ANNALI DI GEOFISICA, VOL. 41, N. 5-6, November-December 1998

Natural catastrophes and point-like processes
Data handling and prevision

Giovanni P. Gregori
Istituto di Fisica dell'Atmosfera (IFA-CNR), Roma, Italy and World Laboratory, Lausanne, Switzerland

Abstract
A frequent approach when attempting to manage a natural catastrophe is in terms of a numerical model, by which we try to forecast its occurrence in space and time. But, sometimes this is difficult or even unrealistic. On more pragmatic grounds we can appeal to a formal analysis of the historical time series of every catastrophe of concern. Only approximately, however, can such series be likened to a point-like process, because the «detector-mankind» experienced substantial changes *versus* time. Nevertheless, such algorithms can be approximately applied by means of a few suitable assumptions. In the ultimate analysis, four basic viewpoints can be considered: i) either by assuming that phenomena are periodic; ii) or by assuming that an event occurs only whenever some energy threshold is attained (calorimetric criterion); iii) or by assuming that it occurs only whenever the system experiences some abrupt change in its boundary conditions; or iv), whenever no such algorithm is viable due to scanty observational information, just by applying fractal analysis, in terms of the box counting method, or some other more or less related and/or equivalent algorithms. The mutual relations, advantages, and drawbacks of any such approach are briefly discussed, with a few applications. They already lead to an apparently successful long-range forecast of a large flood in Northern Italy which occurred in 1994, and to the prevision of the next explosive eruption of Vesuvius. But the success of every application is closely determined by the quality of the historical database, or by the physical information that is fed into the analysis, rather than by mathematics that *per se* have only to be concerned with avoiding some arbitrary input being added, based only on the human need for simplicity. The present paper gives a synthesis of several algorithms that were previously independently applied on a simple intuitive basis to different case studies, although with no comparisons or discussion of their similarities and/or differences.

Key words *natural catastrophes – point-like processes – prevision – periodicity – cyclicity – energy balance – fractals – box-counting method – floods – climate anomalies – solar control – volcanic cycles – volcanic supply*

1. Introduction

The cognitive process involves three stages: i) morphological or descriptive; ii) inductive; and iii) deductive. The first is concerned with the collection of observations, the second one with data handling that is aimed at inferring all evidence that is manifestly intrinsic within the database. In such a respect, paradoxically one can even imagine programming a computer to perform «all» kinds of statistical analyses that a researcher is supposed to carry out. Such a stage, however, has an intrinsic limitation within the available statistics, that are normally insufficient for exploiting all which in principle ought to be carried out. At such a point, the intuition by the researcher enters into play, by envisaging possible solutions, mechanisms, models, etc. by exploiting their eventual computation, and finally by checking them by some

Mailing address: Prof. Giovanni P. Gregori, Istituto di Fisica dell'Atmosfera (IFA-CNR), Via Fosso del Cavaliere 100, 00133 Roma, Italy; e-mail: gregori@atmos.ifa.rm.cnr.it

trial-end-error procedure: this is the deductive stage. Induction (according to Aristotle and Galileo) means inferring the premises from the final evidence, while deduction is from axioms to final conclusions.

The deductive stage can be further divided into two sub-classes, that are here briefly called deterministic and pragmatic viewpoints, respectively. The first recalls the Lagrangian determinism, by which every particle has its own degrees of freedom (d.o.f.), *e.g.* supposedly expressed in terms of canonical coordinates. Wherever we know the exact physical laws, and we are also capable of writing and solving all equations, we can, in principle, know everything about the past, present, and future of our physical system. Since this is a utopia, one normally appeals to statistics, thermodynamics, etc. In contrast, the pragmatic viewpoint wants to limit the amount of inference only to the realm of what can be objectively obtained from the limited amount of available observations.

On a practical ground, the deterministic viewpoint is implicitly used whenever one formulates a numerical model. In fact, one assumes to know some fundamental laws, and defines a few suitable parameters that ought to give a description of a few basic d.o.f.'s of the system. Some equations are written and solved by feeding into them some boundary conditions. The solution gives a detail that is expressed in terms of much wealthier information, compared to the scanty observations that composed the boundary conditions. Differently stated, the researcher overcomes the drawback of limited observational information, by using deduction from a few known fundamental laws. The correctness of his entire approach must, however, be tested by comparing his inferences with additional available observations, the form and content of which were not formally fitted into the model. In general, mathematical difficulties imply some limits. Sometimes, for computational purposes, one can only deal with linear, or linearised, processes, etc.

The pragmatic approach is more akin to induction. Provided that we exploited all which induction alone could provide, we envisage some working hypotheses, some chains of causes-and-effects, concerned with either linear or non-linear mechanisms, by which, although we exploit no detailed modelling, we can carry out some checks by means of the available observations.

