User-Centered Evaluation of the Performance of U.S. Geological Survey Aftershock Forecasts

Max Schneider[1], Michael Barall[1], Jeanne Hardebeck[1], Andrew Michael[1], Morgan Page[2], Nicholas van der Elst[2]

1: United States Geological Survey, Moffett Field, CA 2: United States Geological Survey, Pasadena, CA

USGS Aftershock Forecasts

The United States Geological Survey (USGS) releases public aftershock forecasts following earthquakes larger than M4.0 in the United States (Fig. 1). These forecasts have helped inform response and recovery decisions after damaging earthquakes (Schneider et al., 2025), including:

- emergency managers planning search and rescue operations
- engineers deciding whether and when to repair or rebuild damaged buildings
- risk communicators disseminating information on future hazards.

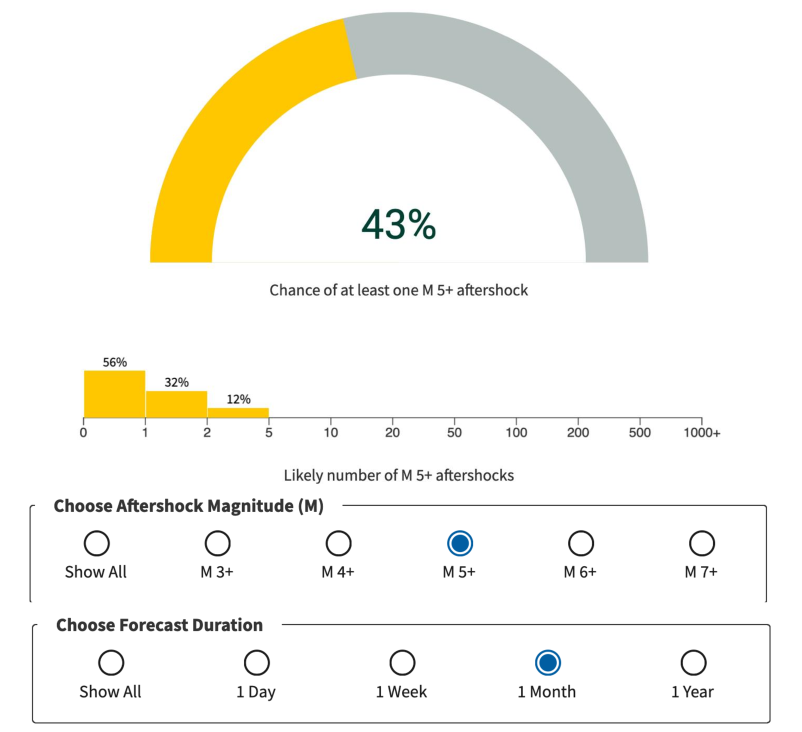

Fig. 1: Example of USGS public aftershock forecast product, showing forecast of M5+ aftershock for the next month following an earthquake

To forecast aftershocks, we fit parametric statistical models to data from past earthquakes (including the mainshock). Simulating from the model produces a probability distribution of aftershock counts (0, 1, …) for a forecast for a specific:

- Magnitude threshold (e.g, M3+, 4+)
- Forecast duration (e.g., one day, month)
- Time point following a mainshock (e.g., 20 minutes, 1 day, 3 months)

We discretize forecast probability distributions in breaks of 1.25% (Fig. 2).

