Parametric time series analysis of geoelectrical signals: an application to earthquake forecasting in Southern Italy

Gerardo Di Bello ([1]), Vincenzo Lapenna ([2]), Maria Macchiato ([3]), Celestina Satriano ([4]), Carmine Serio ([1]) and Valerio Tramutoli ([1])

([1]) *Dipartimento di Ingegneria e Fisica dell'Ambiente, Università della Basilicata, Potenza, Italy*
([2]) *Istituto di Metodologie Avanzate di Analisi Ambientale, Area della Ricerca del C.N.R., Tito (PZ), Italy*
([3]) *Dipartimento di Scienze Fisiche, Università «Federico II», Napoli, Italy*
([4]) *Dipartimento di Chimica, Università della Basilicata, Potenza, Italy*

Abstract
An autoregressive model was selected to describe geoelectrical time series. An objective technique was subsequently applied to analyze and discriminate values above (below) an *a priori* fixed threshold possibly related to seismic events. A complete check of the model and the main guidelines to estimate the occurrence probability of extreme events are reported. A first application of the proposed technique is discussed through the analysis of the experimental data recorded by an automatic station located in Tito, a small town on the Apennine chain in Southern Italy. This region was hit by the November 1980 Irpinia-Basilicata earthquake and it is one of most active areas of the Mediterranean region. After a preliminary filtering procedure to reduce the influence of external parameters (*i.e.* the meteo-climatic effects), it was demonstrated that the geoelectrical residual time series are well described by means of a second order autoregressive model. Our findings outline a statistical methodology to evaluate the efficiency of electrical seismic precursors.

Key words *time series – geoelectrical signals – earthquake prediction*

1. Introduction

Earthquake precursory phenomena of electrical nature have long attracted the attention of scientists. Strong earthquakes in China (Raleigh *et al.*, 1977) have been forecast using geoelectrical measurements and a network devoted to electrical field measurements has been experimented in Greece (Varotsos and Alexopoulos, 1984a,b; Varotsos *et al.*, 1993). However, up to now, a full comprehensive model to explain the physics of the process has not been available. Since the first recognition of natural electrical signals as earthquake precursors, many authors have pointed out a crucial problem: are the anomalies in the geoelectric measurements random fluctuations intrinsic to the phenomenon or are they effectively related to the underground charge motion induced by tectonic activity? (Burton, 1985). The goal of this work is to build a model to describe the self-potential or geoelectrical time series, offering an objective methodology, based on advanced statistical techniques, to evaluate the occurrence probability of the extreme events in the electrical time series recorded in the seismic areas.

Technically a geoelectric or self-potential time series is a sequence of voltage differences measured with a selected sampling interval using a dipole with two unpolarizable electrodes. In geoelectrical sounding, where a current is injected into the ground, it represents the noise (Lapenna *et al.*, 1994); when we use these measurements in a seismic area it is the signal.

A geoelectric time series can be modeled by the following equation:

$$x(t) = s(t) + z(t) \tag{1.1}$$

where $s(t)$ is a deterministic component related to the external effects (climatological variables, anthropic activities, etc.) and $z(t)$ is a stochastic component related to the charge motion in the ground.

The problem is to localize in the $z(t)$ component values above (below) a fixed threshold to study, in a subsequent step, the possible correlation with seismic activity.

In order to obtain an effective precursor signal we have to take the following steps:

a) the deterministic components must be filtered;

b) a model to describe the $z(t)$ term must be selected;

c) on the basis of the model the occurrence probability of seismic electrical anomalies SES (*i.e.* electrical signals possibly connected with seismic events) must be computed.

A suitable tool to analyze the probability of abnormal values in the geoelectric time series is the crossing theory or theory of runs (Cramer and Leadbetter, 1967). Many applications of this methodology can be found in meteorological research (Macchiato *et al.*, 1993).

The term crossing theory is used for continuous series, whereas for discrete series the term run theory is often used. According to such theories extreme events are treated as rare events, then such events can be modeled by suitable stochastic processes. The crossing theory has been successfully employed in the field of flood frequency analysis (Yevjevich, 1972; Bras and Rodriguez-Iturbe, 1985).

2. Data

In this work we use a multi-year geoelectrical time series recorded by an automatic station located in Southern Italy. In May '91 we installed two arrays with copper electrodes inserted in the ground at 1 m depth, along the N-S and E-W directions, with spacing 100 m and 120 m respectively, in the Area della Ricerca of the National Council of Research located in the Basilicata region, Tito [40°35′N, 15°44′E] (fig. 1). The sampling rate is $\Delta t = 5$ min and the time series $x(n)$ are built up with the daily mean voltage values obtained from a set with 288 measurements (fig. 2). Some data are missing only during the period July '91 - September '91 because we upgraded the station, there are not other periods with a high number of data missing, the global ratio between missing and measured value being 10%. To check eventual polarizing effects we also used unpolarizable electrodes, built of ceramic vessels with a saturated solution of copper sulphate. A constant check between measurements obtained with different probes is carried out. During the working period of the station no anomalous patterns, possibly related to polarizing effect, were detected (Di Bello *et al.*, 1994).