In principle, every such stage and/or viewpoint is correct *per se*, provided that it is suitably used. The ancient Greek thinkers were fond of deduction alone, while in the XIV century A.D. the Franciscan school of theology was much more pragmatic (they were the real forerunners of Galileo ~3 centuries in advance, but their logical struggle was defeated for political reasons; Gregori and Gregori, 1998). Leonardo da Vinci (1452-1519) and finally Galileo (1564-1642) succeeded in giving the correct place to induction. Nowadays, a somewhat analogous competition survives between the deterministic and the pragmatic viewpoint.

In this respect, some controversy is sometimes raised concerning an improper mixing of different logical steps (*e.g.* when dealing with the global climate change). In fact, one should distinguish between: i) the collection of the observational database, ii) the knowledge of the basic mechanisms (physical, chemical, biological); and iii) the exploitation of a numerical model to forecast some future scenario. Implementing the third step, with no adequate previous assessment of the knowledge dealing with the first two steps, definitely appears over-ambitious. Moreover, the first two steps can be tackled either from the pragmatic or from the deterministic viewpoint, while the third step can only be deterministic. Several present debates and controversies could thus be promptly solved, whenever one can clearly recognize what is objectively contained within observations, and what is the result of intuition or speculation by every researcher.

The present paper is specifically focused on the logical problems related to the prevision of natural catastrophes. Such a target can be challenged in terms of determinism, that implies working out a suitable model, that by means of a few observed precursors ought to be capable of issuing a reliable forecast. Pragmatists, rather, prefer to emphasise the limits of our knowledge, and to consider the observational evidence *per se*, that is provided by the historical time series of several previous events, that can be more or less related to the catastrophe we are concerned with. Such a pragmatic viewpoint will be here adopted.

A basic warning deals with the definition both of the kind of «event» that is of concern, and of what is meant by «prevision». The definition of natural catastrophe implies that it must cause some relevant damage to mankind and/or causalities. But a catastrophe is the result of two concurrent causes: i) of an unusual natural occurrence; and, mostly, ii) of the fact that a suitable number of people are located in some critical site in order to «detect» the event, in such a way that damage and causalities are produced. In fact, if such a detector is not correctly located, we cannot monitor the event. Differently stated, the anthropic factor (such as the use of territory by mankind) is crucial for the definition of catastrophe: the impact of a flood, of a land- or snow-slide, of a volcanic eruption, of a forest fire, etc. is therefore much more a function of Man than of Mother Nature.

Concerning the definition of prevision, its error-bars ought to be specified both in time and space. For example, a seismic map is a prevision with a good space definition, but with no time specification. Too loose a resolution either in space or time, or too short an advance-time of a precursor, makes a forecast of little practical use. The needs of mankind, and the intrinsic pace and scale of human actions, have exigencies that sometimes hardly fit with the typical pace and scale of natural phenomena. Moreover, a simple-minded (and surprisingly frequent) approach is in searching for some universal precursor that ought to apply, almost «magically», within every environment in all parts of the world (forgetting about the relevant, and still poorly known, inhomogeneities of the Earth).

Both from the pragmatic and from the deterministic viewpoint, a prevision shall never be expected to provide with a detail that is better than the intrinsic information provided by its observational database. Science is the search for truth, but we never reach it: we can just approach it in terms of an iterative process. Therefore, no prevision shall ever be error-free. This is just a matter of common sense, although the legislative system of almost every country seems to be completely unprepared to face such a logical and unavoidable reality, in the fact that decision makers erroneously presume that science is always capable of providing a fully reliable error-free forecast.

Therefore, when defining a logical approach to the search for a prevision, we must first define its basic, and more or less conscious or unconscious, implicit assumptions.

2. Abstraction and rationale

The first abstraction relies on considering a catastrophe a point-like process, *i.e.* a «yes-or-not» event. For example, we have to decide when we state that a flood occurred, or a spate, or an earthquake (what minimum magnitude), or a new volcanic eruption (rather than being a simple continuation of a previous one), or a land- or snow-slide (what is the minimum volume involved and what is the area of interest), etc. The theory of point-like processes is discussed by Brillinger (1975; additional references in Pavese and Gregori, 1985). However, in a strict sense, a catastrophe is not a point-like process according to its formal mathematical definition, in the fact that a catastrophe needs for a recorder (mankind) that has a sensitivity and a precision varying in time, depending on the demographic increase, or on the change in technology and energy needs, or on the territory occupation, that in general depends on the varying relation between Man and territory. Moreover, even psychological factors play an important role, *e.g.*, even in reporting about a volcanic eruption (Gregori *et al.*, 1992), or dryness/wetness in China (Banzon *et al.*, 1992a). That is, a time series of catastrophes is severely biased by the recorder, *i.e.* by mankind, neither is it complete. In contrast, several pertinent examples of point-like process, fitting with its formal mathematical definition, can be easily found *e.g.*, in biology (such as heartbeats, or impulses from a nervous cell, etc.). Therefore, the mathematical theory of point-like processes can be rigorously applied in biology, while in the case of natural catastrophes it can be used only with some warning. Let us distinguish between four different possibilities. Each of them can be applied, provided that the historical information is suited for it.

