

Anomalous Anticipatory Responses in Networked Random Data

Roger D. Nelson* and Peter A. Bancel**

**Global Consciousness Project, Princeton, NJ, USA, rdnelson@princeton.edu*

***Global Consciousness Project, Paris, France, pabancel@free.fr*

Abstract. We examine an 8-year archive of synchronized, parallel time series of random data from a world spanning network of physical random event generators (REGs). The archive is a publicly accessible matrix of normally distributed 200-bit sums recorded at 1 Hz which extends from August 1998 to the present. The primary question is whether these data show non-random structure associated with major events such as natural or man-made disasters, terrible accidents, or grand celebrations. Secondly, we examine the time course of apparently correlated responses. Statistical analyses of the data reveal consistent evidence that events which strongly affect people engender small but significant effects. These include suggestions of anticipatory responses in some cases, leading to a series of specialized analyses to assess possible non-random structure preceding precisely timed events. A focused examination of data collected around the time of earthquakes with Richter magnitude 6 and greater reveals non-random structure with a number of intriguing, potentially important features. Anomalous effects in the REG data are seen only when the corresponding earthquakes occur in populated areas. No structure is found if they occur in the oceans. We infer that an important contributor to the effect is the relevance of the earthquake to humans. Epoch averaging reveals evidence for changes in the data some hours prior to the main temblor, suggestive of reverse causation.

Keywords: global consciousness, psi, random events, RNG, anomalies, earthquake prediction

INTRODUCTION

A large database of consistently generated random numbers has been collected by an international collaboration of researchers called the Global Consciousness Project (GCP). The project maintains a global network of physical random event generators (REGs) distributed widely around the world.[1,2,3] Custom software records data at each node from random sources based on quantum tunneling in solid state junctions, and sends it via the Internet to a central server in Princeton, NJ. Beginning in 1998, the network has been generating a continuous, growing database of true random numbers, collected as parallel sequences from each of a number of REGs producing one 200-bit trial per second. The number of nodes in the network has grown to a current stable array of about 60 units. The data have characteristics that make them especially useful. After rigorous vetting for errors, they are empirically normalized. The computer clocks are synchronized, so the random number sequences are time-locked in association with

historical time. Thus there is a continuous record of nominally random, but potentially correlated data for any major event, and it is possible to do retrospective analyses assessing correlations with individual items or categories of events occurring any time during the history of the project. Techniques such as data mining as well as specific hypothesis testing can be applied to this unique database, with potential for insights in physical and social arenas.

A series of some 218 formal analyses has been conducted over the history of the project, assessing the data corresponding to major events in the world to test the hypothesis that non-random structure may be found correlated with the events. These include natural and manmade disasters and accidents, great celebrations, and organized events of various kinds that are expected to garner attention and engagement from very large numbers of people. The pre-planned analyses are based on canonical parametric statistics, including measures of correlation and variance in the data from the network of synchronized REGs.

The measured correlations are comparable to the level of statistical noise and single events do not reliably show identifiable structure. A notable exception is the response of the network on September 11 2001, which showed strong departures from random expectation, including a possible precursor response beginning some 4 or 5 hours before the first plane hit the World Trade Towers.[4] Figure 1 shows the cumulative deviation of the REG variance for Sept. 10-12, with the times of the attacks marked. The sharp rise beginning about 04:00 EDT and persisting for several hours, and the equally steep drop in the curve after 11:00 constitute a “spike” on 9/11 that is unique in three years of data accumulated up to that date. Permutation analysis indicates that the total deviation of the variance during the 24 hours (EDT) of Sept. 11 is equivalent to 3.12 sigma, or odds of about 1 in 1000.

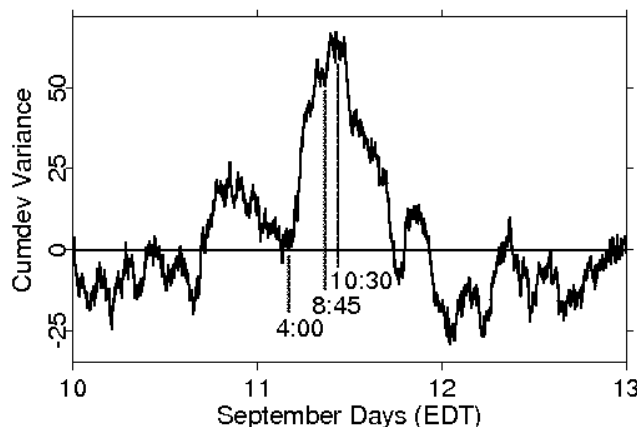


FIGURE 1. Cumulative deviation from expectation of second-by-second device variance across 37 REGs, September 10, 11, and 12, 2001.