Self-potential measurements can be influenced by many geophysical parameters (*i.e.* seasonal effects of the climatological variables, magnetic storms, etc.), so a preliminary filtering procedure is necessary to remove these effects on the data. In order to remove this external noise a filtering procedure described in a previous paper was applied (Di Bello *et al.*, 1994). The result was an estimate of the $z(t)$ component removing the $s(t)$ term from $x(t)$ geoelectrical time series, in the following section we use the discrete form $z(n\Delta t)$ with $\Delta t = 1$ day. The residual time series $z(n)$ was characterized by zero mean and unit variance.

The station is located in a *natural laboratory* to study the electrical precursory phenomena. In fact we monitored a mountain area with very low anthropic noise, where there are no big cities, railways, etc., so the measurements are not affected by electrical noise. On the other hand, the measuring station is located in one of most seismic areas of the Mediterranean region. On November 23, 1980 an earthquake ($M_s = 6.9$) occurred in the Irpinia-Basilicata Apenninic region. It was the largest and most destructive earthquake to occur in this region for over 100 years (Westaway and Jackson, 1984). The historical seismicity pattern (Pantosti and Valensise, 1990) confirms intense

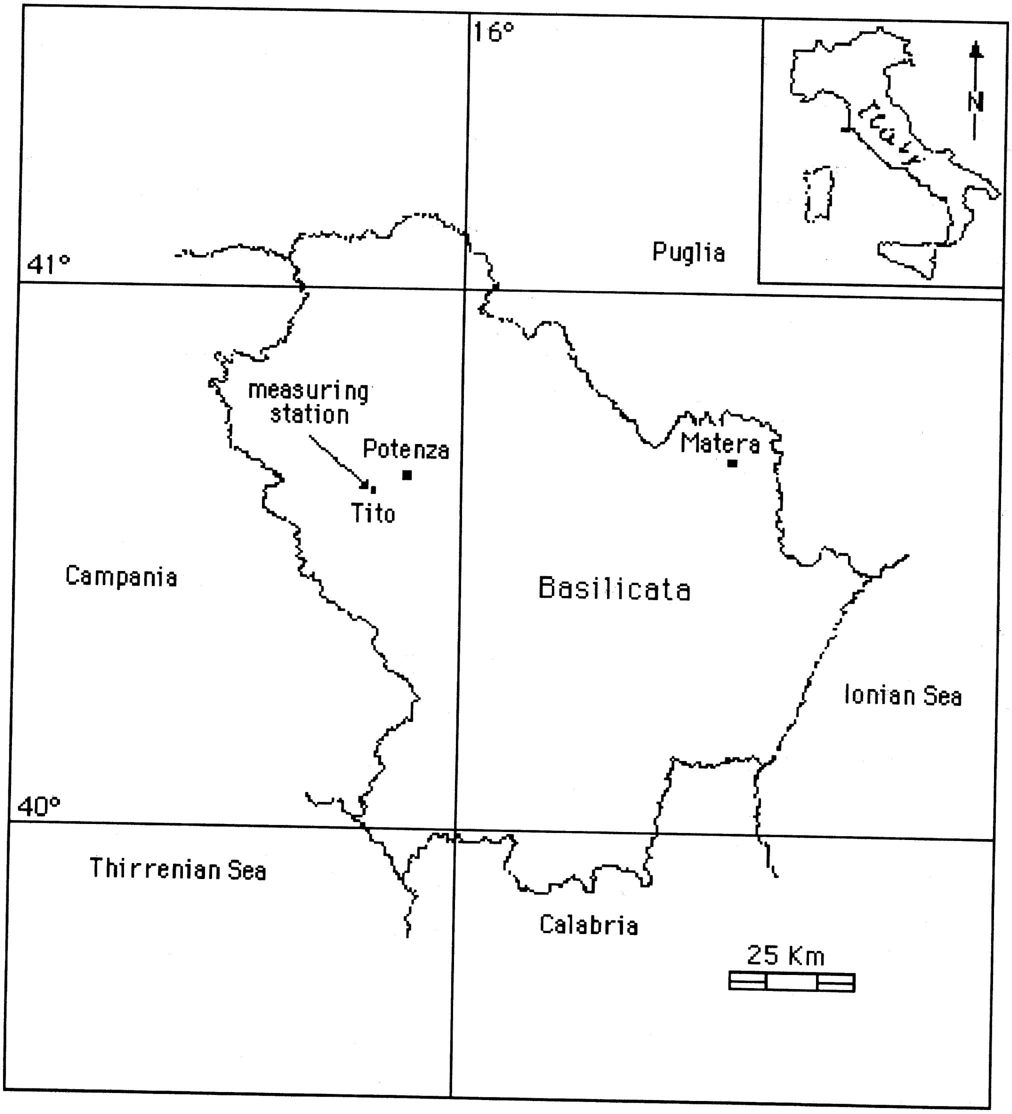

Fig. 1. The map shows the location of the measuring station.

seismic activity and related complexity in crustal faulting. During recent years we have not recorded strong earthquakes in the area under study, but many low magnitude seismic events ($M \leq 4$) occurred. The space pattern of the events which occurred during the last three years is shown in fig. 3.

3. The model

First of all we must demonstrate that the geoelectrical time series can be considered a realization of an autoregressive process.