1) The first viewpoint relies on the assumption that the phenomenon is strictly periodical (Section 3), although in a strict sense nothing is periodical. In fact, even the length of the day (l.o.d.) has some minor fluctuations, although everything that is controlled by the motion of the Earth and of the Moon is approximately periodical, due to the great inertia of their two bodies. It is more correct to state that several phenomena are cyclic, in the fact that there is some regular alternation of relative maxima and minima, although with varying cycle duration (the best known such examples are the solar cycle and the Quasi-Biannual Oscillation, QBO). Therefore, whenever the duration of every such cycle changes by a small percent, one can treat the phenomenon as being approximately periodical, and thus estimate its average period during the time interval spanned by observations. Eclipses are the best known example of natural «catastrophes» that display a closely periodical trend.

2) Another viewpoint is related to the energy balance. We assume that a catastrophe occurs only whenever some suitable, although unknown, energy reservoir, which is located somewhere within our physical system, is filled up beyond some threshold. Such a rationale will be here called calorimetric criterion (Section 4). One specific such application is the Imbò algorithm (Section 5). Such an approach can be heuristically effective, *e.g.*, for comparing different series of point-like processes. In fact, let us suppose that such a rationale is viable for every one of them. Then, either in the case that we can envisage some mechanisms relating them to each other, or in the case that we cannot, we can apply such an algorithm: if the prime mover of all such phenomena, whatever it be, is a common energy source, the pace of their occurrence must display some common feature. Note, however, that the occurrence of the events can even appear uncorrelated with each other, depending on the different amount of energy that has to be stored within different energy reservoirs that eventually control the occurrence of different types of such point-like processes. Nevertheless, the energy balance shows a physical relation that should be evidenced. Moreover, different series of point-like processes can eventually contain different data gaps or vacancies. In such a case, provided that we use a suitably robust algorithm, we can still infer significant physical conclusions.

One specific example of this same rationale deals with the well known interpretation of aftershocks in seismology, by which their distribution *versus* both time and intensity should display a specific exponential decay, according to an intuitive model based on the assumption of some random release of energy by the unknown underground fractures. Such a rationale also has implications on the fractal analysis of their respective time series. This same rationale cannot apply to every other natural «catastrophe» (*e.g.*, it does not fit geomagnetic storms). Therefore, the fact that this same rationale can be successfully applied to the time series of El Niño events appears physically intriguing (A. Palumbo, private communication, 1998; in preparation).

Another related algorithm is concerned with a constant average power (*i.e.,* energy released per unit time by a time series of catastrophes) that apparently holds for several different types of «catastrophes» including large volcanic eruptions of either Vesuvius or Etna, extreme marine floodings in Venice, and storm-surges and floodings in Trieste (Palumbo, 1997, 1997a,b, 1998; Palumbo and Palumbo, 1998).

Within all such applications, the energy balance plays a crucial role, *i.e.* although we cannot know the prime physical process and mechanism, we can afford to perform some energy balance, encompassing the final output in some quantitative frame.

3) However, in general such a calorimetric criterion is not always viable. In fact, in some circumstances an event occurs whenever the physical system experiences an abrupt change of boundary conditions, by which it must re-distribute its internal energy according to some specific requirements by a variational principle. In such a case, one can even state that the energy reservoir of the calorimetric view-

point was constituted by a different form of storage of the internal energy by the system; but, this may appear *ad hoc*. It appears more akin to natural reality to describe the event in terms of a basic difference of the composition of the system. One such example is given by the magnetosphere of the Earth (Gregori, 1998). In fact, the usual way of depicting it in terms of MHD implies assuming that there is a perfectly steady flow of solar wind, by which, if the magnetosphere is supposed to be contained within an ideal huge box, for every electron or proton that leaves the box, there must be another electron or proton entering it, etc. But, whenever there is a plasma cavity within the solar wind, the physical system, that is composed of the particles contained within such a box, is abruptly changed. Very rapid (at light speed) transformations of energy must occur, which also imply by the Hamilton principle a transformation from magnetic into particle kinetic energy. The result is that a magnetospheric substorm is triggered, and several such substorms compose a geomagnetic storm, that is almost like a damped oscillation composed of several substorms of decreasing amplitude. Another example is concerned with the encounters of the solar system with interstellar matter (either galactic clouds, or the Oort reservoir, or the blast wave from a nova or supernova). All events associated with it can be more realistically depicted in terms of a change in the physical system, rather than an energy that was previously stored within such interstellar matter prior to the encounter. Geomagnetic excursions and/or reversals could be associated with one such encounter, in agreement with some authors who claim a possible external origin for such «catastrophes» (Gregori, 1993a, 1994 and references therein), although it is well known that the best agreed theories explain such phenomena in terms of mere endogenous effects (*e.g.*, Jacobs, 1987 and references therein). In the ultimate analysis, the most common viewpoint assumes that the interior of the Earth is basically unaffected by external factors, and one such «catastrophe» ought thus to be explained in terms of some self-amplification, or resonance, by which the system eventually experiences one large instability and the field reverses. Differently stated, the «catastrophe» is identified with a mathematical instability. A somewhat analogous situation deals with El Niño: it is an anomalous seasonal oscillation, clearly detected in the Pacific and Indian Oceans area. The best known approach is by assuming that the general external inputs are basically the same every year, although some suitable perturbation eventually triggers an instability, by which through some kind of a resonance the seasonal oscillation is amplified (Philander, 1990; Diaz and Markgraf, 1992; and references therein). That is, this is a mathematical definition as above, when dealing with geomagnetic excursions and reversals. An alternative possibility relies in searching for some presently unmeasured quantity, by which the physical input to the system experiences a substantial interannual variation that is presently unknown to us. It is premature to state what the correct viewpoint is. As far as the present discussion is concerned, such examples ought to clarify how a different axiomatic assumption, since the very beginning of our investigation, can sometimes result in some kind of unconscious bias. Both viewpoints ought to be exploited, until one reaches the possibility of obtaining an unambiguous experimental test.