Other measures on 9/11 also showed strong departures from expectation, but this event is unusual in the database for its relatively strong signal to noise (S/N) ratio. In order to learn more about the GCP network’s apparent response to major events, and especially to explore the suggestion of reverse causation, noise reduction and signal processing strategies are required. To this end, we have identified well-defined categories containing numerous conceptually similar events, and extracted the

corresponding data. These data are subjected to uniform statistical analyses, including signal averaging, to discover the nature and time course of replicated anomalous departures from expectation.

METHOD

The S/N ratio in the formal series of hypothesis tests is usually too small for individual events to yield deviations as clear as those shown in the data from 9/11. From the formal analyses, we estimate that 50 to 100 replications are needed in a given category in order to achieve a minimum power level of 90%. One subset of the formal analyses, comprising some 51 events, is defined by a sudden and unpredictable impulse at a specific moment in time, for example, a crash, a terror attack, or an earthquake. The original tests of the formal hypothesis for such events specify an analysis period beginning some time before the impulse and ending a few hours afterward. In explorations to discover more about the anomalous structure in the data, additional statistics were calculated for these events over the time period beginning 8 hours before and ending 8 hours after the defining moment.

The statistic most frequently employed for the original series of tests is a measure of network variance we call “netvar”.* It is calculated as a Stouffer Z across REGs each second, and then composed as a Chisquare distributed sequence of Z^2 per second over the duration of the defined event:

$$Z^2 = \left(\sum_{i=1}^N z(i) \div \sqrt{N} \right)^2 \quad (1)$$

Secondly, the distribution variance is calculated for all REG devices operating each second. It is called “devvar”. A third independent measure called “covar” is calculated as:

$$Cov = \sum_{j=1}^N \sum_{i=1}^{j-1} (z(i)^2 - 1)(z(j)^2 - 1) \div \sqrt{(N-1)N} \quad (2)$$

Because the signal to noise ratio is too small for individual cases to yield reliable statistics, we combine the uniformly defined events by epoch averaging.[5] Figure 2 shows the cumulative deviation of superposed, averaged data for the 51 impulse events in the three calculations. Cumulative deviation is a visualization tool used in process control engineering to identify changes in statistical parameters.[6] Its expectation is a level trend, and a change in the measured parameter is readily observed as a positive or negative slope when plotting the cumulative sum of the deviations from expectation. All three statistics show apparent structure persisting over several hours with slopes beginning near $T = 0$. The covar shows a clear positive trend over a period of 6 or 7 hours, possibly beginning about 2 hours before the event. A similar suggestion of an anticipatory change is seen in the devvar graph.

* We have named test statistics mnemonically. The “netvar”, or network variance, is the variance of the network output, taken as the mean REG value/sec. The “devvar”, or device variance, is the sample variance of raw REG values/sec. The “covar”, a complementary variance statistic, is closely related to the variance of the devvar statistic. The netvar and covar statistics are taken with respect to zero, their theoretical mean, a choice which renders them nearly independent of each other and the devvar, as established by empirical distribution analysis. The three statistics are well-approximated by the Chisquare distribution and can be summed accordingly.

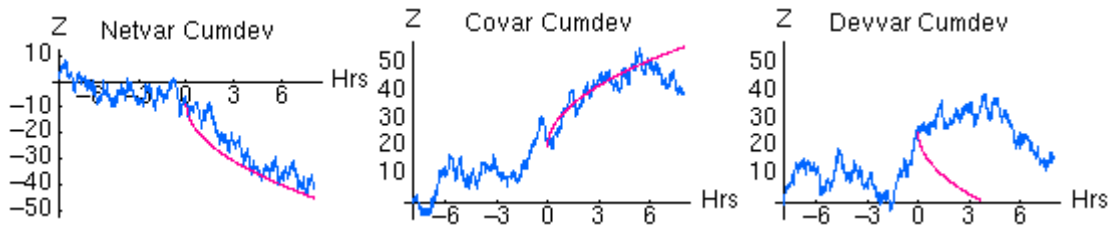


FIGURE 2. Signal average of data for 51 “impulse” events, $T = 0 \pm 8$ hours, plotted as the cumulative deviation from expectation of the Z^2 or netvar statistic (left), the covar statistic (middle), and the devvar statistic (right), as defined in the text. Smooth curves are 5% confidence envelopes.

RESULTS

The analyses of conceptually replicated tests in the historical database provide a *prima facie* case that impulse events, even though they vary in actual content and configuration, are a good place to look for correlated structure in the GCP data. The deviations from expectation accumulate for several hours, and are consistent enough to achieve statistical significance in the aggregate in both the netvar and the covar measures. Moreover, there is a suggestion in the covar measure that the anomalous deviation may begin before $T = 0$.