We briefly summarize the main theoretical aspects of autoregressive processes. A $p-th$

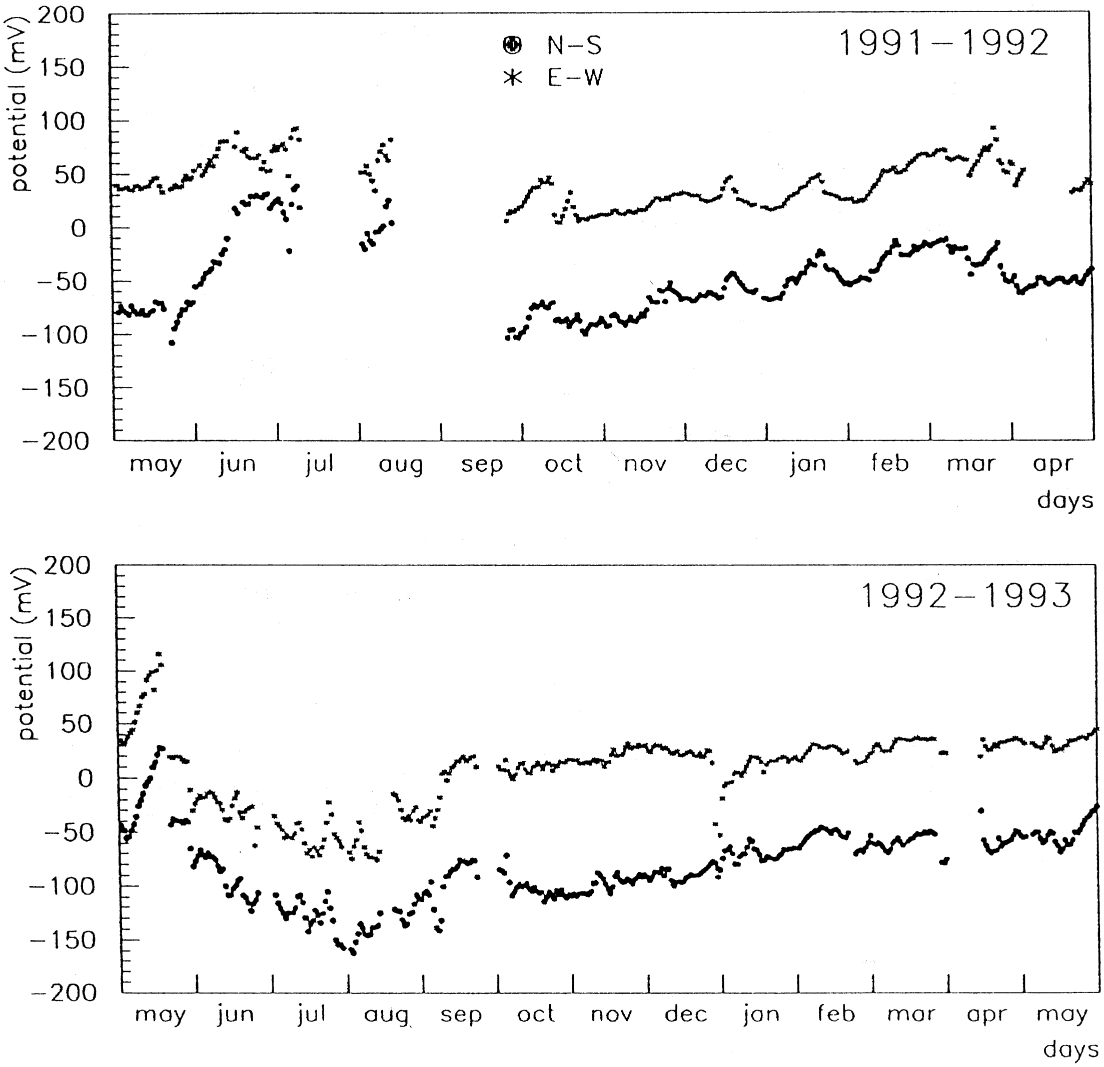

Fig. 2. Plot of the voltage signals measured along the two directions N-S and E-W.

order autoregressive process (Box and Jenkins, 1976), $AR(p)$ can be described as:

$$z(n) = \sum_{j=1}^{p} \phi_j z(n-j) + w(n), \qquad (3.1)$$

where ϕ_1, ..., ϕ_p are the parameters of the model, p is the order of the process, $z(n)$ the time series we must analyze (*i.e.* in our case, the residuals) and $w(n)$ a purely white noise. To fit the experimental data with an autoregressive process is to solve the problem of linear prediction, a forecast value $\hat{z}(n)$ may be considered a linear combination of the previous terms:

$$\hat{z}(n) = \phi_1 z(n-1) + \phi_2 z(n-2) + \ldots\ldots\ldots + \phi_p z(n-p), \qquad (3.2)$$

where the coefficients $\phi_1, ..., \phi_p$ of the linear sum may be evaluated in a way to return an uncorrelated series given by:

$$w(n) = z(n) - \hat{z}(n) = z(n) - [\phi_1 z(n-1) + \\ + \phi_2 z(n-2) + \ldots + \phi_p z(n-p)]. \quad (3.3)$$

Equation (3.3) is equivalent to eq. (3.1), the model parameters may be estimated in a way to give completely independent values $w(n)$.