Summarizing, such a third rationale can be applied whenever we hypothesize some speculated chain of causes-and-effects, that needs to be suitably tested by the available observational information. In such a respect, no general rule can be given, and every case-study needs its own discussion.

4) The fourth rationale applies whenever a point-like process apparently does not even allow for the application of the calorimetric criterion, or for envisaging some adequate speculative chain of cause-and-effect: this is the box counting method or some more complex algorithm borrowed from fractal analysis and that is more or less related to it (Section 6). By this we can infer some intrinsic objective information ultimately related to the pace of events, even in the case that we can work out no other analysis. For example, a well known approach, when dealing with different time series of point-like processes, is by searching for their eventual

bi-variate or multi-variate, linear correlation, or non-linear regression, etc. But, this requires that the wealth of information dealing with the different data series (such as *e.g.* the amplitude of the error-bars, and/or the frequency and distribution of the data gaps) must be reasonably comparable with each other; otherwise, such algorithms cannot work. In contrast, even when such a condition is not satisfied, fractal methods can be effectively applied.

Let us just consider a simple example. A time-varying water source equally supplies two boxes. They have a different size, and each of them, whenever it is filled up, turns upside down and it releases all the water contained inside it. After that, it starts anew filling up as before. The catastrophe is defined as the event by which either one of the two boxes reverses upside down. Suppose that we can monitor separately both of the two time series of such catastrophes and that we have no other information on the system. Although both boxes are supplied by the same time-varying water source, the two point-like processes cannot be correlated with each other. However, there must be some rationale envisaging some common feature of their respective time series, being suggestive of a common prime energy supply. In such a case, in general no periodicity or cyclicity ought to be observed (unless the prime source displays such a feature). The calorimetric criterion, instead, can be applied, provided that each series is not biased by too many gaps. The fractal properties of either series must, however, display a common feature, and the box counting method appears to be the simplest such algorithm.

3. Assuming periodicity – ARP

The best suited algorithm seems to be the superposed epoch criterion, *i.e.* start the stopwatch every time an event occurs, and check that the events are going to be repeated after some fixed time, etc. This is the principle idea of the operator ARP, that works quite effectively provided that the database is not too heavily biased by gaps and/or inhomogeneities. Such a condition can be checked only *a posteriori*, as it results from an intermingling of the intrinsic periodicities of the system, their relative role, the error-bars, the total time span of the observational series, the temporal distribution of its gaps, etc. Such items were extensively discussed elsewhere and will not be dealt with (Banzon *et al.*, 1990, 1992; Gregori 1990, 1997; Gregori *et al.*, 1988, 1994; Pavese and Gregori, 1985; and Pavese *et al.*, 1992). Let us just mention that the superposed epoch criterion implies that, if a phenomenon has an actual intrinsic periodicity of period T_0, it also shows a periodicity of period $T = mT_0$ (for every integer $m > 0$). Moreover, in general one should expect and find also its higher harmonics $T = (m/n)T_0$ (for every integer $m,n > 0$). One thus observes a «period family», or a set $\{(m/n)T_0\}$, having (Gregori, 1990) a «forefather period» T_0. Such families can be easily recognized (Fairbridge, 1996) by plotting log T *versus* log n. This same algorithm can also be effectively applied (Gregori, 1990) to every well-behaved $f(t)$ leading to the operator CETRA, although one must apply a suitable correction to the definition of the supposed time-instant of occurrence of every given event in order to avoid an unwanted bias (*e.g.,* if we know only the year, not the date, of occurrence of a catastrophe, we must also define some hypothetical day for that event by means of a random date; Banzon *et al.*, 1992). The definition of «well-behaved» function denotes an $f(t)$ that is known, monitored, and sampled, at regular time intervals, and that appears almost like a continuous function of time.

