Investigating the effects of smoothing on the performance of earthquake hazard maps

Edward M. Brooks[1,2], Seth Stein[1,2], Bruce D. Spencer[3,2]

[1]Department of Earth and Planetary Sciences, Northwestern University, Evanston, Illinois

[2]Institute for Policy Research, Northwestern University, Evanston, Illinois

[3]Department of Statistics, Northwestern University, Evanston, Illinois

ABSTRACT

To explore how the level of detail in earthquake hazard maps affects their performance, we assess how smoothing Japan's national earthquake hazard maps affects their fit to a 510-year record of shaking. As measured by the fractional exceedance metric implicit in such probabilistic hazard maps -- that the predicted ground motion should be exceeded only at a specific fraction of the sites -- smoothing over progressively larger areas improves the maps' performance such that in the limit a uniform map performs best. However, using the squared misfit between maximum observed shaking and that predicted as a metric, map performance improves up to a ~75-150 km smoothing window, and then decreases with further smoothing, such that a uniform map performs worse than the unsmoothed map. Because the maps were made by using other data and models to try to predict future earthquake shaking, rather than by fitting past shaking data, this result is probably not an artifact of hindcasting rather than forecasting. It suggests that hazard models and the resulting maps can be over-parameterized, in that including too high a level of detail to describe past earthquakes may lower the maps' ability to predict future shaking. We suspect that this effect will contribute to imposing a limit to how good a hazard map can be.

INTRODUCTION

Recent earthquakes that did great damage in areas shown by earthquake hazard maps as relatively safe have generated interest in assessing how well these maps forecast future shaking (Kerr, 2011; Reyners, 2011; Stein et. al, 2011, 2012; Peresan and Panza, 2012; Stirling, 2012; Gulkan, 2013; Marzocchi and Jordan, 2014; Wang, 2015). In particular, Geller (2011) argues that "all of Japan is at risk from earthquakes, and the present state of seismological science does not allow us to reliably differentiate the risk level in particular geographic areas," in which case maps less detailed than present ones would be preferable.

In an earlier paper (Brooks et al., 2015), we examined how well a 510-year-long record of earthquake shaking in Japan is described by the current Japanese national hazard (JNH) maps compared to uniform and randomized versions of these maps. The metric implicit in the probabilistic JNH maps is that during the chosen time interval the predicted ground motion should be exceeded only at a specific fraction of the sites. We found that, as measured by this metric, both uniform and randomized maps do better than the actual maps. However, using as an alternative metric the squared misfit between maximum observed shaking and that predicted, the JNH maps do better than uniform or randomized maps. It appears that although the JNH maps seek to predict shaking levels that should be exceeded at a certain fraction of the sites, the process by which their parameters are chosen made the predicted shaking more closely resemble the maximum observed.

A uniform map is one smoothed (averaged) over the entire country, with all spatial details are removed. Hence these results lead to the question of what the effect of smoothing over a smaller area may be. Is there some level of smoothing that preserves an intermediate level of detail that better describes the shaking?

SMOOTHED MAP PERFORMANCE

A hazard map's performance should ideally be assessed by comparing it to the shaking in earthquakes that occurred after the map was made. However, because the maps forecast the shaking expected over periods of hundreds or thousands of years, the short time period since they began to be made makes assessing how well they work difficult (Beauval *et al.,* 2008; 2010). Hence, we follow earlier studies and examine how well maps describe past shaking (Stirling and Peterson, 2006; Albarello and D'Amico, 2008; Stirling and Gerstenberger, 2010; Kossobokov and Nekrasova, 2012; Nekrasova *et al.,* 2014; Wyss *et al.,* 2012; Mak *et al.,* 2014). Although such assessments are not true tests, in that they compare the maps to data that were available when the map was made, they give useful insight into the maps' performance.

As in our previous paper, we compared a catalog (Miyazawa and Mori, 2009), giving the largest known shaking on the Japan Meteorological Agency (JMA) instrumental intensity scale at each grid point in 510 years (1498-2007) to four Japanese National Hazard (JNH) maps for different return periods (Japanese Seismic Hazard Information Station, 2015) (Fig. 1). The effect of site conditions is included in both the predictions and observations, making the two comparable.