We fitted the data with an autoregressive process with an upper limit of order p_u. The optimal order p, limited in the interval $(0, p_u)$, was selected with the information criterion AIC (Akaike, 1974). The choice consists in the minimization of the following function:

$$AIC(p) = N \log (\hat{\sigma}_w^2 (p)) + 2p \quad (3.4)$$

where $\hat{\sigma}_w^2 (p)$ is the white noise variance of the autoregressive process that can be calculated

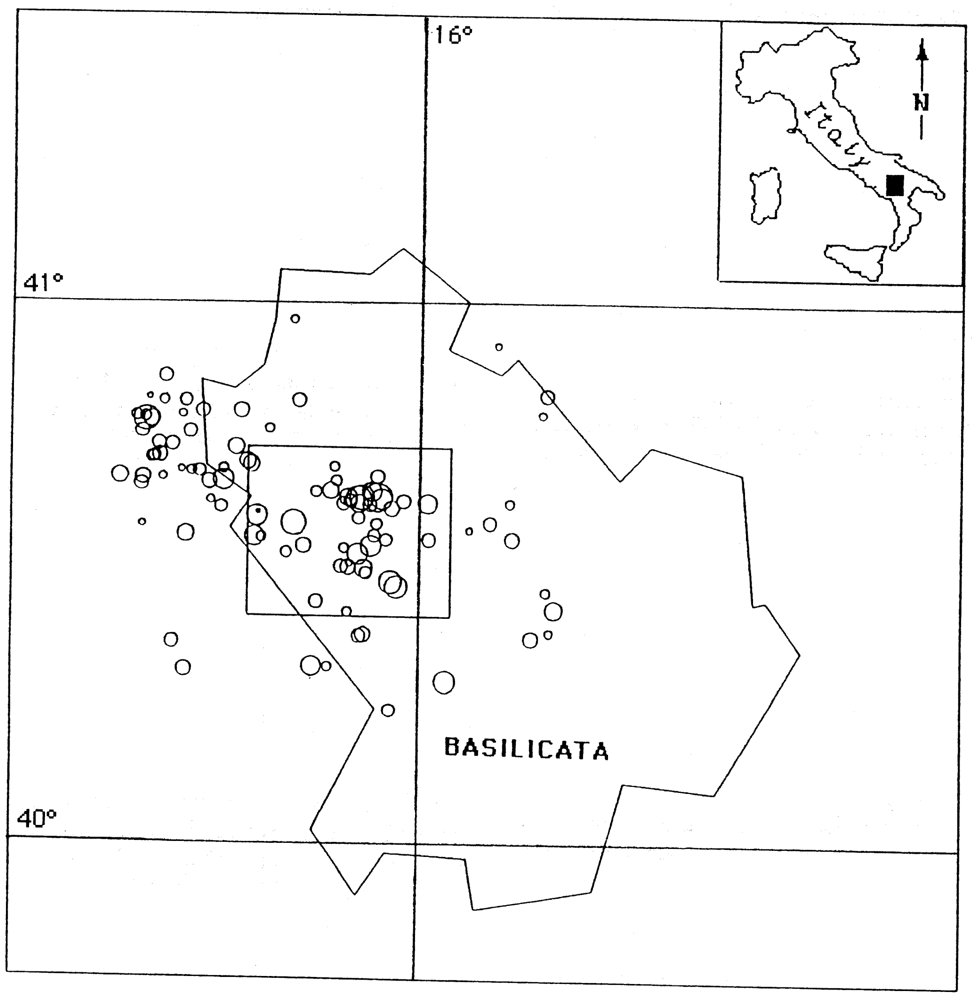

Fig. 3. Seismic events in the Apenninic area during the period May '91-May '93. The magnitude of the earthquakes varies within the range ($2 \leq M \leq 4$).

from eq. (3.2) and N is the number of data available.

The parameters of the process, the coefficients ϕ_1, ..., ϕ_p and $\hat{\sigma}_w^2(p)$, are evaluated using the recursive method of Yule and Walker (Box and Jenkins, 1976).

On the basis of the previous techniques we analyze the residual geoelectric potential obtained from the time series of fig. 2, selecting a second-order autoregressive process. Therefore, both the residual time series (N-S and E-W) can be considered a realization of an AR (2) process:

$$z(i) = \phi_1 z(i-1)+\phi_2 z(i-2)+w(i) \quad (3.5)$$

where ϕ_1 and ϕ_2 are the coefficients of the autoregressive process.