ARP was actually applied to a few time series to: i) the climatic anomalies (XII through XVII century A.D.) in the Tanaro valley, recorded in Alessandria (Po valley, North-Western Italy); ii) the floods and large spates of its affluent, the torrent Belbo (XV century A.D. through present); iii) the large spates and floods in the Po delta (A.D. 1700 through present); and (iv) the Tiber floods in Rome (V century B.C.-XIX century A.D.). CETRA was then applied to the output of ARP (for the case (i) here above). This leads to the correct forecast of the great flood that hit the Tanaro valley in 1994. For details refer to the afore-mentioned references for ARP.

4. The calorimetric criterion

Let us focus on the specific problem of the low-pass filtering of an uneven and/or incomplete time-series, and for comparison purposes let us also refer to the case of a well-behaved $f(t)$, *e.g.*, of the planetary average *versus t* of the variation of the climate temperature. Apply to $f(t)$ a running average by means of some window of 1 year total duration, and obtain the well known yearly trend (fig. 1): this is a mathematical low-pass filter, applied to a well-behaved $f(t)$, that rejects the seasonal variation and keeps only the interannual global change. In contrast, consider the case of a pressure-cooker, with a security valve characterized by a small, given weight. Let the heat-source underneath the pressure-cooker be changed by some unknown amount. Monitor the time interval Δt in between any two subsequent whistlers and obtain fig. 2. The time variation of the average prime heat-supply to the pressure-cooker is proportional to $1/\Delta t$.

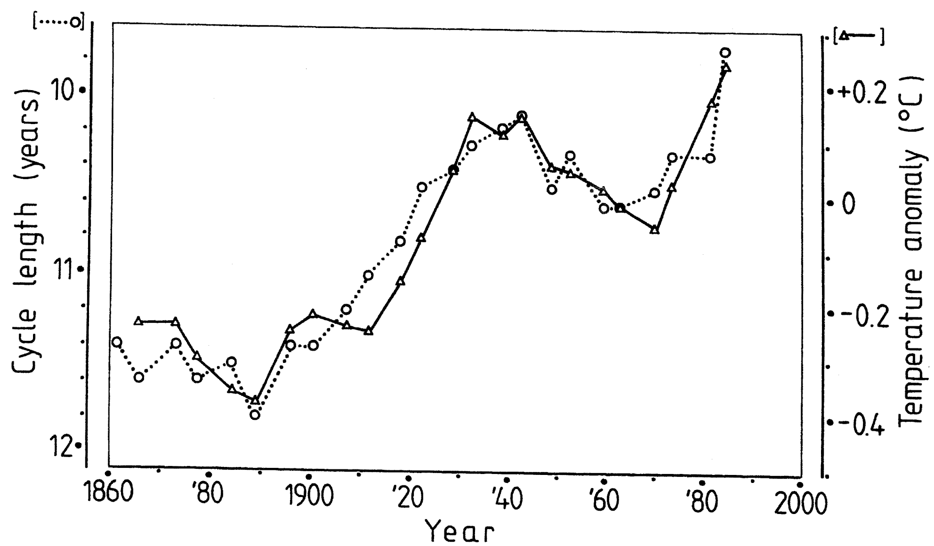

Fig. 1. Yearly global averages of the climate temperature change, and duration of the sunspot cycle, *versus t*. Redrawn after Friis-Christensen and Lassen (1991). A more exhaustive discussion, dealing with a longer time interval, is given by Friis-Christensen and Lassen (1995).

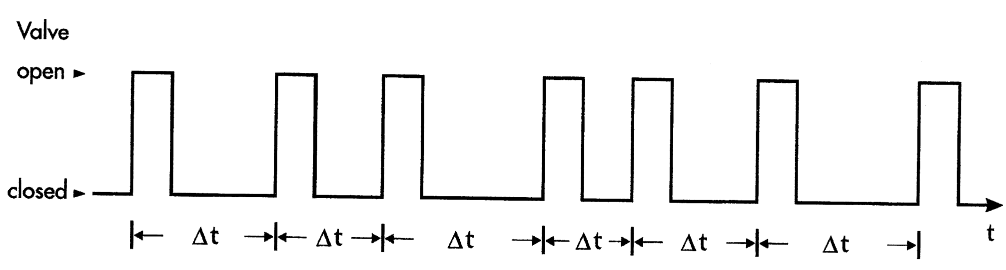

Fig. 2. The opening and closing of the valve of a pressure-cooker occurs at time-lags Δt, the (time-varying) duration of which is inversely proportional to the intensity of its average prime heat-supply.

