GEOPHYSICAL RESEARCH LETTERS, VOL. 23, NO. 11, PAGES 1315-1318, MAY 27, 1996

VAN earthquake predictions – an attempt at statistical evaluation

Y. Y. Kagan

Institute of Geophysics and Planetary Physics, University of California, Los Angeles

Abstract. Statistical verification of the VAN or SES (seismic electric signals) predictions from 1987-1989 is considered. The test is carried out with the updated rules proposed by *Varotsos et al.* [1996]. Although for the Greek (SI-NOA) earthquake catalog the VAN method formally is successful, this high rate of success is due either to the retroactive adjustment of prediction rules or to the non-randomness of seismicity. A simple prediction algorithm accounting for earthquake clustering (foreshock-mainshock-aftershock sequences), yields similar or even better forecast results. If we remove dependent events from the catalog, the 'prediction effect' becomes statistically insignificant. For the PDE (NOAA) catalog the test shows that the VAN predictions' rate of success can be attributed to chance.

Introduction

Varotsos et al. [1996] examine the history of the application of seismic electric signals (known as the VAN or SES precursors) for earthquake prediction, as well as efforts by several groups [*Hamada,* 1993; *Mulargia and Gasperini,* 1992, 1993; *Shnirman et al.,* 1993; *Takayama,* 1993] to verify the prediction results statistically. These validation attempts lead to conflicting conclusions: on the one hand, *Mulargia and Gasperini* [1992, 1993] state that the successful VAN predictions can be attributed to a random chance, whereas *Hamada* [1993], *Shnirman et al.* [1993], and *Takayama* [1993] find a significant correlation between the VAN signals and ensuing earthquakes. *Varotsos et al.* [1996] summarize the discussion, propose more detailed and specific prediction criteria and offer basic rules for testing the predictions. In particular, they argue that magnitude, time and location of the prediction alarm and each subsequent earthquake needs to be analyzed in order to verify the prediction. Moreover, during a statistical test of the prediction, the non-Poisson behavior of earthquake sequences needs to be properly accounted for. These rules take into account the basic requirements for formal, rigorous statistical testing of forecasts; however, an additional distinction should be made between retrospective and forward predictions.

Since we do not have a general comprehensive theory of earthquake occurrence, we apply empirical techniques for earthquake prediction. Therefore, any perspective forecast method should be tested at two stages: (1) a learning period, when suitability of the method is ascertained, and values of adjustable parameters established; and (2) a control stage, where no parameter fitting is allowed [*Jackson,* 1996]. If the VAN method authors had followed the procedure, it would be significantly easier to reach definite conclusions. However, even though the method has been known since 1981, no such test has been carried out. Even now there is some uncertainty about the values of the VAN basic parameters: for example, *Varotsos at al.* [1996] indicate that *Mulargia and Gasperini*'s [1992] analysis is deficient since they used a single time period for the VAN alarms (11 or 22 days), whereas *Varotsos at al.* [1996] suggest three different lead times. However, *Hamada* [1993] and *Shnirman et al.* [1993] also use a single period (22 and 42 days, respectively), and *Varotsos at al.* [1996] seem to accept their results.

The testing of retrospective predictions requires proper accounting for degrees of freedom employed by the forecast model [*Molchan and Rotwain,* 1983]. These degrees of freedom include explicit adjustable parameters of the model and selection of 'hidden' parameters: the earthquake catalog, spatial boundaries, beginning and ending times of the test, etc. Although the statistical theory can accommodate some parameter adjustment, in general, the validation becomes more difficult, especially when the number of degrees of freedom is comparable to the number of successful predictions. For example, by granting an exception to the rules, as in *Hamada* [1993, p. 207], we can arbitrarily increase the success rate. Similarly, *Shnirman et al.* [1993] find four earthquakes successfully predicted by the VAN method, with one failure: at least several variables (magnitude cutoff, alarm duration, catalog selection) have been modified to achieve the result.

