Comment on "Relationship between accelerating seismicity and quiescence, two precursors to large earthquakes" by Arnaud Mignan and Rita Di Giovambattista

Citation for published version:
Greenhough, J, Bell, A & Main, I 2009, 'Comment on "Relationship between accelerating seismicity and quiescence, two precursors to large earthquakes" by Arnaud Mignan and Rita Di Giovambattista' Geophysical Research Letters, vol. 36. DOI: 10.1029/2009GL039846

Digital Object Identifier (DOI):
10.1029/2009GL039846

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Geophysical Research Letters

GEOPHYSICAL RESEARCH LETTERS, VOL. 36, L17303, doi:10.1029/2009GL039846, 2009

Comment on "Relationship between accelerating seismicity and quiescence, two precursors to large earthquakes" by Arnaud Mignan and Rita Di Giovambattista

J. Greenhough,[1] A. F. Bell,[1] and I. G. Main[1]

Received 30 June 2009; revised 30 July 2009; accepted 6 August 2009; published 2 September 2009.

Citation: Greenhough, J., A. F. Bell, and I. G. Main (2009), Comment on "Relationship between accelerating seismicity and quiescence, two precursors to large earthquakes" by Arnaud Mignan and Rita Di Giovambattista, *Geophys. Res. Lett.*, *36*, L17303, doi:10.1029/2009GL039846.

[1] The significance levels of many reported episodes of Accelerating Moment Release (AMR, a cumulative function of earthquake magnitude with time) have been shown to be too low to reject a range of alternative hypotheses [*Hardebeck et al.*, 2008]. While *Mignan* [2008] acknowledges the deficiency of power-law fitting alone for forecasting large events via AMR, this and the proceeding study [*Mignan and Di Giovambattista*, 2008] do not address an underlying problem of applying standard regression methods to cumulative data. We consider this a timely opportunity to emphasize why regression on any cumulative quantity requires the utmost care and is at best avoided. This cautionary comment is relevant to a wide range of applications in geophysics and elsewhere.

[2] Each measurement in a cumulative quantity depends, via summation, on all the preceding measurements in addition to the current increment. Even in the simple case of independent Gaussian increments equally spaced in time, their cumulative sum is the familiar 2-D random walk (in 3-D, Brownian motion) with periodicities on all timescales. Low-frequency components have higher amplitudes (hence the spectral description 'red noise') and any trend one seeks will generally be found somewhere. Figure 1 is an example of such a process, including two arbitrarily chosen random trends whose high R^2 values clearly tell us nothing about the process that generated them. The main problem with the use of cumulative data is that such spurious trends arise everywhere, even without a systematic driver. For example, *Hardebeck et al.* [2008] demonstrate the ease with which apparent AMR can be identified in simulated data containing no real AMR, attributing many such events to careful data selection. Our point here is that this is a general problem caused by the inapplicability of standard regression methods to cumulative data.

[3] While one may in some circumstances be confident that random trends are consistently dominated by a larger systematic trend, a further obstacle is the failure to meet the requirements for conventional regression. Regardless of the details – linear, non-linear, least squares, Poisson – regression will not maximize the likelihood of the data given the fit unless the residuals (differences between data and fit) are independent. Cumulative data, however, are highly correlated, and residuals around a low-frequency fitted function will display correlation due to trends on shorter timescales. Furthermore, if the distribution of residuals is non-Gaussian then the least-squares method is inappropriate even for independent residuals [see, e.g., *Greenhough and Main*, 2008]. Statistics derived from all such regressions are therefore ambiguous. For example, whereas it is common practice to search for AMR by comparing linear and power-law least-squares fits to cumulative functions of magnitude (or of event numbers) against time (the *c* value [see *Mignan et al.*, 2006, and references therein]), finding a lower root-mean-square residual for the power law does not imply that it is the more likely underlying trend. In spite of warnings in standard texts such as *Box and Jenkins* [1970], many other examples of potentially spurious regression have been identified and discussed in the fields of climate reconstruction [*Thejll and Schmith*, 2005], energy consumption [*Lawson*, 2007], global consciousness [*Scargle*, 2002], and economics [*Phillips*, 1986]. Since models of these and other physical systems can carry significant human impact, the fundamental shortcomings outlined here should not be ignored.