Plot *versus* t any one given observed $f(t)$ (either continuous, such as a temperature, or atmospheric pressure, or else, or also a point-like event, such as the afore-mentioned pressure-cooker whistler). For simplicity, refer to two specific such examples: pressure-cooker and sunspots. In general, one can imagine plotting any kind of arbitrary function $F(f(t))$ of $f(t)$ that can be defined in any most general way, with the only restriction that $F(f)$ *versus* f must be always either increasing or decreasing. Owing to such a condition, the maxima (minima) of $f(t)$ are still maxima (minima) for $F(f(t))$ (or maxima and minima are interchanged in the case of a decreasing $F(f)$ *versus* f). One can thus imagine, with no loss of generality, plotting some suitable unknown function $F(f(t))$, such that $F(f(t))$ is an actual physical energy. The maxima and/or minima of such an unknown energy $F(f(t))$ are therefore located at the same t of the maxima and minima of the known $f(t)$. Let us call «event» the occurrence of a maximum (or equivalently the occurrence of a minimum). We can now reason just like in the case of the pressure-cooker: any one given event occurs only provided that the energy of the system (supplied by some unknown cause) reaches some threshold. This is the calorimetric criterion: the relative variation of the Δt between the occurrence of any two such pre-defined point-like events is a reliable indication of the relative time variation of the prime energy-supply to the phenomenon itself. Concerning the example of sunspots, see fig. 1. Let us add just a comment on $F(f(t))$. In the case of the pressure cooker, the most obvious choice is $F(f(t)) = 1/\Delta t$. For sunspots, in fig. 1, they plotted $F(f(t)) = -\Delta t$. However, for $\Delta t \sim 11$ years, it is

$$\frac{1}{\Delta t} = 11^{-1}\left[1+\left(\frac{\Delta t}{11}-1\right)\right]^{-1} = \tag{4.1}$$

$$= \frac{2}{11} - \frac{\Delta t}{121} + O\left[\left(\frac{\Delta t}{11}-1\right)^{2}\right]$$

by which the choice $[-\Delta t]$ is almost identical to $[1/\Delta t]$ apart a change of units on the ordinate axis.

Summarizing, such a well known fig. 1 is the result of two different algorithms for low-pass filtering: one (applied to climate temperature) derived by a simple running-mean, *i.e.* in terms of mere mathematical filtering; the other one (applied to sunspots) in terms of the calorimetric criterion, or in terms of a physical

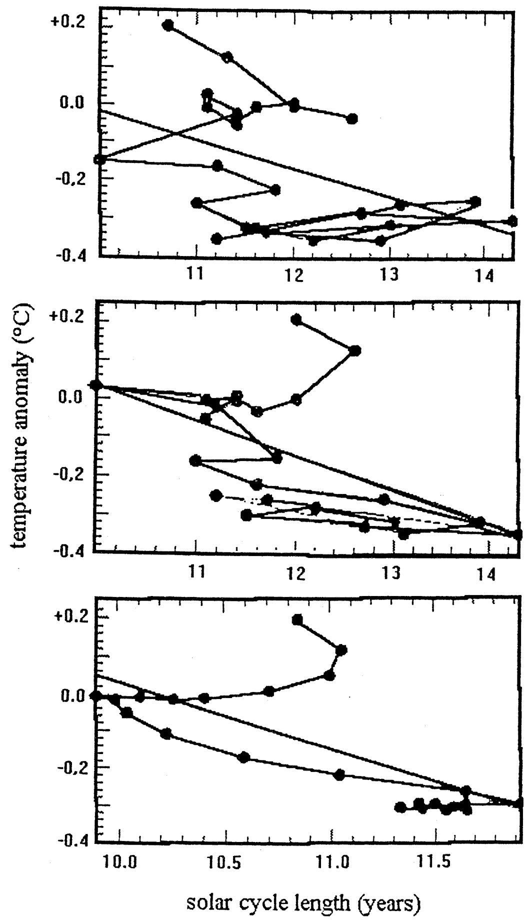

Fig. 3. «*Solar cycle length versus average temperature per solar cycle. Lines connect data points in time sequence.* 3a *is a plot of the 'raw' data with linear fit* ($\rho = -0.43$)*.* 3b *is the same but with temperature lagging by one solar cycle* ($\rho = -0.53$)*.* 3c *includes the one cycle temperature lag and trapezoidal filtering* (1-2-2-2-1)». Figure and captions after Damon and Peristykh (1998).

energy threshold. The conclusion is that, irrespective of the details that can be claimed by any modelling for physical interpretation, such a finding implies that the prime energy-source, that supplies both such phenomena either directly or indirectly, appears likely to be the same. One main focus of the present discussion deals with the complementarity of such two logically different approaches to obtain the same target, *i.e.* low-pass filtering of an observational $f(t)$. This is accomplished in the case either of a well-behaved $f(t)$ according to standard mathematics, or of indirect information (such as sunspots that are an indirect index of the prime, although unknown, energy-supply to the Sun) according to the calorimetric criterion.