Impulse Events and Earthquakes

The indications of non-random structure, possibly including a time-displaced effect, correlated with assorted impulse events motivated a search for a larger and more consistent set of events of similar character. Major earthquakes constitute an ideal corpus. Their documentation includes precise timing, location, and strength estimates, and over the years we wish to examine, there are more than 100 quakes of the sort that garner widespread attention because of destruction and loss of life.[7] Using the tools developed for the impulse events, we did a preliminary assessment of all earthquakes from 1999 through early 2006 with Richter magnitudes $R > 7$. There are 111 such events, and within these there is a subset of 99 that are not included in and don't overlap any of the GCP impulse events. The same epoch averaging analysis was performed on the earthquake data and the impulse events. That is, the events were aligned relative to their defining point in time, $T = 0$, and superposed for the period $T = -4$ hrs through $T = +8$ hrs.

Figure 3 compares the cumulative deviation traces of the test statistics for the two event sets. Both the netvar and the covar statistics yield traces for the earthquakes that are similar to the impulse events. Again there is an indication that the covar measure begins to depart from a normal random walk well before the occurrence of the defining impulse, the primary temblor of the earthquake. While we cannot rule out the possibility that the similarities between the traces may be accidental, this outcome suggests that a careful assessment of GCP data corresponding to earthquakes can provide useful insight into the character of the network response to major events. In particular, we wish to know whether quake parameters such as magnitude, intensity

of damage, location and attendant news coverage correlate with the statistical deviations. We present an approach to addressing these aspects below.

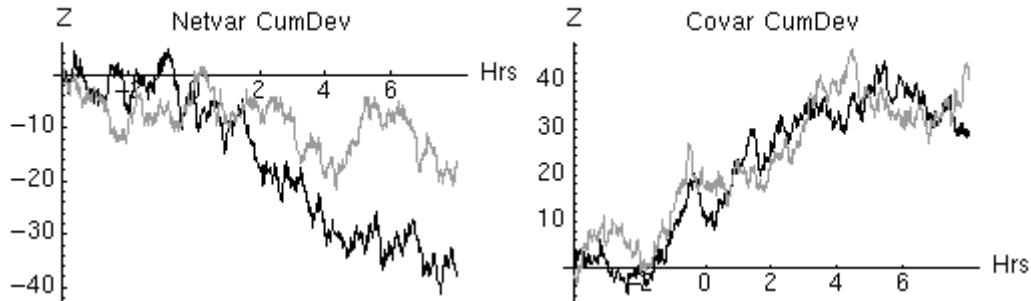


FIGURE 3. Cumulative deviation of 99 signal averaged earthquakes, Richter magnitude > 7.0 (upper, light grey curves), compared with cumulative deviation of 51 “impulse” events. Left panel shows Z^2 or netvar statistic, right panel shows covar.

Earthquake Database

We would like to select earthquakes which draw the attention of large numbers of people and test these against the occurrence of earthquakes having little human impact. While there is no obvious way to make such a determination, we can guess that likely candidates will be strong quakes occurring in populated areas. Accordingly, we have divided the database of major earthquakes[7] into those on land that may cause destruction and deaths, and a larger number that occur in the oceans, with little consequence to humans. We choose only quakes with measured magnitudes of $R \geq 6$, the conventional criterion for designating strong earthquakes in seismology. The selection yields over 180 earthquakes in the period from 2000 to early 2006, during which time the GCP network has had at least 25 nodes in continuous operation. In the same time period, there were over 500 quakes in the oceans and regions of low population. The map in Figure 4 shows rectangles circumscribing populated regions prone to quakes. The locations of most of the REGs in our network are within these four areas, and are indicated in the map by light

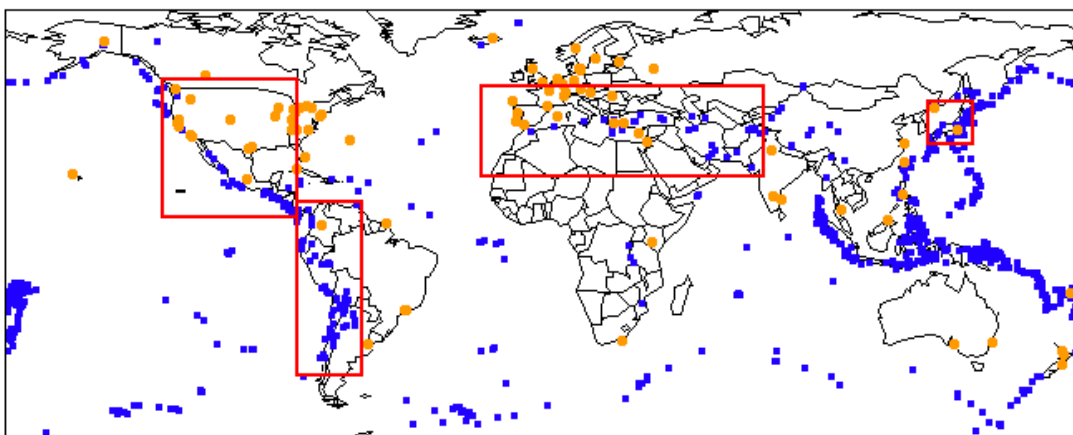


FIGURE 4. Black dots are earthquakes with magnitude $R \geq 6$ from 2000 to 2006. Rectangles enclose populated region subsets defined for analysis. Lighter dots show location of the network nodes.

