Testing forecasts specified as synthetic catalogs

David Rhoades

SCEC Workshop on Predictive Skill Across Tectonic Settings and Planning CSEP 2.0

8 September 2018, Palm Springs, California

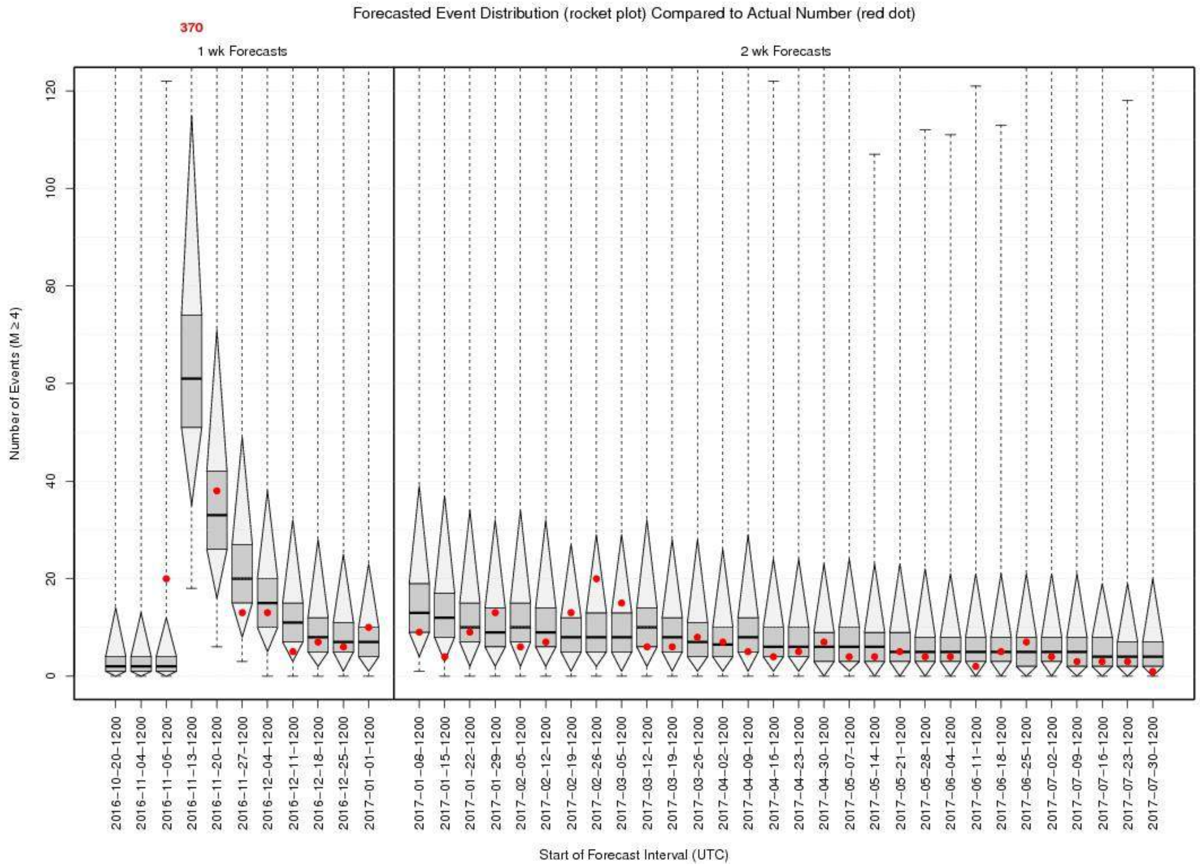
GNS SCIENCE
TE PŪ AO

Summary

- Testing of forecasts specified as synthetic catalogs requires different methods from CSEP1.0, but is straightforward in principle.
- Advantages: Greater flexibility, including tests incorporating space-time correlations.
- Challenges: May still need gridding for some tests, e.g. information gain statistics, and issues around updating and storage of synthetic catalogs.

