: Forecast skill and verification



Joshua Talib, Research Scientist, ECMWF



What is forecast skill and verification?

Forecast skill is a quantifiable metric used to assess whether a forecast adds value compared to a reference prediction. A sub-seasonal forecast is considered “skilful” if it consistently performs better than climatology (average conditions) or persistence (assuming the same conditions as the start of the forecast). Even when forecasts aren’t perfect, they can still be useful for planning and risk reduction, especially compared to the decisions taken when acting without a forecast.

Forecast verification is the process of systematically comparing forecasts with observed outcomes to assess their accuracy, reliability, and usefulness. It’s essential for understanding how forecasts behave under different conditions, for example, whether they tend to overpredict rainfall, or perform better in some regions or seasons than others. Verification builds forecast confidence and helps users know what level of certainty they can place on a forecast.



Different forecast types need different verification methods

Not all forecasts are the same, and neither are their verification approaches. Choosing the right method ensures the forecast verification reflects what the forecast is trying to communicate.

Deterministic forecasts (i.e. predicting tropical storm strength) are typically evaluated using measures of error or correlation.

Categorical forecasts (i.e. predicting tropical storm strike) are verified using hit rates, false alarms or skill scores.

Probabilistic forecasts (i.e. the likelihood of tropical storm strike) use statistical scores such as the Brier Score and ROC curves.

Ongoing learning process

Verification is part of a continuous cycle of learning and improvement. Regular evaluation helps identify where models perform well and where they need refinement. It also helps forecasters’ communication.

Verification supports scientific progress and better decision-making by creating an ongoing feedback loop between forecasts, verification, and improvement.



Choosing the correct baseline

Probabilistic forecast skill is often measured relative to a reference forecast. Using an inappropriate reference forecast can give a misleading picture of forecast performance. For example, a forecast might look skilful compared to persistence, but not much better than climatology. Selecting the right reference forecast, ideally one that represents typical user behaviour, ensures a fair and meaningful assessment. This is crucial for communicating realistic expectations of forecast quality.

Verification results in context

Verification statistics alone don’t tell the whole story. A “good” score in one situation might be “poor” in another, depending on the variable, region, or decision being made. What matters most is fitness for purpose, whether a forecast performs well enough to support the decision it’s designed to inform. Communicating verification results in approachable language, and linking them to real-world examples, helps users understand what verification scores mean and how to use them effectively.