In order to check the $AR(2)$ model the following differences are computed:

$$\hat{w}(i) = z(i) - \hat{\phi}_1 z(i-1) - \hat{\phi}_2 z(i-2) \quad (3.6)$$

where $\hat{\phi}_1$ and $\hat{\phi}_2$ are the coefficients of the autoregressive model fitted to the data. In our case for an AR model (Box and Jenkins, 1976) we have:

$$\hat{\phi}_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \quad (3.7)$$

$$\hat{\phi}_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2} \quad (3.8)$$

where ρ_1 and ρ_2 are the autocorrelation coefficients. The coefficients of the autoregressive model fitted to the data result $\hat{\phi} = 0.947$ and $\hat{\phi} = -0.052$, for the N-S data, and $\hat{\phi} = 0.971$ and $\hat{\phi} = 0.190$, for the E-W data. Our model is an optimal choice when the $\hat{w}(i)$ obtained by eq. (3.6) is a purely white noise. In the frequency domain when a purely white noise is analyzed a completely flat power spectrum is obtained. In this case the values of the cumulative periodogram

$$I(j) = \sum_{i=1}^{j} P(i) \quad (3.9)$$

(where $P(i)$ is the power for each harmonic component), plotted versus the frequency number j, are aligned around a straight line. According to the classical Kolmogorov-Smirnov test (Jenkins and Watts, 1968) we can consider our data a realization of a white noise if the experimental values are inside a band, whose limits are strictly related to the chosen probability level. In the case under study we selected a 95% confidence interval. The results of the test are plotted in fig. 4a,b: all the values fall within the band, the residual obtained from eq. (3.6) is a white noise. The geoelectric time series can be considered a realization of an $AR(2)$ process.

4. Occurrence probability of extreme events

In this section we study the statistics of the extreme events in the geoelectrical time series using the crossing theory. Generally this technique is applied to very long historical records to obtain a good estimation of occurrence probability curves of these abnormal values. In many practical applications this condition is not satisfied.

In our case, the length of the time series is long enough to assess the structure of the geoelectrical time series, but not sufficiently long to have good statistics about abnormal events. When a time series model is identified, it is possible to simulate a time series with the same statistical properties as experimental data. In this way an estimate of the occurrence probability of extreme events can easily be obtained from the simulated data. In section 4.1 we briefly summarize the theoretical aspects of the crossing theory.

4.1. *Theoretical background*

A sequence of successive values above (below) a threshold is defined *run*, different statistics are associated with runs in many geophysi-

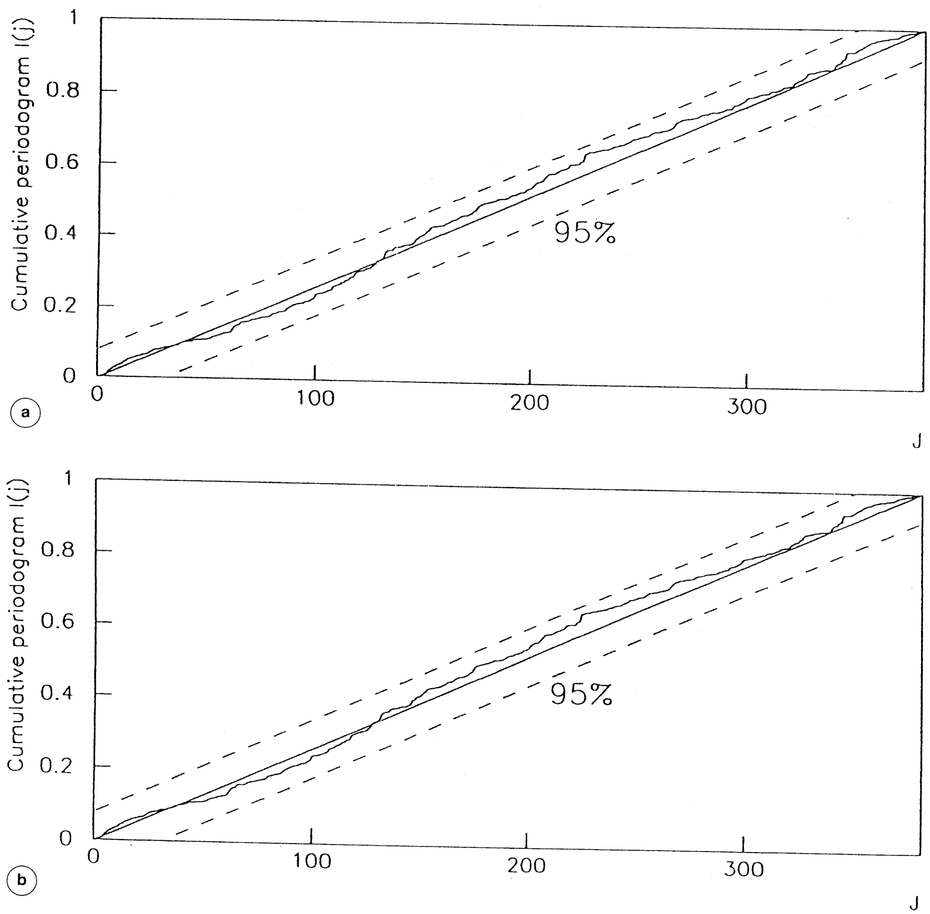

Fig. 4a,b. Kolmogorov-Smirnov test on $\hat{w}(i)$ data. The figures a) and b) refer to the results for the N-S and E-W array directions. The dashed lines delimit a band within which the values computed from a white noise process should be included with a probability of 95%.

cal fields. In this work we deal with the statistic related to the length, representing the period in days, m, that the variable under study is above (below) the selected truncation level z_0 (fig. 5). The probability distribution $P(m \geq j; z_0)$ which for an arbitrary level z_0 gives the probability that m is greater than a j-day period. The function $P(m \geq j; z_0)$ is normalized so that $P(m \geq 1; z_0) = 1$. In this paper we use a variable with unit variance, from which the deterministic component $s(i)$ has been removed and consequently the truncation level is not related with the time.