Consistency tests of simulation-based models

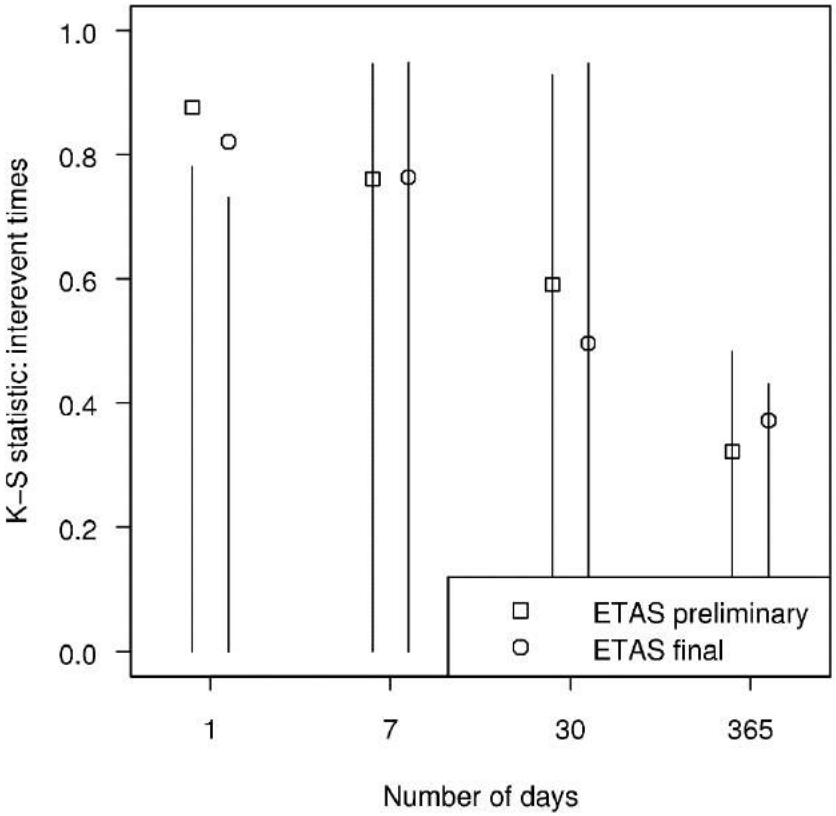
- A wide range of consistency tests is possible.
- Tests can include clustering behaviour
- General approach is to compare a statistic computed from the real catalogue with the distribution of the same statistic computed from numerous simulations of the model.
- Each time-period results in a p-value. Combined distribution of p-values can be compared to a uniform distribution.



Example: (David Harte, GNS Science)
Number of events N(M ≥ 4) observed in each time period compared to distribution of N(M ≥ 4) in ETAS model synthetic catalogs for each time-period.
ftp://ftp.gns.cri.nz/pub/davidh/NZ-OEF/

Consistency tests of simulation-based models

- **For a distribution**:
  Comparing the observed distribution in the real catalog to the combined distribution from the synthetic catalogs gives a K-S statistic.
- Comparing the distribution in each synthetic catalog to the same combined distribution gives a family of K-S statistics.
- Comparing the K-S statistic of the real catalog to the family of K-S statistics gives a p-value.
- Each time-period thus results in a p-value. Combined distribution of p-values can be compared to a uniform distribution.



Example: Tests of inter-event time distribution based on a family of Kolmogorov-Smirnov statistics for forecasts of period 1, 7, 30 and 365 days starting 1-hour before the Christchurch M6.3 earthquake using preliminary and final input catalogs. (Christophersen et al., SCEC poster, 2016)

Consistency tests continued

- **For multiple time-periods compare distribution of p-values to a uniform (0,1) distribution.**
- **The shape of the cumulative distribution of p-values shows how the model deviates from the data.**