[4] It is therefore preferable to examine incremental data, especially where their independence allows rigorous testing for secular changes; a wealth of material on basic data analysis exists both on line and in text books. For the example in Figure 1, we could begin by checking for (i) independence via a scatter plot and (ii) Gaussianity via a quantile-quantile plot, and possibly try more sophisticated diagnostics. If these assumptions are met, a Student's t-test could answer the question, 'Is the mean increment during the trend significantly different from the mean increment preceding it?'. As expected for this artificial example, we cannot reject the hypothesis that the mean increment is constant. If the increments are correlated, however, then this property must be modelled and subsequently removed. When examining earthquake rates, for instance, one must account for the clustering of events in space and time before testing for changes in the underlying 'background' rate [*Marsan and Nalbant*, 2005]. This requires the fitting of a point-process model such as ETAS [*Ogata*, 1999], for which software is readily available; though it is helpful to plot the integrated fit alongside cumulative event numbers, it is the individual event times and magnitudes that are used to optimize correctly the parameters of the model. To test

---

[1] School of Geosciences, University of Edinburgh, Edinburgh, UK.

0094-8276/09/2009GL039846

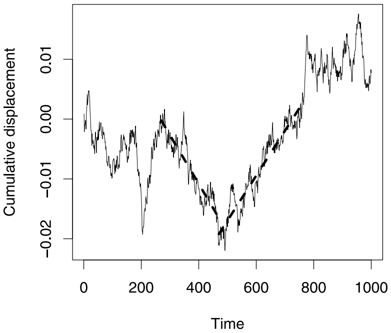

Figure 1. Solid line: random walk consisting of an independent increment (displacement) at each of 1000 intervals of 1 time unit. Increments are Gaussian-distributed with mean 0 and variance 0.001. Dashed lines: arbitrary examples of linear trends (R^2 = 0.84 and 0.92 respectively) arising randomly with no change in mean increment.

for AMR in data free from aftershock sequences, one could fit a non-stationary Poisson model whose rate increases as a power law with time. Although the parameters remain sensitive to data selection, they represent the most likely model for that selection which cannot, for the reasons discussed above, be claimed for least-squares fitting to cumulative data. Meaningful regression on cumulative data is very challenging [*Mizon*, 1995] and can indeed be avoided.

[5] **Acknowledgments.** The authors gratefully acknowledge financial support from the New and Emerging Science and Technologies Pathfinder program 'Triggering Instabilities in Materials and Geosystems' (NEST-2005-PATH-COM-043386), and from the EC FP6 NERIES project (RII3-CT-2006-026130).

References

Box, G. E. P., and G. M. Jenkins (1970), *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, Calif.

Greenhough, J., and I. G. Main (2008), A Poisson model for earthquake frequency uncertainties in seismic hazard analysis, *Geophys. Res. Lett.*, *35*, L19313, doi:10.1029/2008GL035353.

Hardebeck, J. L., K. R. Felzer, and A. J. Michael (2008), Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant, *J. Geophys. Res.*, *113*, B08310, doi:10.1029/2007JB005410.

Lawson, J. (2007), *Energy savings prediction: A case study, in Case Studies in Business, Industry and Government Statistics*, vol. 1, pp. 54–60, Bentley Univ., Waltham, Mass.

Marsan, D., and S. S. Nalbant (2005), Methods for measuring seismicity rate changes: A review and a study of how the M_W 7.3 Landers earthquake affected the aftershock sequence of the M_W 6.1 Joshua Tree earthquake, *Pure Appl. Geophys.*, *162*, 1151–1185, doi:10.1007/s00024-004-2665-4.

Mignan, A. (2008), Non-critical Precursory Accelerating Seismicity Theory (NC PAST) and limits of the power-law fit methodology, *Tectonophysics*, *452*, 42–50.

Mignan, A., and R. Di Giovambattista (2008), Relationship between accelerating seismicity and quiescence, two precursors to large earthquakes, *Geophys. Res. Lett.*, *35*, L15306, doi:10.1029/2008GL035024.

Mignan, A., G. King, D. Bowman, R. Lacassin, and R. Dmowska (2006), Seismic activity in the Sumatra-Java region prior to the December 26, 2004 (M_W = 9.0–9.3) and March 28, 2005 (M_W = 8.7) earthquakes, *Earth Planet. Sci. Lett.*, *244*, 639–654.

Mizon, G. E. (1995), A simple message for autocorrelation correctors: Don't, *J. Econometrics*, *69*, 267–288.

Ogata, Y. (1999), Seismicity analysis through point-process modeling: A review, *Pure Appl. Geophys.*, *155*, 471–507.

Phillips, P. C. B. (1986), Understanding spurious regressions in econometrics, *J. Econometrics*, *33*, 311–340.

Scargle, J. D. (2002), Was there evidence of global consciousness on September 11, 2001?, *J. Sci. Explor.*, *16*, 571–577.

Thejll, P., and T. Schmith (2005), Limitations on regression analysis due to serially correlated residuals: Application to climate reconstruction from proxies, *J. Geophys. Res.*, *110*, D18103, doi:10.1029/2005JD005895.

A. F. Bell, J. Greenhough, and I. G. Main, School of Geosciences, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JW, UK. (a.bell@ed.ac.uk; john.greenhough@ed.ac.uk; ian.main@ed.ac.uk)