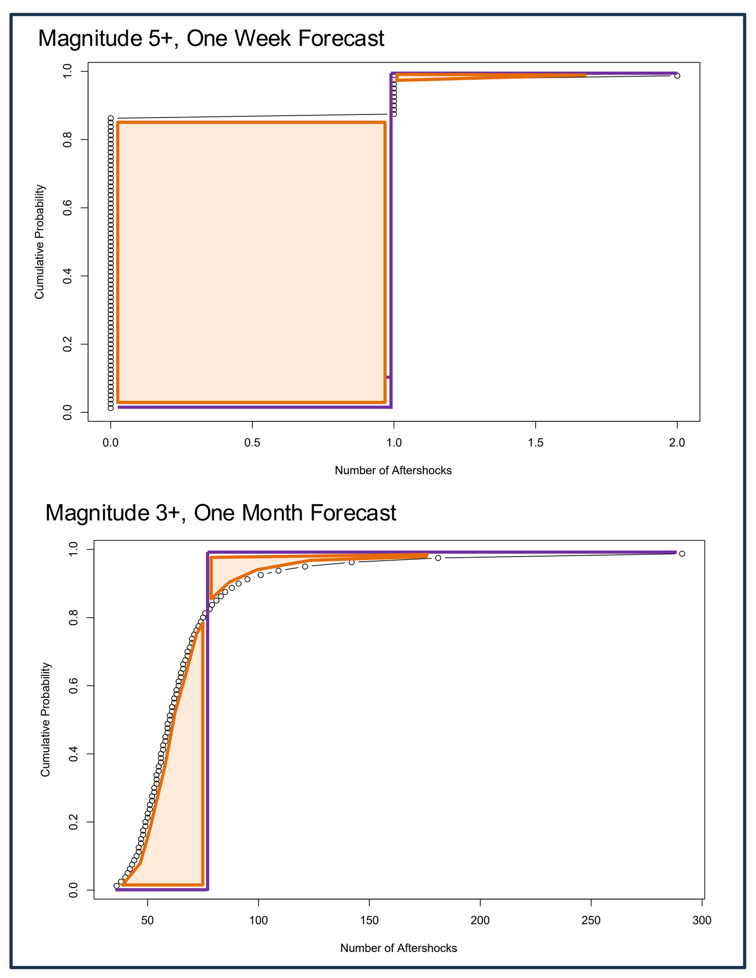

Fig. 2: Example cumulative distribution functions for forecasts for two bins, with the observed number in purple and the CRPS in orange

Forecasted probability distributions are released on the USGS website, with interactive graphics for exploring different magnitude-duration combinations. Forecasts are updated for one year on a fixed schedule, with more frequent updating early in the sequence when more aftershocks tend to occur.

The forecasting system allows for several operational choices (Page et al., 2016):

- Model parameters can vary over time (they are repeatedly estimated during the sequence) rather than stay fixed at specific values. Early in time, model parameters are drawn from distributions corresponding to the tectonic region of the mainshock and these distributions are updated based on the sequence's activity, in a Bayesian way
- Magnitude of completeness (Mc) can be set to vary over time, by adopting parameter values based on the mainshock region
- Mainshock productivity can be parameterized separately from the productivity of the rest of the aftershocks in the sequence

USGS aftershock forecasts have only been evaluated for key sequences (e.g., Michael et al., 2020). Here, we systematically quantify how the above forecasting ingredients affect forecast skill, aiming to improve forecasting for audiences with diverse needs.

Evaluating Forecasts for Synthetic Sequences

We simulate aftershock sequences following mainshocks of magnitudes 6, 7 and 8, using a model with two productivity parameters and no short-term aftershock incompleteness, taking canonical values for model parameters. We then compute aftershock forecasts under each combination of choices within the three operational variables in the table below. We repeat this 10 times for each mainshock and evaluate forecasts using multiple metrics. We aggregate metrics across forecast bins and forecasted sequences, considering the 75th percentile in each bin, due to heavy zero-inflation in all scores.

Variable	Operational Choice
Mag. of completeness (Mc) parameters	Baseline: None, Mc(t)=Mc CA values: F=1.00; G=4.75; H=0.75 World values: F=0.50; G=0.25; H=1.00
Productivity parameters	Baseline: One productivity for mainshock and aftershocks Innovation: Separate productivity for mainshock
Sequence-specific model parameters	Baseline: No updating (parameter estimates fixed to generic tectonic dist'n) Innovation: Generic parameters updated by sequence-specific parameter estimates using Bayesian approach

We evaluate the operational USGS aftershock forecast system, finding that both forecast accuracy and precision improve if forecasts use model parameters fit to the ongoing aftershocks.