Having a model that describes the empirical time series, the run analysis proceeds straightforwardly. Analytical relations of distributions of run length of the AR model are available, but they are difficult to handle in practical computation. It is easier to estimate $P(m \geq j; z_0)$ by simulating a very large residual time series from the selected AR process.

The above mentioned simulation approach can produce artificial time series that reflect any desidered run length compatible with the set of observations. Finally, the $P(m \geq j; z_0)$ value is estimated as the sample relative fre-

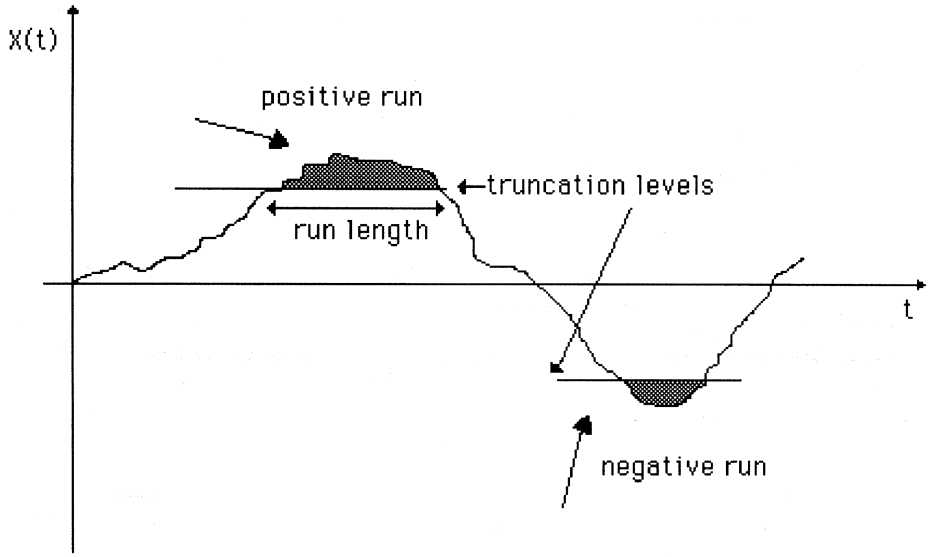

Fig. 5. An example of extreme events or runs picked out in a time series.

quency of run lengths that are greater than j-day periods:

$$P(m \geq j;\, z_0) = \frac{\left(\sum_{i=j}^{\infty} F_i\right)}{\left(\sum_{i=1}^{\infty} F_i\right)}, \qquad (4.1)$$

where F_i indicates the number of runs below (above) the threshold z_o that are i days long. In the previous equation a truncation point must take the place of the theoretical infinity limit appearing in the summations. An example regarding the occurrence probability curves, $P(m \geq j;\, z_0)$, for an $AR(2)$ process is reported in fig. 6. The curves are obtained computing the runs from a simulated time series using eq. (4.1).

4.2. *Estimate of SES occurrence probability*

A suitable application of the previous methodology to the time series analysis devoted to earthquake forecasting is the evaluation of the occurrence probability of a sequence of abnormal experimental values in the geoelectrical data. We use the following procedure:

i) an AR model to describe the time series is selected;

ii) a simulation is performed in order to obtain a very large amount of data;

iii) the theoretical occurrence probability curves are computed according to eq. (4.1);

iv) for each run length we selected on the experimental data, an occurrence probability (*i.e.* easily obtained using the theoretical curves) is associated.

In such a way an objective methodology to discriminate the SES anomalies from the random fluctuations is determined: we consider as SES anomalies only the runs with a very low occurrence probability (*i.e.* rare events).

Finally we give an example using the geoelectrical time series recorded by our automatic station in Tito (Southern Italy) (fig. 7). In the graphs of fig. 7 the runs above (below) 2σ truncation level are picked out and the earthquakes in a circular area, with a radius of 20 km centered in Tito, are indicated by arrows. Obviously the choice of the radius is empirical, but it is necessary to mark the boundary of the area in which variations in geophysical field induced by the seismic events are detectable (Dobrovol'skiy, 1993).

During May '91 and May '92 many abnormal values were picked out from residual geoelectrical time series, the length of the selected runs (*i.e.* the number of consecutive values above (below) the fixed threshold) is so large to give a very low occurrence probability for each of them. Using the selected $AR(2)$ model and eq. (4.1) it results lower than 10%. It is certain that the anomalies are not random fluctuations of the stochastic process under study, but they are related to deep underground charge motion.

In the same period, many seismic events occurred in the area, therefore it seems there is a possible correlation between anomalous electrical signals and seismic activity during these two particular periods. On the other hand during early April there was an earthquake and we did not observe runs.