I define M earthquake magnitude; M_c magnitude cutoff of a catalog; M_a magnitude of predicted earthquake; ΔM magnitude difference between prediction and earthquake ($M_a - M$); ΔT time difference in days between prediction and earthquake; N_q the total number of earthquakes during an observation period (i.e., those events which could be in principle predicted); n_q the number of successfully predicted earthquakes; N_a the total number of alarms; t_a alarm time; T time span of a catalog; R_{max} radius of an alarm circle; $\nu = (N_q - n_q)/N_q$ fraction of failures to predict [*Molchan and Kagan,* 1992]; $\tau = \sum t_a/T$ fraction of alarm time, or more generally fraction of alarm time-space; μ expected number of prediction successes; α significance level.

Paper number 95GL03417
0094-8534/96/95GL-03417$05.00

VAN predictions and tests

Varotsos et al. [1996] propose that VAN alarms refer only to earthquakes with magnitude greater or equal to 5 ($M_c = 5.0$, $M_a \geq 5.0$), $\Delta M \leq 0.7$, and three limits for t_a (11 days for a single SES, 22 days for several SES signals, and t_a '... of the order of 1 month...' for "gradual variation of electric field" – GVEF). They mention two distance limits R_{max}, 30 km and 50 km without indicating their final choice; I take $R_{max} = 100$ km as in *Hamada* [1993]. Thus, there are 7 formal adjustable parameters in the VAN method.

Presently there is no consensus on the rules for earthquake prediction validation. The often mentioned failure-to-predict (ν) versus false alarm criterion is clearly unsatisfactory, since a trivial strategy of alarm declaration for the entire region leads to zero errors. *Molchan and Kagan* [1992] propose the following three criteria to judge the effectiveness of earthquake prediction: τ and ν (fraction of alarm time and fraction of missed events), as well as N_a, the number of alarms. The ratio $(1 - \nu)/\tau$ is a probability gain of a prediction technique; another way to measure the effectiveness of a prediction is the sum $\nu + \tau$ [*Molchan and Dmitrieva,* 1992].

A significance level of 5% or less is usually considered sufficient to reject a null hypothesis. However, if adjustable parameters have been used in a test, or several different cases have been investigated, the α level must be reduced [*Molchan and Rotwain,* 1983; *Molchan and Dmitrieva,* 1992].

I use the list of the VAN predictions from January 1, 1987 to December 1, 1989, ($T = 1065$ days) [*Dologlou,* 1993]. The predicted epicentral area is indicated in terms of azimuth and distance from Athens. *Mulargia and Gasperini* [1992] convert these directions into geographical coordinates. I have removed from the list all the predictions with $M_a < 5.0$. Only one prediction (on April 27, 1987) was issued using the GVEF rule, (item 'b' in Table 1 of *Dologlou* [1993]), and the only qualifying earthquake in the SI-NOA catalog occurred on May 29, 1987, i.e., separated by more than 32 days from the prediction. The statement defining this prediction (see above) is not numerically specific enough to be used in the formal tests. Therefore, I also exclude this prediction from consideration. Thus we consider 28 predictions. I assume that an alarm can predict more than one earthquake.

I use two earthquake catalogs: the Greek SI-NOA catalog (Table 1 in *Geller* [1996]) for 1987-1989, and the PDE (NOAA) catalog [*U.S. Geological Survey,* 1993] for 1965-1993 (Table 2 in *Geller* [1996]). The magnitude in the SI-NOA catalog is used directly; for the PDE catalog I add 0.3 to magnitude m_b [*Hamada,* 1993] to make it compatible with the VAN predictions. The catalogs' space window is 17-27°E, 35-42°N; the hypocenter depth limits 0-70 km. The total number of earthquakes $M \geq 5.0$ for the prediction time period in the SI-NOA catalog is 46 (36 of them are less than 100 km away from any of the prediction points, $N_q = 36$). The PDE catalog has 39 events with $m_b \geq 4.7$ in the space window, $N_q = 33$.