Figure 1 has raised some debate. Additional discussion is given by Landscheidt (1997) and Feng and Tang (1997), plus several papers in preparation by Tang Maocang *et al.* (private communication, 1997). Let us mention that some concern is even related to the applied filtering technique (Kelly and Wigley, 1992), a concern further emphasised by Damon and Peristykh (1998), who did not confirm fig. 1; rather, they obtained fig. 3 that manifestly disagrees with fig. 1. Such a disagreement shows how critical the basic methodological discussion is. However, it is sometimes very difficult to get rid of such a drawback, as not all authors have published detailed logs of their respective original input database. In fact, all such data handling is likely to be equally reliable and conceptually correct, although they eventually relied on almost subconscious different basic assumptions of their respective rationale and procedure.

5. Imbò's algorithm

This is a special case of a calorimetric criterion that, up to the author's knowledge, dates back to Imbò (1928), who applied it to the series of the subset of the historical eruptions of Etna that opened a new *boca*. He considered the volcanic cycles that apparently result when realizing that Etna regularly alternates epochs of frequent eruptions with epochs of comparatively less frequent activity. He considered the total time Δt elapsing in between any two subsequent eruptions, and he plotted such a Δt *versus* the order number of every eruption in his eruption log (thus obtaining an «Imbò's histogram»). He found an oscillatory trend, by which he could clearly recognize some volcanic cycles, delimited by epochs of comparatively greater volcanic activity. Every cycle had, however, a varying total duration. His results were subsequently controverted, basically because his database was incomplete. Moreover, one basic concern relates to the definition of a «new» eruption, compared with an event that is only the continuation of a previous one (*e.g.*, Gregori, 1993, 1996, 1996a, and Gregori *et al.*, 1992, 1994a). There was also some misunderstanding, dealing with a supposed 11-year solar-cycle control (claimed by some authors) about Hawaiian vulcanism. Such volcanic cycles were therefore finally abandoned (Bullard, 1984, and references therein).

The author and co-workers (Gregori *et al.*, 1992, 1994a; Gregori, 1995a, 1996, 1996a) resumed such a procedure by using the currently available more complete database. They called it *Imbò's algorithm*, and applied it to the historical eruption series (of either all kinds of eruption, or only the lava-producing ones) for either Etna or Vesuvius. Finally, they applied the same energy constraint of the calorimetric criterion by which the relative time variation of the prime heat-supply is proportional to $1/\Delta t$. By this, they drew fig. 4 (bottom plot, that therefore is the result of a physical low-pass filtering) that was interpreted (Gregori, 1991) by guessing that the apparent ~ 500-600% increase in the prime heat-supply to either Etna or Vesuvius, and that occurred during the last ~ 4-5 centuries, ought to explain the transition from the Little Ice Age to the present climate (in fact, a higher volcanic activity implies a higher concentration of greenhouse gases injected into the atmosphere, hence a higher climate temperature). Moreover, an unexpected by-product resulted in the prevision of the next eruption of Vesuvius (Gregori, 1993, 1995a, 1996, 1996a).

A most surprising result, however, is (Gregori *et al.*, 1992) the close correlation between such a plot for Etna and Vesuvius, and either the geomagnetic or solar activity index, ob-