colored dots. An indicator which may be of importance is the distance of the earthquake location from the GCP nodes. Two of the selection regions, North America and Eurasia, encompass the majority of the network nodes. Quakes within these regions are at a median distance of 6 or 7 thousand kilometers from the REGs. We will focus on them for most analyses. The South American and Nippon regions include only one or two GCP nodes, and they circumscribe quakes which are on average more distant from REG devices. Their median quake-to-network-node distances are over 8 thousand kilometers.

Figure 5 looks at strong quakes using three independent measures. All calculations, regardless of regional selection, use the full network of REGs (25 to 60 REGs depending on the year). They are normalized as Z scores, which are in units of standard error or sigma. The plots compare the means of the netvar, covar, and devvar statistics for three periods before and after the main temblor of the quakes. The left panel plots Z scores for data blocks extending 1.5, 3, and 7 hours before time $T = 0$. The right panel of the figure shows the same calculations for times after the primary temblor.

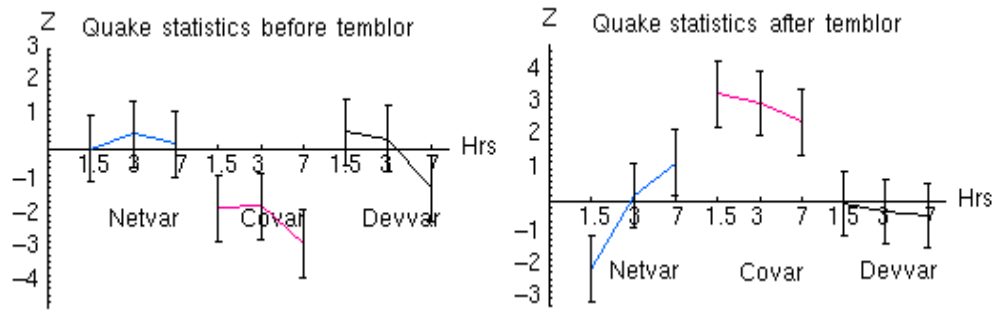


FIGURE 5. Response of netvar, covar, and devvar statistics before and after earthquakes in North America and Eurasia. Left panel, pre-temblor; right panel, post-temblor responses. Error bars are 1σ .

The covar measure shows a larger and more consistent deviation than the other statistics, and we will focus on it for subsequent analyses. For this dataset, the covar statistic is significantly lower than expected for all three periods prior to the quake. For the 7-hour pre-temblor period the point estimate is $Z \sim -3.2$. The post-quake deviation of the covar measure also departs substantially from expectation. It is higher than expected for each of the three periods, with Z scores ranging from about 2.7 to 3.3.

The structure is more starkly apparent in Figure 6, which shows the covar measure for North American and Eurasian quakes, compared with ocean quakes, calculated for -8 hours, $+8$ hours, and ± 8 hours centered at the time of the temblor. The difference between land and ocean quakes is visually obvious, and calculation shows it is significant. Whereas the data collected during quakes in populated land areas shows strong deviations ($Z = -3.3$ before and $+2.6$ after), the data for quakes in the oceans are unremarkable, with Z scores less than 0.5 for all three calculations.

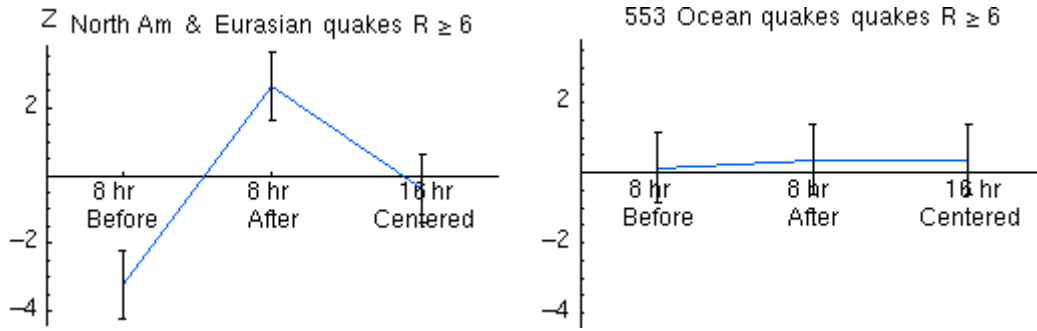


FIGURE 6. Deviation of covar for quakes with $R \geq 6$, 8 hours before and 8 hours after, and 16 hours centered at the time of the quake. Left panel, North America and Eurasia, $N=86$. Right panel, ocean regions, $N=553$.