The maps show for each site i the level of shaking x_i (which we refer to as the predicted shaking) that has probability p of being exceeded during the return period. We assessed their performance first by the fractional exceedance metric

$$M0(f,p) = |f - p|,$$

the difference between p , the expected number of sites at which the observed shaking should exceed that predicted by the map, and f, the actual number of such sites. Because this metric does not consider the magnitude of the predicted and actual shaking, we also assessed the maps with a squared misfit metric

$$M1(s,x) = \sum_{i=1}^{N} \frac{(x_i - s_i)}{N}$$

which compares x_i and s_i , the predicted and maximum observed shaking at each of the N sites. The two metrics characterize different aspects of map performance, in that *M0* is more sensitive to how well the map predicts average shaking levels, whereas *M1* is more sensitive to how well it predicts the spatial variations in shaking (Stein et al., 2015a). Visually comparing maps of predicted and observed shaking amounts to using *M1*.

The JNH maps were smoothed by placing a square composed of cells over each point on the map, averaging the predictions within the square, and assigning that value to the central cell. Iterating over all points on the map using progressively larger squares yielded maps smoothed to greater degrees. For regions close to the coast we used only values on land Japan, disregarding values from the surrounding ocean. This procedure preserves the number of points in each map, so successive iterations can be compared to the observed history of shaking via the two metrics. The smallest smoothing square was 3 x 3, and each individual cell was ~1.5 km on a side.

Smoothing over a small area preserves many details of the hazard maps, suppressing only the sharpest high and low hazard features. Progressively larger smoothing areas suppress more of the details (Fig. 2). Fig. 3 shows plots of the change in map performance as a function of smoothing area, for each of the four maps using both metrics.

The fractional exceedance metric (*M0*) generally improves as the smoothing area increases. Fluctuations are present for smaller smoothing areas, but performance increases steadily for smoothing areas above 200 cells (300 km on a side) across. This reinforces our earlier result, in that smoothing over all of Japan produces uniform maps, which we found perform better than the JNH maps as measured by *M0*.

In contrast, as measured by the squared misfit metric (*M1*), map performance improves somewhat up to a 50-100 cell (75-150 km) smoothing window, and then decreases with further smoothing. This reinforces our earlier result that by this metric uniform maps perform worse than the unsmoothed map.

We repeated these comparisons for an updated map that incorporated shaking from the 2011 Tohoku event. Brooks et al. (2015) noted that adding these high shaking values improved the JNH maps' performance as measured by both metrics, but their performance relative to uniform and randomized maps remained the same. Similarly, we found that the effects of smoothing on performance remained essentially the same.

IMPLICATIONS

These results suggest that including too high a level of detail to describe past earthquakes may lower hazard maps' ability to predict future shaking. Such an effect seems plausible given the variability in space and time of earthquake recurrence, so previous earthquakes do not completely show what will happen in the future. Longer records including paleoseismic data, complemented with inferences from geological and geodetic data about faults, are naturally better. However, even a very long record is unlikely to fully capture the variability.

An analogous phenomenon is recognized in other applications and termed "overfitting" or "overparameterization." Figure 4 shows an example of using a model derived from past data to predict the future evolution of a function. A linear model fits the past data and predicts the future reasonably well, and a quadratic does both even better. However, an 8th order polynomial that fits the past data perfectly does a poor job of predicting the future. The more detailed model seems better because it matches the past so well, but imposing that level of detail makes the model predict the future worse. Hence to forecast the future, the goal should be not to build the most detailed model, but instead one that is robust or stable in the sense that small changes in the past data do not dramatically change the model's forecasts (Parker, 1977; Box, 1979)

Our results showing an improved fit resulting from smoothing do, however, have other possible interpretations. First, the fact that the smoother models fit better could result from some features of the historical shaking dataset used. Second, our approach involves comparing a time-dependent hazard model to past data (hindcasting) rather than the more desirable comparison with future data (forecasting). As discussed in our earlier paper comparing these maps and data, we do not believe either problem is large enough to invalidate our approach. Most crucially, the maps were made by using other data and models to try to predict future earthquake shaking, rather than by fitting past shaking data. In particular, the earthquake magnitudes assumed in the maps were inferred from the fault lengths (Fujiwara *et al.,* 2009), rather than from past intensity data. Because the hazard map parameters were not chosen to specifically match the past intensity data, comparing the map and data is a useful comparison.

