# Loma Prieta seismicity: prediction errors differ significantly from independent identically distributed surrogates

Nick Fitton [1] and Frank Horowitz[2]

(1) Australian Geodynamics Cooperative Research Centre, Perth, W.A.
(2) CSIRO Division of Exploration and Mining, PO BOX 437 Nedlands WA, Australia
(e-mail frank@ned.dem.csiro.au; phone: +61 8 9389 8421; fax: +61 8 9389 1906)

A class of time series prediction techniques has previously been applied to "postdicting" (simulating predictions, but after the fact) the seismic activity in the Loma Prieta region of California [Horowitz, EOS, v72, n44, p321, 1991]. Briefly, all hypocenters recorded within a fixed polyhedron around the Loma Prieta mainshock, above a depth of 25 Km., and above (coda duration) magnitude 0.1 are retrieved from the Northern California Seismic Network database. All magnitudes within 2 week intervals are converted to scalar moment, and their sum is converted back to a magnitude. This results in an equal-interval time series representing the physical output of the fault system. This time series is Principal Component filtered to retain the strongest components. The filtered time series is then put into Time-Delay-Coordinates (TDC) by passing a moving window through it: each entry in the window becomes a component of the corresponding point in the N-dimensional TDC space. Treating the position of one of these TDC points as the independent variable $\mathbf{X}_i$, we define the dependent variable $f(\mathbf{X}_i)$ as the value of the filtered time series one position past the end of the moving window (i.e. one step ahead). This yields a set of points $\mathbf{X}_i$ and their corresponding values $f(\mathbf{X}_i)$ in TDC space. Making the hypothesis that these control points are sampling a mapping function in TDC space, the mapping function's value is estimated elsewhere by Radial Basis Function (RBF) interpolation of the control points. (It can be shown that RBF interpolation is identical to the dual of simple Kriging with a suitable choice of variogram; thus we are merely Kriging in TDC space. This raises the possibility of applying more sophisticated geostatistical techniques to time series prediction.) The technique requires adjustable parameters, yet we wish to avoid *a posteriori* fitting. To accomplish this, we segment the PC filtered time series into 3 sequential parts: a "training" set, from which we form the control points in TDC space; a "tweaking" set, over which we allow ourselves the freedom to adjust parameters to find the best fitting predictions; and a "predicting" set, during which we run predictions keeping the parameters frozen. Below, we report prediction errors only from the "predicting" set. To test the significance of these predictions, we construct 1000 surrogate datasets, each as an Independent, Identically Distributed (IID) stochastic process (the distribution used is the one from the equivalent magnitude time series). Using the same techniques described above, with the parameters chosen from the actual "tweaking" dataset, we predict the behaviour of these surrogate datasets. We compare the distribution of errors from the actual predictions to the distribution of errors from the surrogate datasets using a Kolmogorov-Smirnov test. The result suggests that the actual distribution differs from the surrogate distribution with better than 99% confidence. We infer that our actual predictions are exploiting something significantly different from the distribution. An auxiliary inference is that Loma Prieta seismicity is not merely IID stochastic. The text of Fitton's M.S. thesis chapter on this topic can currently be found in <http://www.ned.dem.csiro.au/HorowitzFrank>.