Obviously, the above possible correlation is only an example, our goal being to build an objective methodology to select extreme events

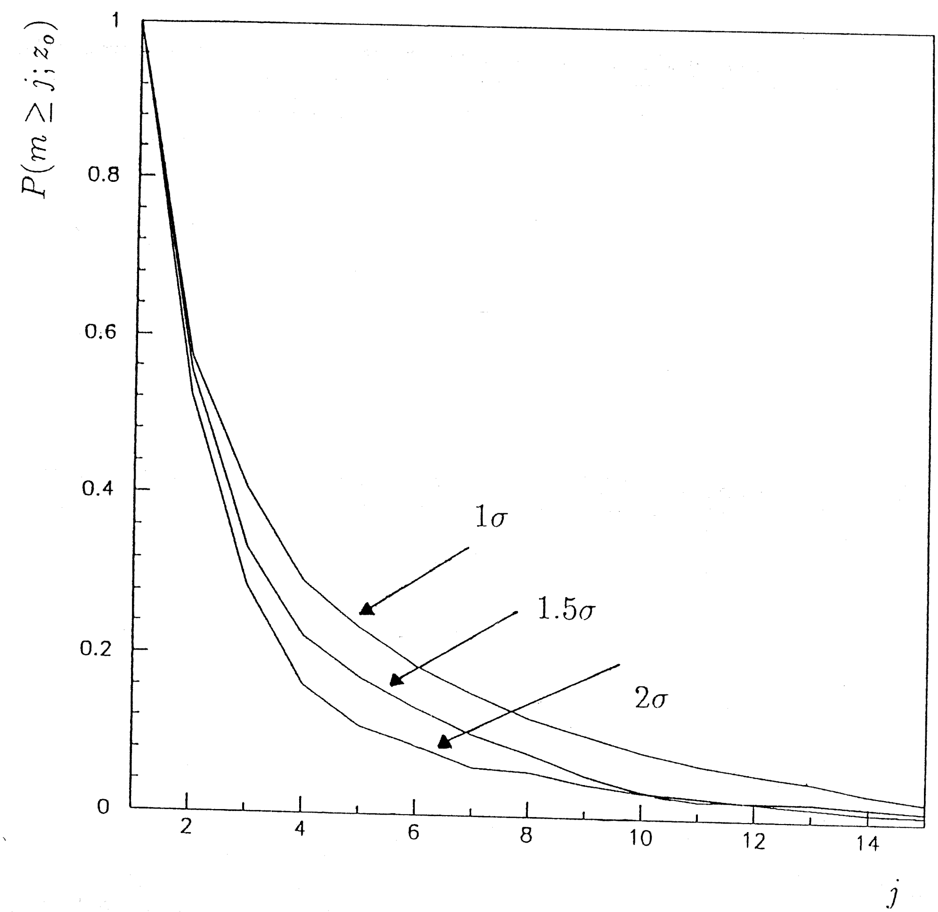

Fig. 6. The probability occurrence curves for an autoregressive process; the different lines refer to the selected thresholds.