These results are for a particular area, much of which has a high earthquake hazard, and a particular set of maps and data. However, these results, combined with the fact that in many applications overfitting past data leads to poorer future predictions, suggests that similar effects could arise for earthquake hazard maps elsewhere. Our approach involved smoothing maps resulting from a hazard model. Alternative approaches could involve using less detailed models to produce maps. In our view, smoothing may be valuable and is certainly worth further exploration.

DATA AND RESOURCES

The Japanese hazard maps are from *http://www.j-shis.bosai.go.jp/map/?lang=en* (last accessed February 2015). The catalog of historic intensity from Miyazawa and Mori (2009) was provided by M. Miyazawa.

ACKNOWLEDGEMENTS

We thank Masatoshi Miyazawa and James Mori for sharing their intensity data. Brooks thanks the Institute for Policy Research at Northwestern University for his graduate fellowship.

REFERENCES

Albarello, D., and V. D’Amico (2008). Testing probabilistic seismic hazard estimates by comparison with observations: an example in Italy. *Geophys. J. Int.* **175**, 1088–1094.

Beauval, C., P.-Y. Bard, S. Hainzl, and P. Guéguen, (2008). Can strong motion observations be used to constrain probabilistic seismic hazard estimates? *Bull. Seismol. Soc. Am.* ***98*,** 509–520.

Beauval, C., P.-Y. Bard, and J. Douglas (2010). Comment on “Test of seismic hazard map from 500 years of recorded intensity data in Japan” by Masatoshi Miyazawa and Jim Mori. *Bull. Seismol. Soc. Am.* ***100*****,** 3329-3331

Box, G. E. P. (1979) Robustness in the strategy of scientific model building. *Robustness in statistics* **1,** 201-236.

Brooks, E. M., S. Stein, and B. D. Spencer (2015). Comparing the performance of Japan’s earthquake hazard maps to uniform and randomized maps. *Seismol. Res. Lett.* in Press.

Fujiwara, H. *et al.* (2009) Technical reports on national seismic hazard maps for Japan, *Technical Note of the National Research Institute for Earth Science and Disaster Prevention* No. 336.

Geller, R. J. (2011). Shake-up time for Japanese seismology. *Nature* **472**, 407–409.

Gulkan, P. A. (2013). A dispassionate view of seismic-hazard assessment. *Seism. Res. Lett*. **84,** 413-416.

Japanese Seismic Hazard Information Station (J-SHIS) (2015). *http://www.j-shis.bosai.go.jp/map/JSHIS2/download.html?lang=en*. (Last accessed May 2015).

Kerr, R. A. (2011). Seismic crystal ball proving mostly cloudy around the world. *Science* **332,** 912-913.

Mak, S., R.A. Clements, and D. Schorlemmer, (2014). The statistical power of testing probabilistic seismic-hazard assessments. *Seismol. Res. Lett.* **85**, 781–783 (2014).

Marzocchi, W. T.H. and Jordan (2014). Testing for ontological errors in probabilistic forecasting models of natural systems. *Proc. Natl. Acad. Sci. U.S.A.* ***111*****(33),** 11973-11978.

Miyazawa, M. and J. Mori (2009). Test of seismic hazard map from 500 years of recorded intensity data in Japan. *Bull. Seismol. Soc. Am.* **99,** 3140-3149.

Nekrasova, A., V. Kossobokov, A. Peresan, A. and A. Magrin (2014). The comparison of the NDSHA, PSHA seismic hazard maps and real seismicity for the Italian territory. *Natural Hazards* **70,** 629–641.

Peresan, A., and G.F. Panza (2012). Improving earthquake hazard assessments in Italy: An alternative to “Texas sharpshooting.” *Eos, Transactions, American Geophysical Union* **93,** 538.

