Developing reference models for earthquake predictability experiments: First Results

Thomas Göbel
University of Southern California, Los Angeles, CA 90089-0742, USA

06/10/2010

Principal Investigators:

Matt Gerstenberger
GNS Science, PO-Box 30-368, Lower Hutt, New Zealand

Danijel Schorlemmer
Southern California Earthquake Center
Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089-0742, USA

J. Douglas Zechar
Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964

Motivation and Goals: Reliable long and intermediate term earthquake forecasts are fundamental in reducing seismic risk, raising public awareness, and encouraging the development of emergency responses in case of major earthquakes. Probabilistic earthquake forecasts exhibit the potential of advancing the physical understanding of the faulting process if predictions based on specic models can be formulated and tested with statistical rigor. Our aim is to provide well documented, easily adaptable models that are based on a minimum of physical assumptions and have been tested extensively hence can serve as possible null hypothesis in proceeding investigations. They will be an integral part of the Collaboratory of the Study of Earthquake Predictability (CSEP) earthquake-testing-distributions, utilizable for standardized testing regions around the world.

	Model class	Parameters	Number of Parameters	Variable Learning Period
1	UNIFORM		0	-
2	UNIFORMACTIVEZONE	$M_{\min}$	1	x
3	WEIGTHEDACTIVEZONE	$M_{\min}$	1	x
4	SMOOTHEDCONSTANTσ	$M_{\min}$, σ	2	x
5	SMOOTHEDVARIABLEσ	$M_{\min}$, σ, d	3	x

Table 1: Overview of model groups with corresponding parameters and indication if the learning period of the input earthquake catalog was varied or the longest available record was chosen. Model groups are listed with increasing complexity from top to bottom. The first two (UNIFORM, UNIFORMACTIVEZONE) were used for reference purposes and requiered no or only a minum amount of parameter optimization. The last three model groups were tested and optimized extensively and later compared to the current best RELM model. d and σ control the local and overall smootheness of a model and $M_{\min}$ is the minimum magnitude cutoff of the input catalog.

Method: Currently five different models have been developed and their performances analyzed through tests deployed within the scope of CSEP. Models were based on the assumption that earthquakes are more likely to occur in areas with previously observed seismic activity. Complexity and number of parameters were increased stepwise (Table 1) starting from a (1) uniform rate model, (2) a uniform rate model where only seismic active regions were considered, (3) a weighted model according to the level of local activity, (4) a constant bandwidth smoothed model and (5) an adaptive bandwidth smoothed model where the bandwidth depends on local event density. (see *Helmstetter et al.* [2006] for more details an adaptive bandwidth smoothing). Model parameter were chosen to maximize the likelihood of the observed rate given a particular forecast rate assuming a Poissonian distribution. Simple model ratings and hypothesis testing were extended by introducing flexible learning periods, magnitude specic event forecasts, a minimum magnitude cutoff for seismic events that where utilized to compute forecasts and through evaluating spatial performances in different tectonic regimes of California. Based on the results of these tests we created a pseudo-prospective forecast for the period of the Regional Earthquake Likelihood Model (RELM) experiment, enabling a comparison between our and officially submitted models. We tested forecasts with the aim to identify a best performing model but also to clarify in what regions, due to what parameters or modeling steps a forecast can be preferred over others. Some vital insights into the physics of faulting, the characteristics of earthquake catalogs and seismicity can be obtained through evaluating the role and performance of each parameter and model class individually.

Results: Subsequent increase of model complexity resulted in a steady improvement of forecast test scores with smoothed models clearly outperforming other model groups (for example the uniformActiveZone or weightedActiveZone model groups). The evaluation of parameters showed a strong influence of forecasts on learning periods and minimum magnitude cutoffs $M_{\min}$ with best results if more recent seismic records were combined with fairly small $M_{\min}$. Shorter learning periods better represented currently active seismic regions and received highest test scores in areas of aftershock occurrences as observed for example in the Eastern Californian Shear Zone. The optimized $M_{\min}$ were always larger or equal to 2.0 and increased with longer learning periods. On the one hand this emphasizes that locations of very small earthquakes below magnitude 2.0 are diffuse (which might be due to larger uncertainties and events occuring preferedly away from major faults) thus representing longterm seismicity inadequately. On the other hand it is interesting to realize that deficiencies due to short seismic records could be compensated by extending the considered magnitude range to smaller events (small but above magnitude 2.0) indicating that small earthquakes play an important role in forecasting large events. To evaluate if this tendency is only observable for catalogs that include aftershocks we removed all dependent events from the seismic record employing an algorithm based on *Gardner and Knopoff* [1974]. As a results the dependency of the forecast performance on the learning period was largely removed with slightly better test scores for longer periods. Nevertheless the previously mentioned tendency of lower $M_{\min}$ with decreasing learning periods was still observed but for the declustered catalog the smallest minimum magnitude was 3.1 for the best model.

Furthermore we compared our best model, which was a constant bandwidth smoothed model with a learning period from 1997 to 2005 and $M_{\min} = 2.0$, to the current best model of the RELM experiment [*Helmstetter et al.*, 2007]. Our best model received slightly better absolute test scores while the Helmstetter model exhibited a better overall consistency with the observed events. This could be due to locally stronger varying test scores for our model with poor results in some areas that were compensated by very good results in other regions. We noticed that differences between the forecasts could be tracked down to three main areas: the eastern Californian Shear Zone, the Mendocino Fracture Zone and the Imperial Fault Region. Both forecasts received unsatisfying scores in the Imperial Fault Region due to three events that occurred in one magnitude-space bin but Helmstetter predicts higher rates resulting in better local scores. The aftershock in the Eastern Californian Shear Zone was better forecast by our model while the HELMSTETTER model had slightly better test scores close to the Mendocino Fracture Zone. Thus the current results could not provide means of favoring one or the other model.

We state that smoothed model perform signicantly better then UNIFORMACTIVEZONE or WEIGHTEDACTIVEZONE model. The minimum magnitude of the input increases for almost all model groups with larger learning catalog time spans while the smoothing bandwidth remains fairly stable. Our best model is comparable in its performance to the current best model of the RELM experiment. The removal of dependent events strongly de-emphasizes the time dependence of forecasts while the previously mentioned trend in the minimum magnitude is still observable.

References

Gardner, J. K., and L. Knopoff (1974), Is the sequence of earthquakes in southern california, with aftershocks removed, poissonian?, *Bull. Seismol. Soc. Am.*, *64*(5), 1363–1367.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006), Comparison of short-term and long-term earthquake forecast models for southern California, *Bull. Seismol. Soc. Am.*, *96*, 90–106.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2007), High-resolution time-independent grid-based forecast for $M \geq 5$ earthquakes in California, *Seismol. Res. Letts.*, *78*(1), 78–86, doi:10.1785/gssrl.78.1.78.