Table 1 displays an abbreviated list of predictions, and seven earthquakes from the SI-NOA catalog which formally satisfy the prediction criteria. How can we evaluate whether the seven successful predictions shown in Table 1 are a chance coincidence? *Mulargia and Gasperini* [1992, 1993], *Hamada* [1993], *Shnirman et al.* [1993], and *Takayama* [1993] have considered this problem. I see certain deficiencies in the above tests: (1) Since *Varotsos et al.* [1996] specified the forecast parameters, it is necessary to repeat the analysis with updated rules. (2) Earthquake clustering, which manifests itself mostly in aftershock sequences, has not been properly taken into account. (3) The probability of earthquake occurrence is often calculated using $N_a t_a$ over all the predictions issued. The value is incorrect when $N_a t_a$ is comparable in size with T. For example, for $t_a = 22$ days and 29 predictions [*Varotsos et al.,* 1996], $N_a t_a = 0.6\ T$. Using formula (2.17) of *Molchan and Rotwain* [1983], I calculate that the average total alarm time is 0.45 T, i.e., significantly smaller than the value above. The actual measurement for the 29 VAN predictions in 1987-1989, using $t_a = 22$ days, yields the value $\tau = 0.42$. For the t_a values suggested by *Varotsos et al.* [1996] and 28 predictions, $\tau = 0.25$. (4) Loca-

Table 1. List of VAN earthquake predictions and earthquakes

VAN Predictions					Earthquakes						
No	Date yr/mo/da	Coordinates degrees		M_a	Date mo/da	Coordinates degrees		M	R km	ΔT days	ΔM
1	1987/2/26	37.94	20.32	6.5	2/27	38.40	20.42	5.9	48.4	2.0	0.6
2	1988/5/15	37.94	20.32	5.3	5/18	38.35	20.47	5.8	47.5	3.2	–0.5
					[a]5/22	38.35	20.53	5.5	49.4	7.3	–0.2
3	1988/5/30	37.94	20.32	5.4	[b]6/2	38.27	20.37	5.0	36.8	3.4	0.4
					[b]6/6	38.30	20.48	5.0	42.6	7.3	0.4
*4	1988/9/1	37.96	21.01	5.8	9/22	37.98	21.12	5.5	10.0	21.5	0.3
*5	1988/9/30	37.96	21.01	5.3	[b]10/16	37.90	20.97	6.0	7.2	16.5	–0.7

* signifies series of SES predictions, $t_a = 22$ days;

[a] dependent event (aftershock) is removed according to Equations (1) and (2) with $T_c = 1$;

[b] dependent events are removed when $T_c = 0$.

tions of earthquake epicenters and alarms were either not taken into consideration [*Shnirman et al.*, 1993] or are accounted for in an approximate fashion [*Hamada*, 1993]. (5) In all of these tests, the significance level has been calculated using a single average value for μ. This might lead to overestimation of α, as the following example shows. Suppose five successful alarms with $t_a = 0.1\ T$ have been issued in five spatially separate regions; only one earthquake occurred in each of regions 1-4, whereas in the fifth region 21 earthquakes are registered. (Such different success levels might be attributed to a high level of cultural noise in the fifth region.) The average $\mu = 2.5$; a test based on this value would yield $\alpha = 11\%$ (using the Poisson approximation), invalidating the method. However, if we calculate the probability of a chance success in each region separately, the combined $\alpha < 10^{-4}$.

The test method which does not have the defects described above is Monte Carlo simulation: we randomize the occurrence time of earthquakes in a catalog, keeping their epicentral position and magnitude unchanged, and then perform calculations such as in Table 1, with a synthetic catalog. This procedure is repeated many times in order to obtain a statistical distribution of 'success rate' with a randomized catalog. This distribution serves as a pattern against which we test the real prediction (Table 1).

Using 10^6 realizations, I find that $\mu = 3.26$, μ/N_q corresponds to the fraction of space-time alarm, τ. The ratio n_q/μ is the probability gain. According to simulation results, the probability that seven or more earthquakes occurred in the prediction zones by chance is 3.6%; the Poisson approximation using the μ value above yields $\alpha = 4.8\%$. In interpreting these results one needs to take into account that the above seven successful VAN predictions were obtained after the prediction rules were retroactively adjusted using all the available data. There are at least six explicit degrees of freedom in the VAN model tested here (see above). In principle, this should disqualify any statistical test result [*Jackson*, 1996]. Let us disregard the above problem for a while, to see whether we can reach any conclusion with the available information. The following tests could be used as a model for the statistical analysis of future VAN and similar predictions.