Reyners, M. (2011). Lessons from the destructive Mw 6.3 Christchurch, New Zealand, earthquake. *Seismol. Res. Lett.*, ***82*****(3)**, 371-372.

Stein, S., R.J. Geller, and M. Liu (2011). Bad assumptions or bad luck: why earthquake hazard maps need objective testing. *Seismol. Res. Lett.*, ***82*****(5)**, 623-626.

Stein, S., R.J. Geller, and M. Liu (2012). Why earthquake hazard maps often fail and what to do about it. *Tectonophysics* **562–563,** 1–25.

Stein, S., B.D. Spencer, and E.M. Brooks (2015a). Metrics for assessing earthquake hazard map performance, *Bull. Seismol. Soc. Am. 1*05, 4, doi: 10.1785/0120140164 .

Stein, S., B.D. Spencer, and E.M. Brooks (2015b). Bayes and BOGSAT: issues in when and how to revise earthquake hazard maps. *Seismological Research Letters*, ***86*****(1),** 6-10.

Stirling, M. W. (2012). Earthquake hazard maps and objective testing: the hazard mapper's point of view. *Seismol. Res. Lett.* **83,** 231–232.

Stirling, M.W., and M. Gerstenberger (2010). Ground motion-based testing of seismic hazard models in New Zealand, *Bull. Seismol. Soc. Am.* **100,** 1407-1414 (2010).

Stirling, M.W. and M. Petersen (2006). Comparison of the historical record of earthquake hazard with seismic-hazard models for New Zealand and the continental United States. *Bull. Seismol. Soc. Am.* **96**, 1978-1994.

Wang, Z. (2015). Predicting or forecasting earthquakes and the resulting ground motion hazards: a dilemma for earth scientists, *Seismol. Res. Lett.* ***86,*** *1-5*.

Wyss, M., A. Nekraskova, and V. Kossobokov, V. (2012). Errors in expected human losses due to incorrect seismic hazard estimates. *Natural Hazards* **62,** 927–935.

MAILING ADDRESSES

Edward M. Brooks
Department of Earth and Planetary Sciences
Technological Institute
2145 Sheridan Road
Evanston IL, 60208-3130

Seth Stein
Department of Earth and Planetary Sciences
Technological Institute
2145 Sheridan Road
Evanston IL, 60208-3130