Timing of the Response

Given the suggestions of a precursor response in the networked random data corresponding to major earthquakes in populated areas, a more focused analysis is in order. First we wish to confirm that there is a correlated response, and then look carefully at the question of timing. In particular, we seek to determine whether there is solid evidence for a response that precedes large earthquakes.

As before, the analysis proceeds by aligning the GCP data corresponding to the quakes on the moment of the main temblor. Then an epoch average is calculated to look specifically at the timing of the GCP network reaction to the quake events. The quakes are again separated into two categories: those in heavily populated regions where we have REGs (North America and Eurasia) and a complementary set in the oceans and less populated areas. Aftershocks and other quakes that overlap within 20 hours are removed to clarify the signal average centered on the primary temblor. The process minimizes discards and keeps the strongest quakes; most aftershocks are relatively weak.

Figure 7 presents the epoch averaged raw covar data over 60 hours centered on the primary temblor of 86 quakes with $R \geq 6$ occurring in the two selected regions, and a smoothed version of the same data. Close examination reveals a hint of structure in the center of the raw data graph, near time zero. This is clarified in the smoothed data,

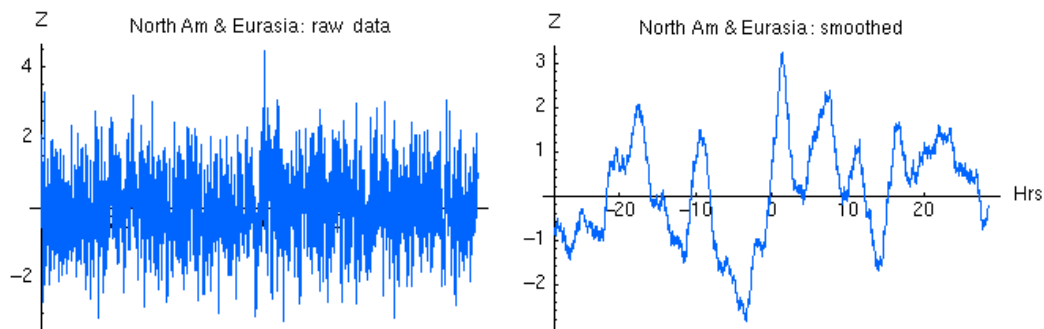


FIGURE 7. Left panel, raw data for 86 quakes in populated regions, North America, Europe, Asia. Right panel, the same data with Gaussian smoothing using a three hour window. $T = 0 \pm 30$ hours.

which shows that for several hours prior to the quake the covar statistic is persistently low, then changes to persistently high values with the crossover exactly at $T = 0$. The spikes before and after $T = 0$ are extreme relative to the usual variation, and both reach Z scores of about 3.

Figure 8 presents the same superposed data in the cumulative deviation format, which displays a 60-hour history of the level of covar response, plotting the accumulating sum of deviations from expectation. The steep trend beginning approximately 8 hours before the quake represents a period of unusually persistent low values of the covar statistic, and it indicates that a change is registered in the GCP data before the quakes strike.

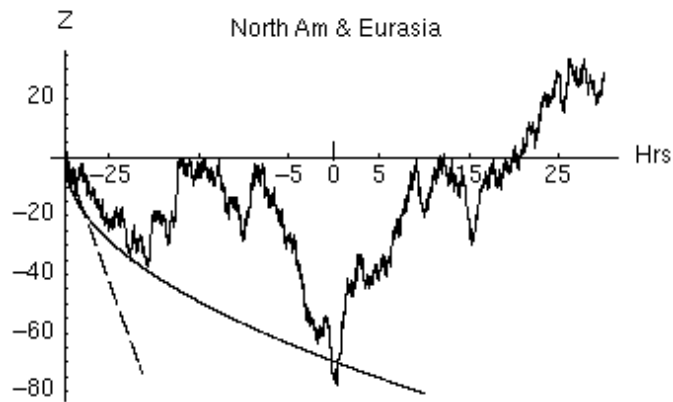


FIGURE 8. Cumulative deviation of covar from expectation. Signal average of 86 earthquakes, Richter 6 or more, North America and Eurasia. Resolution is one Z score per minute. The smooth curve is a 95% confidence envelope for the cumulative deviation from the origin. A dashed line approximates the slope of the pre-quake trend.