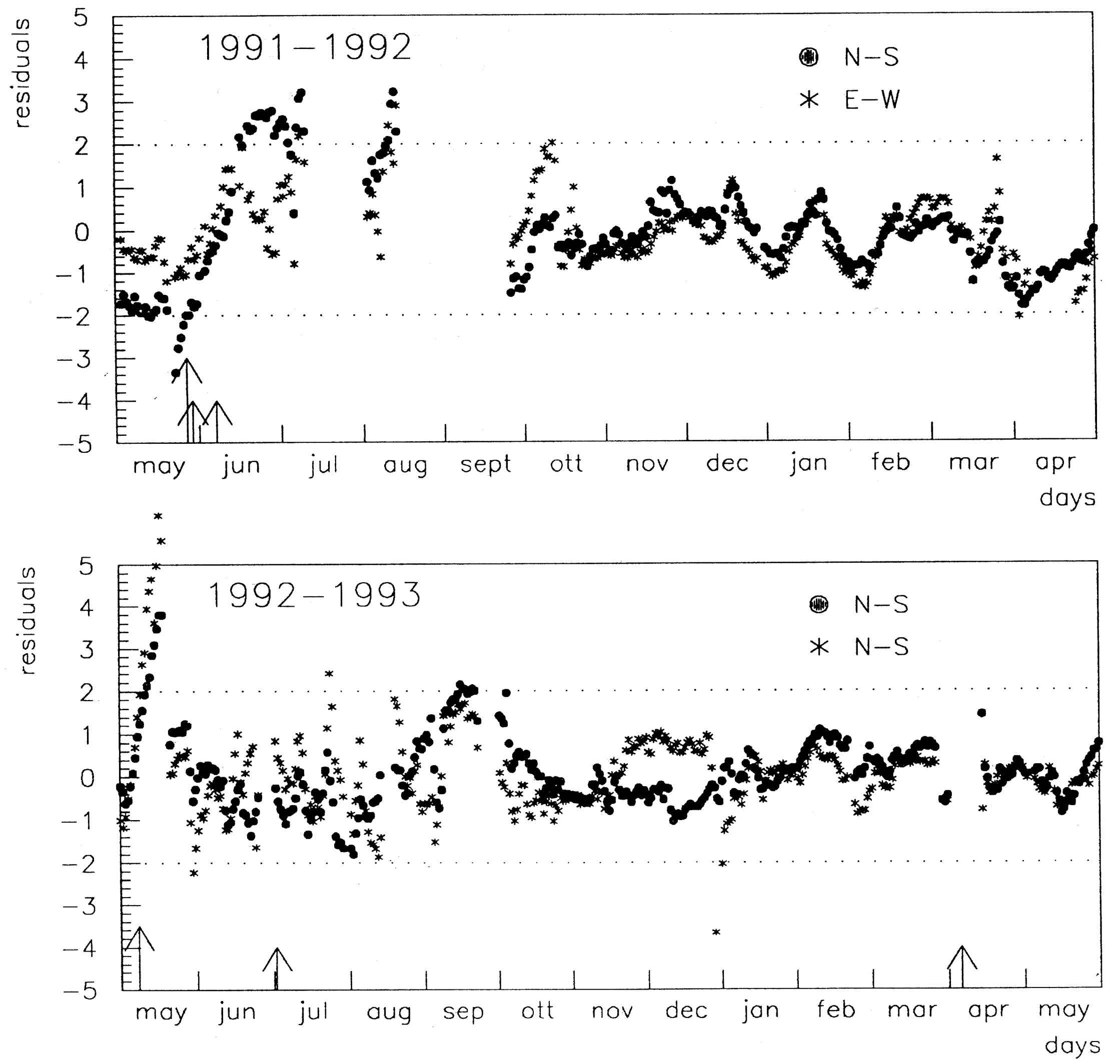

Fig. 7. Residual potential time series during the period May '91-May '93. The arrows refer to the earthquakes in a circular area (r = 20 km) surrounding the station.

in the geoelectrical time series. Actually we do not have enough run events and seismic events to study the efficiency of the SES precursory technique in this area from a statistical point of view. In the near future, on the basis of the methodology discussed above, we will be able to evaluate the efficiency of electrical precursors.

5. Conclusions

In this work we demonstrate that, after removing the periodic components related to meteo-climatic parameters, the self-potential time series can be considered a realization of autoregressive processes. A complete statistical check of the proposed model is performed using the time series recorded by an automatic station located in Southern Italy. On the basis of this model and using the crossing theory an objective methodology to evaluate the occurrence probability of extreme events in the geoelectrical time series is discussed. Finally, the selected extreme events and the seismic events which occurred in the area are correlated.

REFERENCES

AKAIKE, H. (1974): A new look at the statistical model identification, *IEEE Trans. Autom. Control*, AC-19, 716-722.

BOX, G.E.P. and G.M. JENKINS (1976): *Time Series Analysis Forecasting and Control* (Holden Day, Oakland, California), pp. 350.

BRAS, R. and I. RODRIGUEZ-ITURBE (1985): *Random Fluctuations and Hydrology* (Addison-Wesley), pp. 599.

BURTON, P.W. (1985): Electrical earthquake prediction, *Nature*, **315**, 370-371.

CRAMER, H. and M.R. LEADBETTER (1967): *Stationary and Related Stochastic Processes* (Wiley and Sons), pp. 348.

DI BELLO, G., V. LAPENNA, C. SATRIANO and V. TRAMUTOLI (1994): Self-Potential time series analysis in a seismic area of Southern Italy: first results, *Annali di Geofisica*, **37** (5), 1137-1148.

DOBROVOL'SKIY, I.P. (1993): Analysis of preparation of a strong tectonic earthquake, *Phys. Solid Earth*, **28** (6), 481-493.

JENKINS, G.M. and D.G. WATTS (1968): *Spectral Analysis and Its Applications* (Holden-Day, San Francisco), pp. 525.

LAPENNA, V., M. MACCHIATO, C. SATRIANO, C. SERIO and V. TRAMUTOLI (1994): Statistical analysis of non-stationary voltage recordings in geoelectrical prospecting, *Geophys. Prospect.*, **42** (8), 911-954.

MACCHIATO, M., C. SERIO, V. LAPENNA and L. LAROTONDA (1993): Parametric time series analysis of cold and hot spells in daily temperature: an application in Southern Italy, *J. Appl. Meteorol.*, **32** (7), 1270-1281.

PANTOSTI, D. and G. VALENSISE (1990): Faulting mechanism and complexity of the November 23 1980, Campania-Lucania earthquake, inferred from surface observations, *J. Geophys. Res.*, **93** (15), 319-341.

RALEIGH, B., G. BENNET, H. CRAIG, T. HANKS, P. MOLNAR, A. NUR, J. SAVAGE, C. SCHOLZ, R. TURNER and F. WU (1977): Prediction of the Haicheng earthquake, *EOS (Trans. Am. Geophys. Union)*, **58**, 236-272.

VAROTSOS, P. and K. ALEXOPOULOS (1984a): Physical properties of the variations of the electric field of the Earth preceding earthquakes, *Tectonophysics*, **110**, 73-98.

VAROTSOS, P. and K. ALEXOPOULOS (1984b): Physical properties of the variations of the electric field of the Earth preceding earthquakes 2: determination of epicentre and magnitude, *Tectonophysics*, **110**, 99-125.

VAROTSOS, P., K. ALEXOPOULOS and M. LAZARIDOU (1993): Latest aspects of earthquake prediction in Greece based on seismic electric signals. *Tectonophysics*, **224**, 1-39.

WESTAWAY, R. and J. JACKSON (1984): Surface faulting in the Southern Italian Campania-Basilicata earthquake of November 1980, *Nature*, **312**, 436-438.

YEVJEVICH, V. (1972): *Stochastic Processes in Hydrology* (Water Resources Publication), pp. 276.