Bruce Spencer
Department of Statistics
2006 Sheridan Rd
Evanston, IL 60208

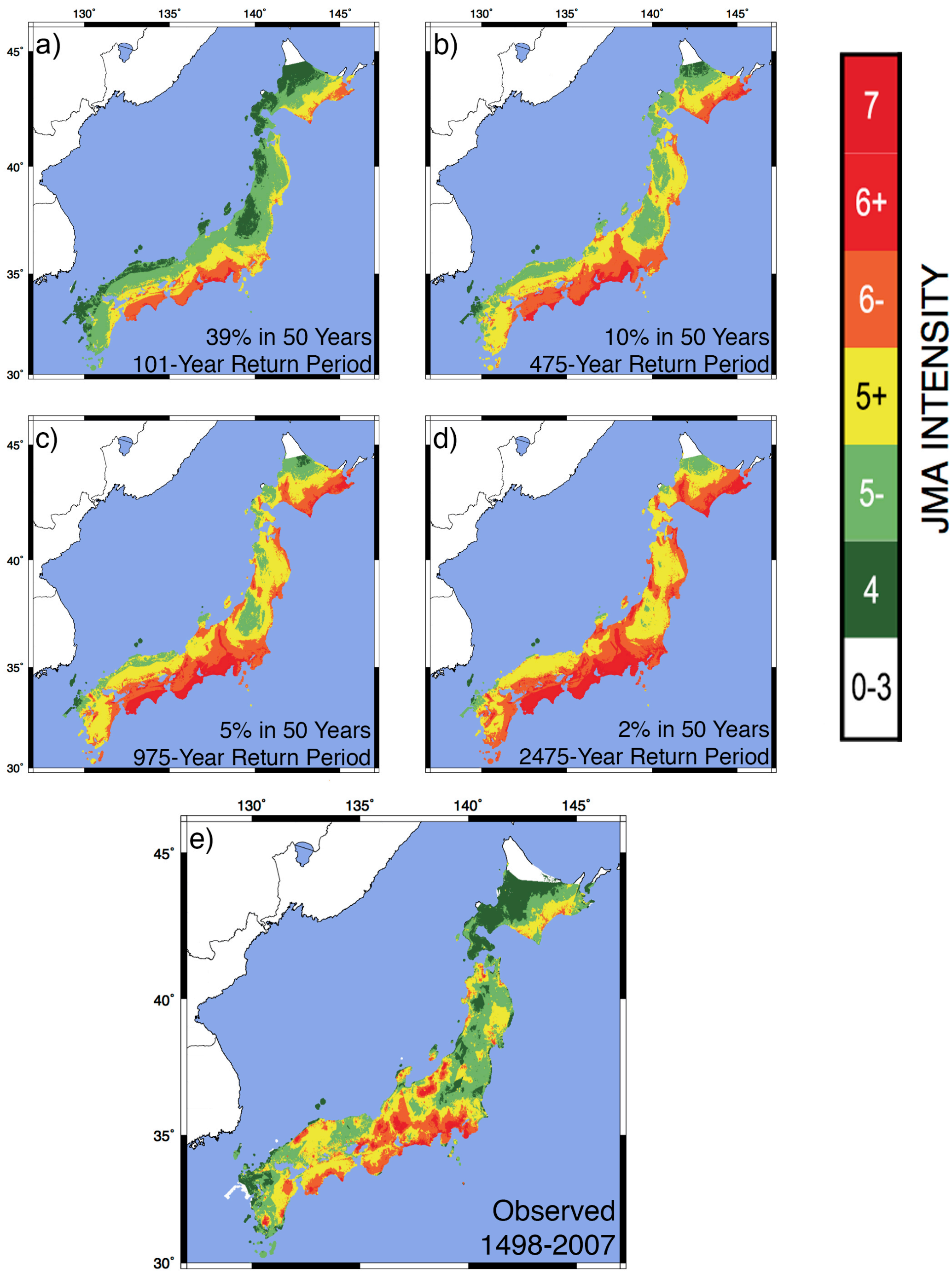

Figure 1. (a-d) Probabilistic seismic hazard maps for Japan, generated for different return periods in 2008. (e) Largest known shaking on the Japan Meteorological Agency intensity scale in 510 years (Miyazawa and Mori, 2009).