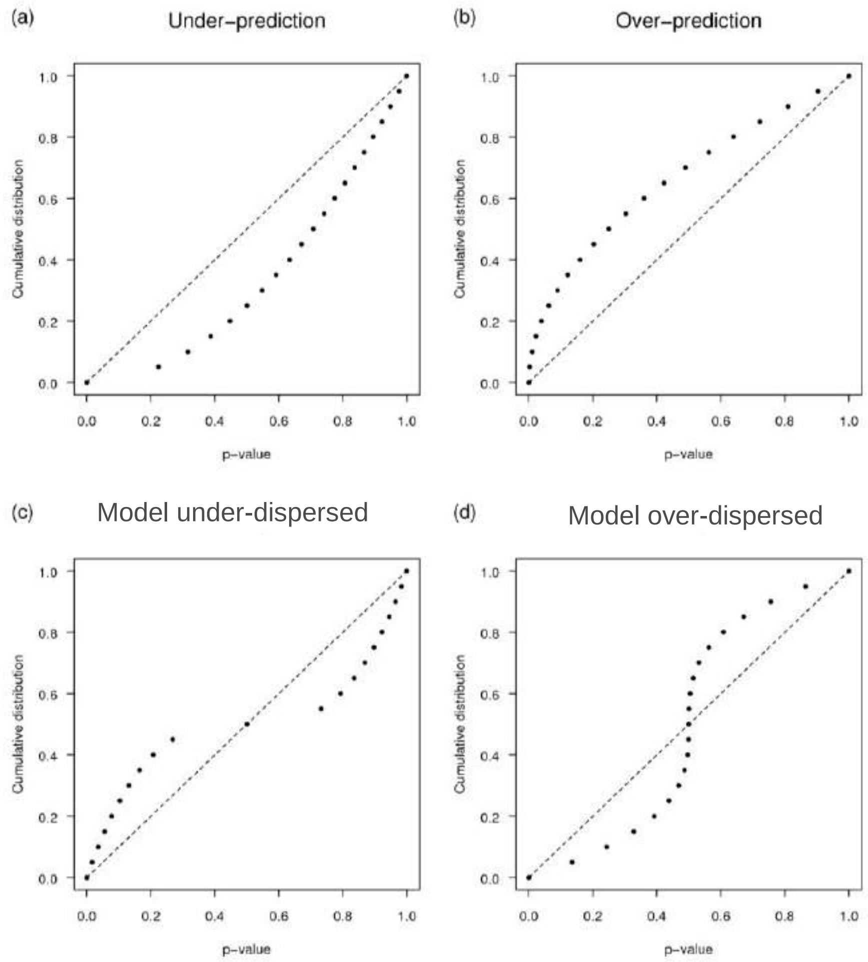


Figure. Schematic of cumulative distribution of p-values for a single statistic over successive test periods (points) and uniform distribution (dashed line): (a) Under-prediction and (b) over-prediction of statistic by the model; (c) Under-dispersion and (d) over-dispersion of statistic in the model simulations

Simulation statistics to do similar job to existing CSEP1.0 consistency tests

- Total number of earthquakes in real and synthetic catalogs (N-test), or counts within particular space-time bins.
- Magnitude distribution of real and synthetic earthquakes (M-test)
- Location distribution of aftershocks projected onto 1D in real and synthetic catalogs (S-test)

Other statistics can explore correlations, e.g. inter-event time distribution (IETD), inter-event distance distribution (IEDD), total earthquake rate distribution over cells (TERD).

Information gain

- For simulation-based forecasts, a transparent estimate of information gain seems difficult without resorting gridded estimates of expected earthquake numbers.
- How can bin expectations best be computed? Rely on simulations, or exploit known regularities (e.g. magnitude distribution, background model)?
- If bin expectations depend on counting simulated events, need to beware of possible zeroes in bin expectations (e.g. underpin with non-zero background).
- How much should computation of bin expectations be modeler’s responsibility?

Synthetic catalog updating issues

- Length of simulated catalogs do not need to match updating interval.
- Simulate for the longest period of interest (e.g. 10 years).
- Update frequently (e.g. daily, weekly monthly, quarterly?) depending on storage capacity.
- Updating frequency controls storage requirements.
- Models can be tested for various catalog lengths and lags (time between end of input catalog and beginning of forecast interval).

Selective synthetic earthquake updating

- Is it feasible to update only those parts of the synthetic catalogs that need updating?
  - delete offspring of earthquakes that failed to occur in the interim;
  - add offspring of earthquakes that occurred in the interim.
- From a computing and storage point of view, is it better to re-compute the synthetic catalogs from scratch, or to selectively update?

Key decisions

- If testing is separated from forecast generation, the most important thing to get right initially is generation and storage of forecasts.
- What updating periods and magnitudes are most important for CSEP testing of models?
- How much effort should CSEP put into testing models of small-earthquake occurrence in the short-term, which potentially consumes most available resources?
- Scientific vs operational forecasting requirements.