The values of α indicate that the null hypothesis – the success of the VAN predictions due to chance – is rejected. However, the above conclusion is crucially dependent on the condition that earthquakes occur according to a Poisson process, but shallow earthquakes display strong non-Poissonian features [*Varotsos et al.*, 1996].

The test described above, has one potentially serious flaw: earthquakes during three year period 1987-1989 are assumed to represent the seismicity in Greece. Earthquakes cluster both in time and space; therefore the use of a relatively short catalog can cause a bias in our results. With the PDE catalog we can test whether a 3-year sample is sufficient. Analysis of the PDE catalog yields six successful predictions (entries 17, 18, 19, 20, 22, and 24 in Table 2 [*Geller*, 1996]). I produce two sets of simulations: one using 1987-1989 data, and the other for the full catalog. Both simulations yield similar results: for the PDE catalog $\alpha = 11.6\%$ in the first case and 9.9% for 1965-1993 data, i.e., the prediction of all six events could be attributed to chance.

Earthquake clustering

It is relatively simple to devise an earthquake prediction scheme which would perform better than a random chance: we need to declare an alarm after each strong earthquake; most of the following earthquakes would be aftershocks, but occasionally an ensuing event would be stronger than the first earthquake, thus making the first earthquake a foreshock [*Molchan and Kagan*, 1992, and references therein]. Any prospective earthquake prediction technique needs to demonstrate that its success is not due to the influence of earthquake clustering (foreshock-mainshock-aftershock sequences).

Two techniques could be applied to account for earthquake temporal clustering: (1) we can decluster a catalog and repeat the measurements; (2) we can show that by using the information about clustering, it is possible to replicate 'successful' prediction results.

Declustered catalogs

Since there is no standard procedure for aftershock removal, it is important to agree in advance which method will be used to avoid possible bias. As a preliminary scheme I apply a variant of *Reasenberg*'s [1985] method for aftershock identification. In this procedure we calculate a time-distance window around each earthquake in a catalog, and declare all earthquakes inside the window to be dependent events.

In particular, the distance limits (in km) are calculated according to the formula [cf. *Reasenberg,* 1985]

$$R(M) = R_c \times 10^{(1.2M-4)/3}. \qquad (1)$$

The aftershock zone sizes are 4.6 km for $M = 4$ and 46 km for $M = 6.5$, if we take $R_c = 2.5$ km. For the time limits I use the following formula which is a variant of that used by *Reasenberg* [1985]

$$T(M) = \frac{10}{3} \times 10^{2(M-4-T_c)/3}. \qquad (2)$$

Equation (2) yields a time limit of 3.33 days for $M = 5$ and 33.3 days for $M = 6.5$ for $T_c = 1$; for $T_c = 0$ the above limits are obtained for $M = 4$ and for $M = 5.5$, respectively.

If an earthquake is in the time-distance window of another event, it is deleted from the declustered catalog. As a result of the application of (1) and (2) to the original catalog for $T_c = 1$ we delete only one aftershock from the SI-NOA catalog; for $T_c = 0$ one foreshock and three aftershocks are found (see Table 1). We also need to correct the prediction list for the clustering, otherwise the application of the predictions to a declustered catalog would yield a biased result. It is not easy to devise a formal, unambiguous scheme for the prediction 'declustering.' I keep only predictions that are separated by more than 11 days from the previous forecast. As a result, nine out of 28 predictions have been removed (entries 'g,' 'h,' 2, 4, 5, 8, 12, 16, and 19, see Table 1 in *Dologlou* [1993]). The simulation process, similar to the process described above, produces the following outcome: for $T_c = 1$, $N_q = 30$, $n_q = 6$, $\mu = 2.5$, and $\alpha = 3.3\%$; for $T_c = 0$, $N_q = 27$, $n_q = 3$, $\mu = 2.1$, and $\alpha = 35\%$. This means, that depending on the declustering algorithm, we can either reject the null hypothesis or accept it.

Alternative prediction

As another test I declare an alarm after each of 46 earthquakes with $M \geq 5$ in the SI-NOA catalog in a time-space window defined by Equations (1) and (2). After brief experimentation with two adjustable parameters, the following values $T_c = 0$ and $R_c = 6.25$ km yield 7 'successful' predictions, the number of successes for the VAN forecasts (see Table 1). The value of μ is 2.20, smaller than that found by the VAN method, $\alpha = 0.4\%$. Although the total number of alarms (46) is larger than that in the VAN procedure (depending on how one counts double predictions, we get either 28 or 43 VAN alarms in 1987-1989), this technique uses only two degrees of freedom versus 6 degrees for the VAN method. Moreover, we do not need any additional (electrical) measurements to achieve the prediction performance of the VAN method.