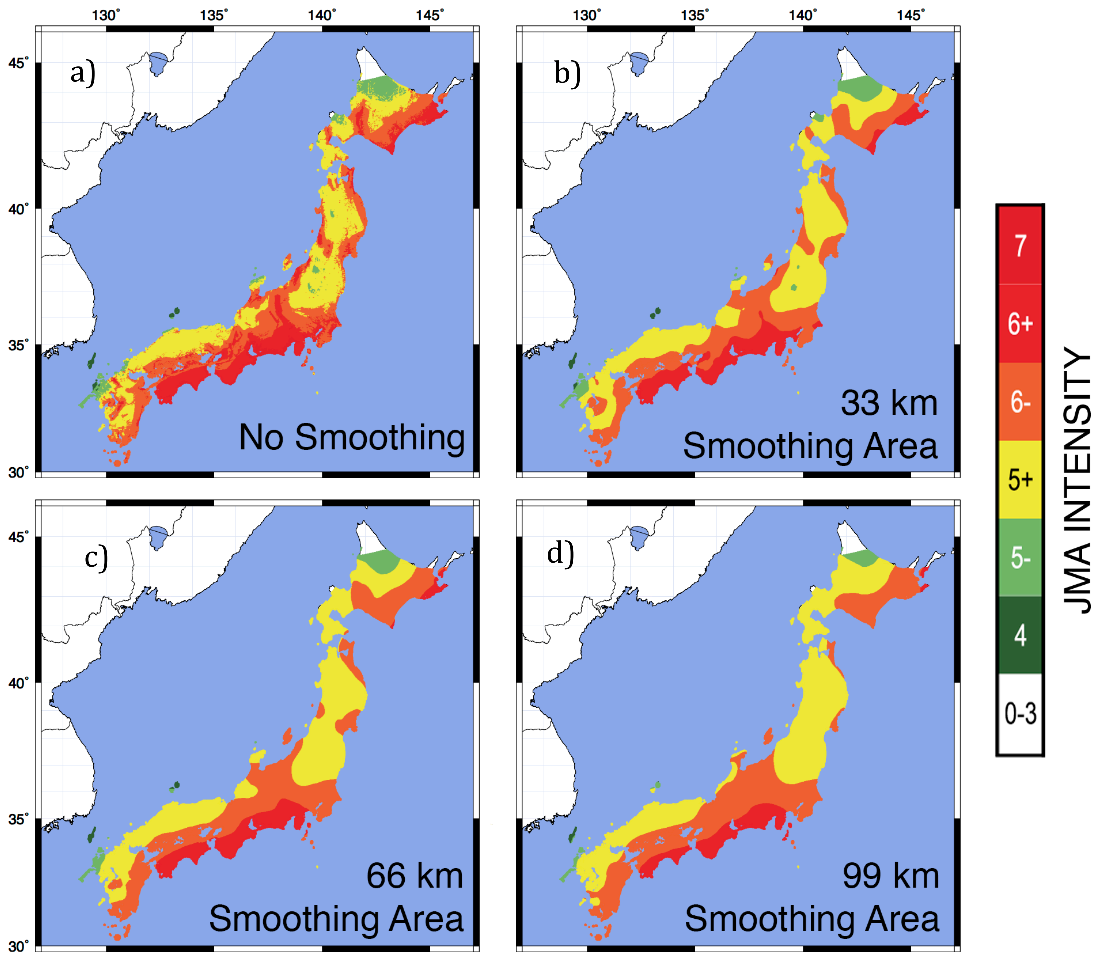

Figure 2. Effects of smoothing the JNH map with 2475-year return period over progressively larger areas (b-d).

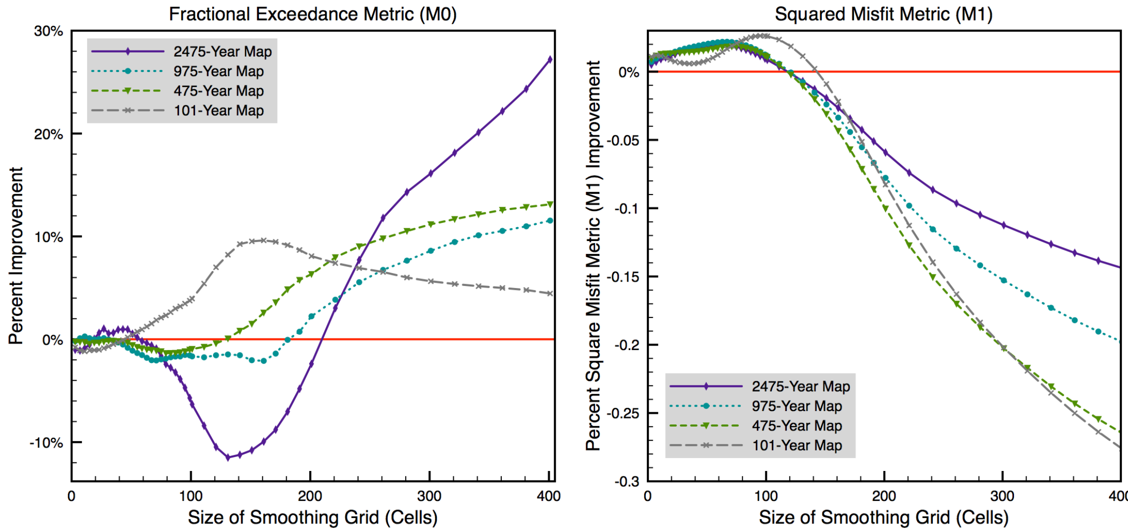

Figure 3. Relative improvement in map performance described by the fractional exceedance (a) and squared misfit (b) metrics compared to the original map, for different amounts of smoothing.