Since these analyses are exploratory, rather than tests of pre-stated hypotheses, we do not have rigorous statistics to support strong interpretations. However, it is worthwhile to consider what is visually apparent. For the first 20 hours shown in the figure, the data have the expected character of a random walk, but the overall picture shows striking departures from the general trend during several hours centered on the defining temblor of the quakes, the zero point. The scale of the effect can be gauged by comparing the slope of the trends against the smooth curve representing 95% confidence. A dashed line parallel to the pre-quake trend, and of equal duration, gives a visual comparison standard that suggests the steep drop beginning about 8 hours before $T = 0$ is significant.

A numerical estimate for the likelihood that the symmetrical dip around $T = 0$ is a chance fluctuation was made using a permutation analysis (10,000 iterations) to calculate a probability for the depth of the minimum and its proximity to $T = 0$. A statistic representing the likelihood of the focused deviation was constructed by combining the corresponding Z scores as $(Z_{\text{depth}} + Z_{\text{prox}}) / \sqrt{2}$. For the cumulative deviation of covar shown in Figure 8, the result is $Z \sim 3.852$. Correcting for multiple analysis and the sign of the excursion reduces this Z score to about 3.2.

When we look at the same kind of analysis for quakes that occur in the ocean or in parts of the world where the population density is much smaller the result is very different. There is no apparent structure related to the time of the quake. Figure 9

shows the result for all quakes that occur in ocean areas not included in our North America-Eurasia regions. There are 553 such quakes of Richter magnitude 6 or larger during the period of our analysis. The cumulative deviation of the covar for this

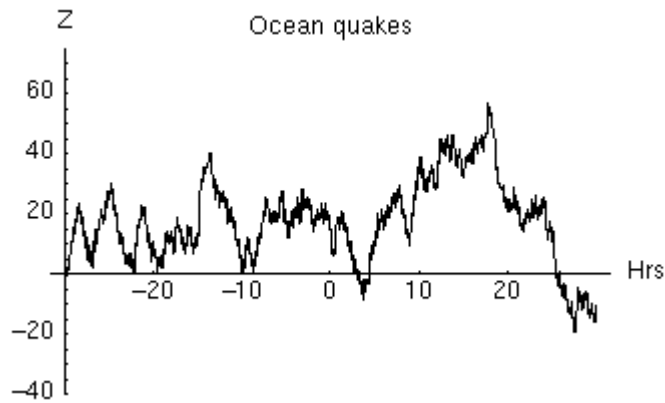


FIGURE 9. Cumulative deviation of covar statistic, 553 signal averaged earthquakes, Richter 6 or more, unpopulated and ocean regions. $T = 0 \pm 30$ hours.

subset of quakes, using the same epoch analysis as in the preceding plots, shows no notable structure around the central point of $T = 0$, and the permutation-based estimate for the combined Z scores representing depth and proximity is $Z \sim 0.177$.

Location of the quakes in populated areas as opposed to unpopulated ocean regions is important in determining the effect on the GCP network, and we can link this observation inferentially to the effects on human populations of severe, destructive earthquakes. The network responds to quakes on land that can affect people, but not to quakes in the oceans where they go largely unnoticed except by seismologists.

The consistency of the response in the two REG-proximate, populated regions is apparent in Figure 10. It shows the period of 16 hours centered on $T = 0$ for the 37 Eurasian and 49 North American quakes separately. Both subsets show the same general shape albeit with differences in detail, suggesting that the anomalous effect is widely distributed geographically.

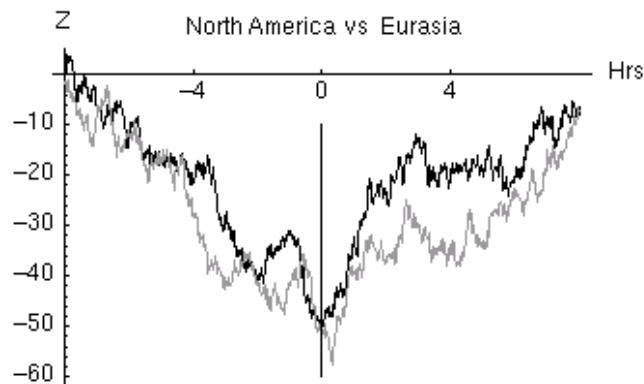


FIGURE 10. Cumulative deviation of covar for North America (black upper curve, 49 quakes), Eurasia (grey lower curve, 37 quakes). The span of time is ± 8 hrs around the primary temblor.

Discussion

The earthquake data are complex and it appears that a simple criterion of large Richter magnitude for continental earthquakes is only one of multiple predictors of fluctuations in the statistics. Further progress will require testing specific hypotheses relating the deviations to the magnitude, location and significance of future earthquakes.