Discussion and conclusions

The results of statistical tests of the VAN earthquake prediction method suggest that although the technique formally is successful, its success might be due either to the posterior adjustment of the prediction rules, or to clustering of shallow earthquakes. A simple prediction algorithm, accounting for non-random seismicity, yields similar forecast results. Since the processing of electric signals is not formalized, it is possible that electric signals were interpreted differently during seismically quiet periods than during periods of seismic activity. This could possibly explain the large difference between the numbers of successful predictions in forward and reverse time [*Mulargia and Gasperini*, 1992, 1993].

There is still a possibility that precursory electric signals are registered before strong earthquakes as well as during aftershock sequences, and that the connection is thus real. Statistical tests do not usually give a final answer. However, it is clear from the tests that if such a correlation exists, it is weak, and it will be difficult to establish its statistical significance. *Jackson* [1996] considers in detail the conditions that need to be taken into account if the VAN method is to be tested rigorously in the future.

Acknowledgments. I appreciate partial support from the National Science Foundation through Cooperative Agreement EAR-8920136 and USGS Cooperative Agreement 14-08-0001-A0899 to the Southern California Earthquake Center (SCEC). I am grateful to Dave Jackson (UCLA) for useful discussion and to Dr. F. Mulargia and to Dr. P. Varotsos for sending me the necessary materials. Publication 185, SCEC.

References

Dologlou, E., A 3 year continuous sample of officially documented predictions issued in Greece using the VAN method – 1987-1989, *Tectonophysics, 224*, 189-202, 1993.

Geller, R. J., this issue, 1996.

Hamada, K., Statistical evaluation of the SES predictions issued in Greece – alarm and success rates, *Tectonophysics, 224*, 203-210, 1993.

Jackson, D. D., this issue, 1996.

Molchan, G. M., and O. E. Dmitrieva, Interaction of seismic events for short times and great distances, *Doklady Russian Acad. Sci., 325*, 56-59, (in Russian), 1992.

Molchan, G. M., and Y. Y. Kagan, Earthquake prediction and its optimization, *J. Geophys. Res., 97*, 4823-4838, 1992.

Molchan, G. M., and I. M. Rotwain, Statistical analysis of long-term precursors of large earthquakes, In: Keilis-Borok, V. I., and Levshin, A. L., (eds.), *Computational Seismology, 15*, Nauka, Moscow, 52-66, (in Russian, English translation: Allerton Press, New York, 56-70, 1985) 1983.

Mulargia, F., and P. Gasperini, Evaluating the statistical validity beyond chance of VAN earthquake precursors – Reply, *Geophys. J. Int., 115*, 1199-1201, 1993.

Mulargia, F., and P. Gasperini, Evaluating the statistical validity beyond chance of VAN earthquake precursors, *Geophys. J. Int., 111*, 32-44, 1992.

Reasenberg, P., Second-order moment of Central California seismicity, 1969-1982, *J. Geophys. Res., 90*, 5479-5495, 1985.

Shnirman, M., S. Schreider, and O. Dmitrieva, Statistical evaluation of the VAN predictions issued during the period 1987-1989, *Tectonophysics, 224*, 211-221, 1993.

Takayama, H., Evaluating the statistical validity beyond chance of VAN earthquake precursors – Comment, *Geophys. J. Int., 115*, 1197-1198, 1993.

U.S. Geological Survey, Preliminary determination of epicenters (PDE), Monthly Listings, U.S. Dep. of Inter., Natl. Earthquake Inf. Cent., Denver, Jan. 1993.

Varotsos, P., K. Eftaxias, F. Vallianatos, and M. Lazaridou, this issue, 1996.

Y. Y. Kagan, Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90095-1567. (e-mail: kagan@cyclop.ess.ucla.edu)

(received May 18, 1994; revised September 5, 1994; accepted October 27, 1995.)