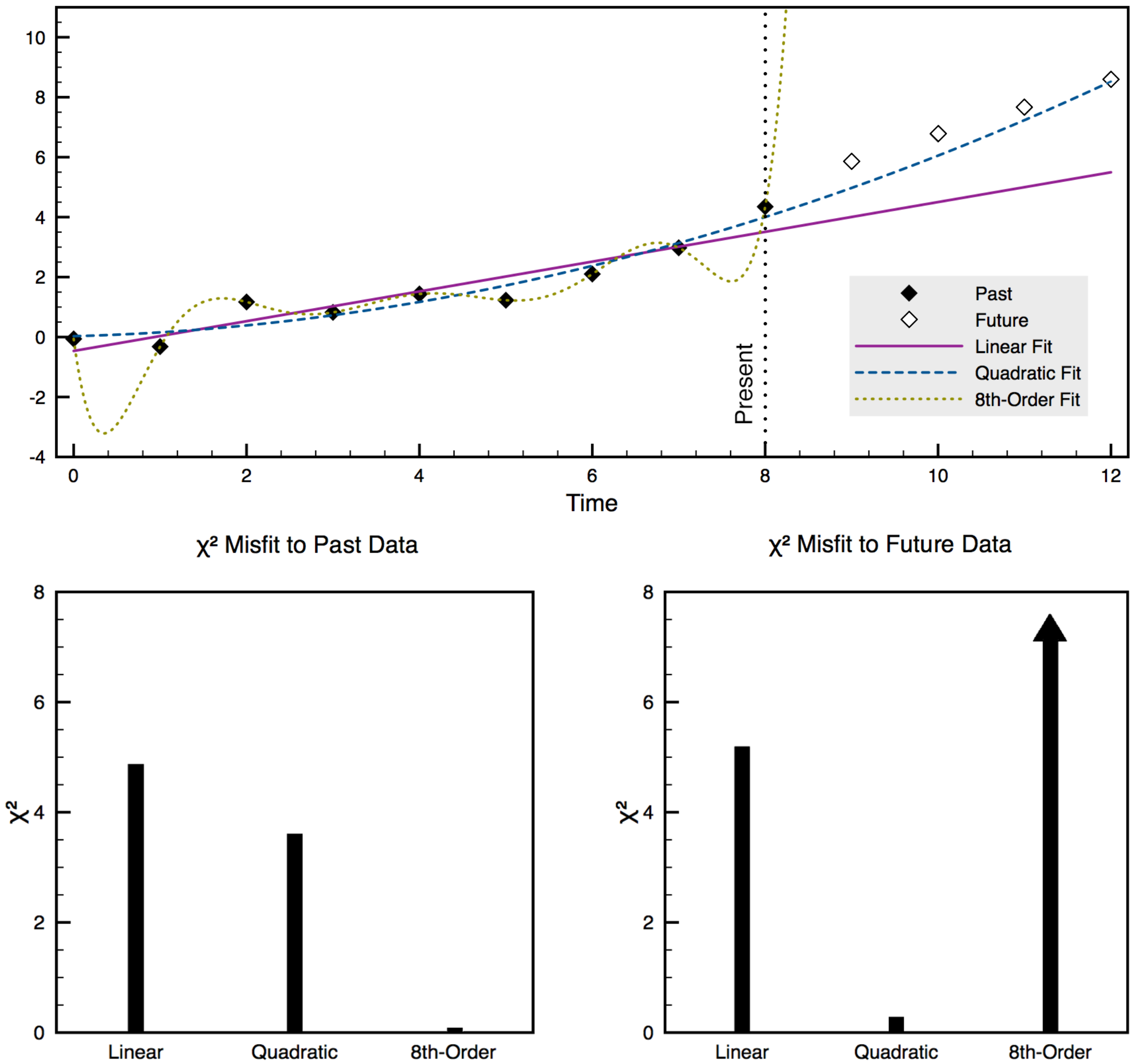

Figure 4. Example of the effect of overparameterization on forecasting. A high order polynomial fits past data better than linear or quadratic models, but this more detailed model predicts the future worse than the simpler models.