We have demonstrated that deviations occur in the covar statistic for data from the global REG network around the time of strong land-based earthquakes in North America and Eurasia. We note, however, that data corresponding to earthquakes with $R \geq 6$ in South America and the Nippon regions do not show the characteristic change in sign at the time of the main temblor found in our analysis of the data for North America and Eurasia. In the left panel of Figure 11, the cumulative deviation for 53 South American quakes exhibits a significant decline prior to $T = 0$, similar to the North American and Eurasian quakes, but it does not revert to a positive trend immediately after the quake. In the Nippon region, a sample of 47 quakes with $R \geq 6$ shows deviations within expectation (albeit with some segments showing persistent trends over periods of a few hours).

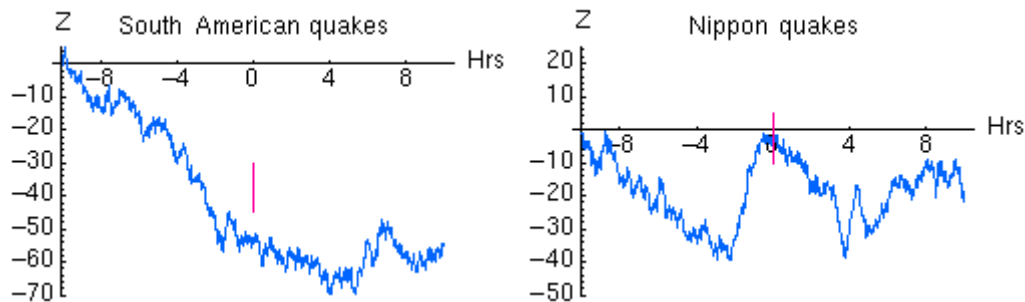


FIGURE 11. Cumulative deviation of covar from expectation. Left panel shows 53 South American Quakes, right panel shows 47 quakes in the Nippon region. $T = 0 \pm 10$ hours, $R \geq 6$.

It may be important that these regions are further removed from the center of mass of the network than the North American and Eurasian quakes, and that they have relatively sparse REG network representation. We are currently studying whether the statistical deviations depend on the distance of the quake location relative to the geographical deployment of the network.

We also have looked at quakes that definitely resulted in fatalities ($N > 5$). In order to accumulate a usable sample we had to accept a smaller Richter magnitude ($R > 5$ instead of $R > 6$), and expand the geographical selection to zones of lower population. Figure 12 shows the cumulative deviation of the covar statistic for the resulting dataset of 40 quakes. The overall curve lies within expectation, although there is a persistent trend from $T = -6$ hours to $T = 0$ that is marginally significant. This decline prior to temblor time is reminiscent of the data for the North American and Eurasian regions, and the overall picture is similar to the South American subset.

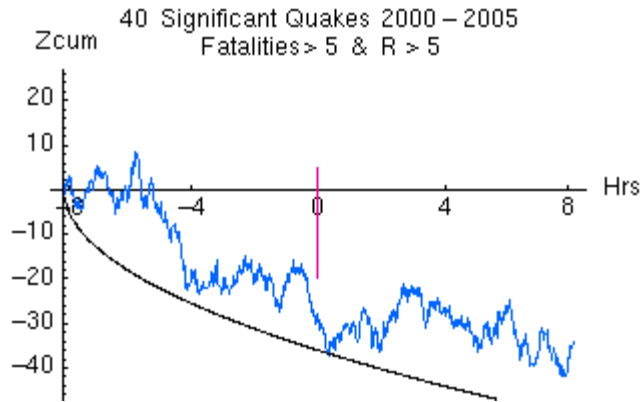


FIGURE 12. Fatal quakes with more than 5 deaths. The magnitude criterion is lowered ($R > 5$) and the selection area enlarged to accumulate 40 quakes. $T = 0 \pm 8$ hours.

It is clear that our analyses are hampered by the small effect size. For the fatal quake subset, which arguably should be a good test case, we are forced to relax the selection criteria in order to obtain an adequate number of samples. Unfortunately, this complicates comparisons because the character of the dataset is categorically changed. Although earthquakes are numerous, these concerns indicate the limits of what can be achieved with the database currently in hand.