User-Driven Evaluation Methods

Past work on earthquake forecast testing has used tests established by an academic consortium, the Collaboratory for the Study of Earthquake Predictability (CSEP; Schorlemmer et al., 2018). Our study differs from traditional CSEP studies because we evaluate:

- full forecast probability distributions, not forecast means
- operational models (which allow the analyst to make modelling choices), not fixed CSEP-style models
- mainshock-triggered (aftershock) not continuous (earthquake) forecasts
- user-relevant components of forecast skill, not CSEP-style test results

We use evaluation metrics that are relevant to user needs for aftershock forecasts. To evaluate the *accuracy* of forecasts represented by cumulative probability distributions of a forecasted variable against observations, we use the continuous ranked probability score (CRPS), or the integrated difference between the forecast distribution and the observation (see orange areas in Fig. 2). The lower the CRPS, the more accurate the forecast and the CRPS is minimized when the forecasted distribution matches the observed distribution. The CRPS has been used widely to summarize forecast accuracy for probabilistic forecasts for other hazards, e.g., weather (Zamo and Naveau, 2018).

The CRPS does not specifically quantify the spread of the forecast distribution (*forecast precision*), nor when a forecast severely over- or under-predicts the observed count. We thus complement it with the following metrics:

- Range: the range of the middle 95% forecast interval, which is provided in USGS' public forecast product
- Underprediction: whether the forecast's maximum is below the observed number and by how much
- Overprediction: whether the forecast's minimum is above the observed number and by how much

Tradeoffs often exist between forecast accuracy vs. precision and forecast overprediction vs. underprediction, which relate to different use cases. While many users would prefer accurate forecast distributions (even at a cost to their precision), a risk communicator may prefer more precise forecasts. Similarly, search and rescue operations may prioritize forecasts that do not underpredict, while engineers delaying building inspections to wait out aftershocks may prefer forecasts that do not overpredict. Quantifying these tradeoffs can inform more user-driven forecasting strategies.

References

Helmstetter, Agnes, et al. "Comparison of short-term and time-independent earthquake forecast models for southern California." Bulletin of the Seismological Society of America 96.1 (2006): 90-106.

Michael, Andrew J., et al. "Statistical seismology and communication of the USGS operational aftershock forecasts for the 30 November 2018 Mw 7.1 Anchorage, Alaska, earthquake." Seismological Research Letters 91.1 (2020): 153-173.

Ogata, Yosihiko. "Statistical models for earthquake occurrences and residual analysis for point processes." Journal of the American Statistical association 83.401 (1988): 9-27.

Page, Morgan T., et al. "Three ingredients for improved global aftershock forecasts: Tectonic region, time-dependent catalog incompleteness, and intersequence variability." Bulletin of the Seismological Society of America 106.5 (2016): 2290-2301.

Schneider, Max, et al. "Meet the People Where They Are: Assessing User Needs for Aftershock Forecast Products in El Salvador, Mexico and the United States." International Journal of Disaster Risk Reduction (2025).

Schorlemmer, Danijel, et al. "The Collaboratory for the Study of Earthquake Predictability: Achievements and Priorities." Seismological Research Letters, 89(4) (2018): 1305–1313,.

Zamo, Michaël, and Philippe Naveau. "Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts." Mathematical Geosciences 50.2 (2018): 209-234.

Preliminary Results and Discussions

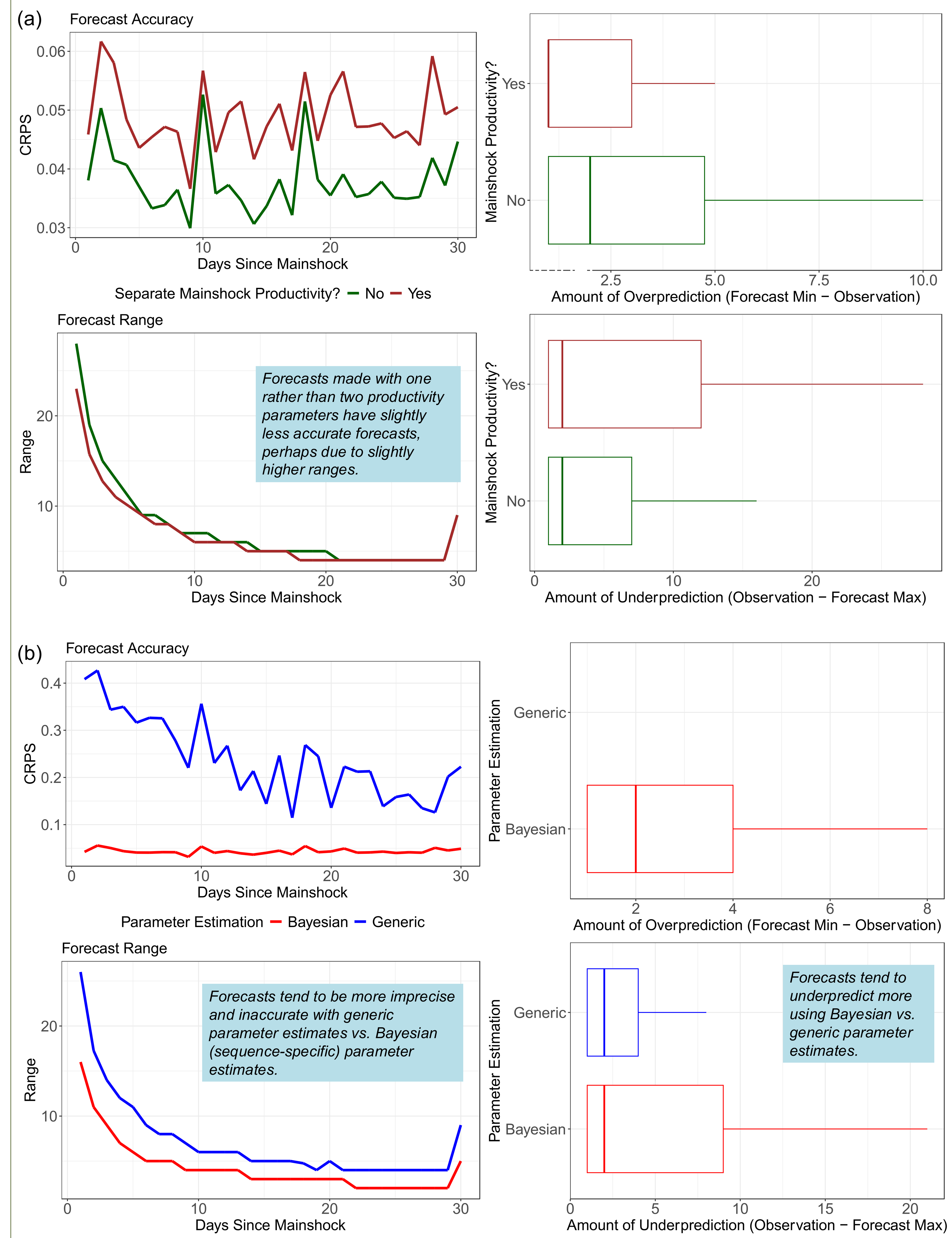

Fig. 3: (a) Testing metrics for forecasts made with mainshock productivity fixed to aftershock productivity or estimated separately.
(b) Testing metrics for forecasts made with model parameters drawn either from generic parameter estimates or Bayesian combination of generic and sequence-specific parameter estimates.

Future Work

For the simulation, we plan to:

- Simulate sequences with different values for model parameters, operational choices
- Calculate forecasts with all combinations of operational choices (supercomputing necessary)
- Sanity-check forecast metrics, for forecasts of all durations and on the time schedule of the operational USGS forecast system

For the observational study, we plan to:

- Compute hypothetical forecasts under each operational combination for several important recent US sequences (e.g., 2019 Ridgecrest, 2020 Magna, UT)
  - Compare to the real USGS published forecast for each sequence
- Evaluate all forecasts against observations with multi-metric approach
- Analyze evaluation results to understand which operational choices consistently lead to superior forecast skill (accuracy, precision, over/under-prediction)
- Present evaluation of USGS forecasts to diverse user groups, emphasizing connection between evaluation metrics and that group's use cases for aftershock forecasts

Contact

Please get in touch with any questions, comments or suggestions!
Email mschneider@usgs.gov.