To assess more directly the effect of changing the quake selection criteria, we identified a set of quakes in the North American-Eurasian regions with smaller magnitude to compare with the main analyses shown in Figs. 8 and 10. Over the range of $R = 5$ to $R = 5.4$ we find 360 land-based quakes. The covar response during these quakes is shown in Figure 13. The results exhibit no structure around $T = 0$, in contrast to the database of quakes with magnitude $R > 6$, which show a statistically significant pattern of deviations using the same analyses. Of course the Richter scale is logarithmic, and the level of destructive energy in the $5 > R > 5.4$ range is considerably less than for $R > 6$. The difference in energy released during quakes

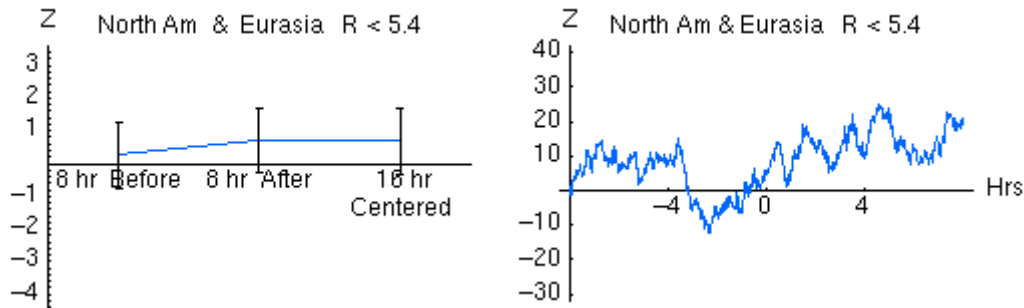


FIGURE 13. North America and Eurasia, 360 quakes, Richter magnitude 5 to 5.4. The left panel shows point estimates for 8 hours before, 8 hours after, and 16 hours centered on $T=0$. The right panel shows the cumulative deviation of the covar from expectation over ± 8 hours.

with $R = 5$ compared with $R = 6$ is a factor of 100 or more. While we will need to work carefully toward a better understanding of the factors that underlie the anomalous structure in GCP data associated with large earthquakes, this analysis

nominally supports the view that the amplitude of measured structure correlates with Richter magnitude.

CONCLUSIONS

The data archive examined here should exhibit no non-random structure, and arbitrary calibration probes show results in accord with this expectation. But there are periods of statistically significant structure in the GCP data, correlated with a class of momentous events defined by the attention given to them by very large numbers of people. These include mass celebrations like New Years and religious holy days, as well as disasters, terror attacks, and major earthquakes. The structure corresponding to events defined by an “impulse” at a specific time shows an important additional aspect: the departures from random expectation appear to begin prior to the nominal precipitating event. This anticipatory response is evident for large earthquakes occurring on land, but not in the oceans. The effect is subtle, as is the basic correlation of structure with events. While the anticipatory pattern is largely consistent, there are variations for differing geographic regions. Although we cannot interpret our results as unambiguous evidence for retro-causation, they do constitute an empirical conundrum. We do not have any other viable explanation for the appearance of structure associated with, but prior to a precipitating event.

We have established that the results we see are not artifacts of electromagnetic fields or physical biases.[4,8] The repeatability of the findings in independent datasets, and in alternative statistical measures, helps assure that we have not made fundamental errors. Nevertheless, the nature of the finding is sufficiently out of the mainstream of physical understanding that it needs to be studied further, and by independent researchers. The data from the GCP are fully documented and publicly available, allowing independent assessments. We welcome comments and invite replication of our analyses.

ACKNOWLEDGEMENTS

The Global Consciousness Project is a collaborative effort of individuals at many institutions around the world. We are especially indebted to Richard Adams, Paul Bethke, Greg Nelson, Dean Radin, John Walker, and the hosts for the REG network nodes.

REFERENCES

-
- 1 Nelson, R. D., “Global Consciousness Project,” (1998), [Hhttp://noosphere.princeton.edu](http://noosphere.princeton.edu), accessed June 16 2006.
 - 2 Nelson, R. D., “Correlation of Global Events with REG Data: An Internet-Based, Nonlocal Anomalies Experiment” *Journal of Parapsychology* **65**, 247-271 (2001).
 - 3 Nelson, R. D., “Coherent Consciousness and Reduced Randomness: Correlations on September 11, 2001.” *Journal of Scientific Exploration* **16**, 549–570 (2002).
 - 4 Nelson, R. D., Radin, D. I., Shoup, R., and Bancel, P. A., “Correlations of Continuous Random Data with Major World Events.” *Foundations of Physics Letters* **15**, 537-550 (2002).

-
- 5 Applications Weekly, "Epoch Averaging to Extract Weak Signals from Noisy Data," (1998) <http://www.gage-applied.com/resource/apps1998/981028.htm>, accessed Feb. 25 2004.
 6. Woodall, W. H., and Adams, B. M., The Statistical Design of CUSUM Charts, *Quality Engineering* **5**, 559-570 (1993).
 - 7 Advanced National Seismic Survey, <http://earthquake.usgs.gov/research/monitoring/anss/>, accessed June 16 2006.
 - 8 Radin, D. I., Exploring Relationships Between Random Physical Events and Mass Human Attention: Asking for Whom the Bell Tolls. *Journal of Scientific Exploration* **16**, 533-548 (2002).