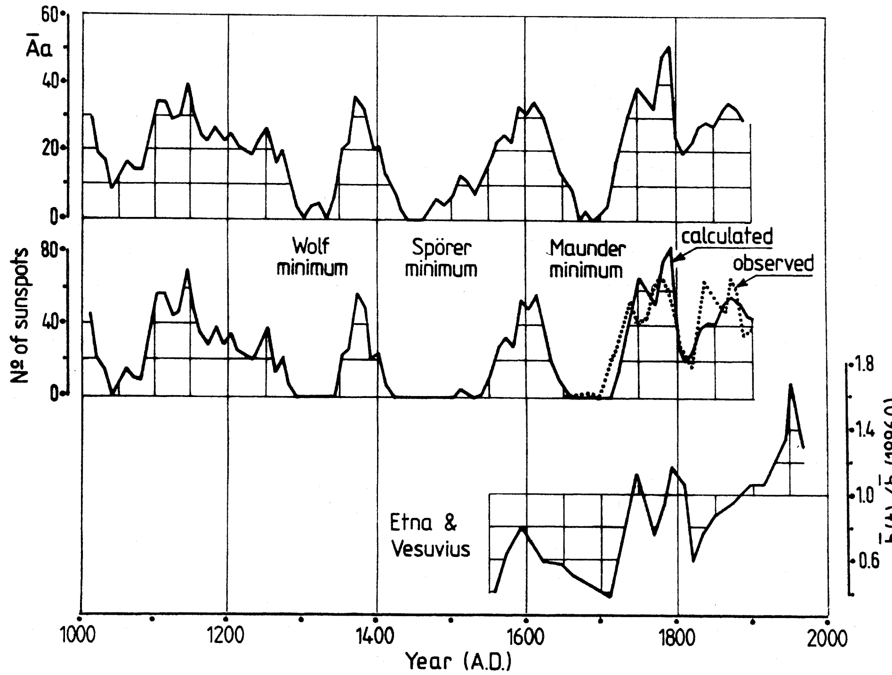

Fig. 4. Top two diagrams: geomagnetic and solar activity derived from the total original ^{14}C concentration within samples of historical wood (after Stuiver *et al.*, see text). Bottom diagram: apparent prime heat-supply $h(t)$ to Etna and Vesuvius *versus* t. The unit heat-supply h (1986.0) is conventionally pre-chosen. This trend was derived by considering the varying duration of the volcanic cycles (*i.e.* delimited by epoches of comparatively greater volcanic activity within an Imbò's histogram; see text). There appears to occur an increase by ~ 500-600% of the prime heat-supply during the last ~ 4-5 centuries, that was speculated to be responsible for the transition between the Little Ice Age and the present climate. See text (figure after Gregori, 1996).

tained by means of the total concentration of ^{14}C within historical wood (Stuiver and Quay, 1980; Stuiver and Becker, 1986; Stuiver and Braziunas, 1988, 1989). The principle idea is as follows. A comparatively higher solar activity implies a lower cosmic ray flux (Forbush decrease), hence a lower production rate of ^{14}C. We know sunspots by visual observation since A.D. ~ 1700, by which we can draw a regression line between ^{14}C concentration and sunspots. Then, for any older epoch we can measure the ^{14}C concentration, and by such a regression line we can infer the corresponding sunspot number. Stuiver and co-workers (*ibidem*) applied such a method for ~ 9600 years BP. Part of their results is shown in fig. 4, top and intermediate diagrams, that are therefore the result of a mathematical low-pass filtering on the well-behaved $f(t)$ of the ^{14}C record. Note that the correlation between solar activity and vulcanism shown in fig. 4 does not apply to the 11-year solar cycle (thus confirming the afore-mentioned 11-year controverted dependence).

Moreover, the afore-mentioned Little Ice Age inference is correct only provided that all volcanoes of the world display a synchronous relative time-variation of their respective prime energy-supply. This fact was successfully checked for several volcanoes of the world, although in general the historical information is not as wealthy as for Etna and Vesuvius (which are the historically best documented volcanoes): upon consideration of all other volcanoes (36 case histories) for which at least 30 historical eruptions were reported (up to 1980), it was possible to check that the hypothesis of such a synchronous variation is consistent with observations (Gregori *et al.*, 1994a, who used the catalogue by Simkin *et al.*, 1981; an updated edition is now available, Simkin and Siebert, 1994). Later on, it was realized that the same trend of maxima and minima displayed in fig. 4 was also reported: i) for a tree-ring index derived from the pines of the Urals (Cook and Briffa, 1995; Gregori, 1997); and ii) for the acidity measured within an ice-core drilled on Monte Rosa in the Western Alps (Tomadin *et al.*, 1996).

This same physical mechanism (for the control by global vulcanism on global climate) was proposed (Ronov, 1983; Budyko *et al.*, 1987) on the geological time scale during the entire Phanerozoic, in terms of the so-called principle of life preservation, by which the carbon cycle has one main source (volcanoes) and one main sink (life). In the case of no vulcanism, life ought to consume the entire CO_2 atmospheric budget, and the Earth should become a desert, much like Mars. In the case of an excess production of CO_2 the Earth should be a greenhouse oven, much like Venus. Concerning real time phenomena, one can attempt to monitor by satellite the interannual variation of the chemical composition of the atmosphere, but this still appears to be technically difficult (Gregori, 1995; Ferrucci *et al.*, 1996, 1997).

A few additional properties of the Imbò algorithm appear relevant. The calorimetric criterion is not directly applied to the database, rather to the time lags in between any two subsequent new eruptions. Moreover, the Imbò histogram that is thus obtained for Vesuvius immediately shows an unambiguous cyclicity, unlike Etna's for which a low-pass filtering (moving average) had to be applied in order to obtain clear evidence of a cyclicity. That is, consistently with the recommendation by Tuckey (1977), the information hidden within an observational data series must be evidenced in any suitable way, and it must be clear, unambiguous, self-evident, and no general algorithm applies to every database. The exploratory analysis should never be identified with the confirmatory analysis.

6. The box-counting method (or some related more complex fractal algorithm)

The Imbò algorithm has a logical link with the box-counting method of fractal theory (*e.g.*, Feder, 1988; Peitgen *et al.*, 1992; Turcotte, 1992; and references therein). In fact, the ruler method implies considering an *ND* space (*e.g.*, a 2D geographic map) and an (*N*-1)D variety $S_{(N-1)D}$ in it (*e.g.*, a coastline) that, however, must be continuous, and it must have either a start- and an end-point, or it has to be a loop (*e.g.*, an island). Moreover, define an (*N*-1)*D* rigid ruler $\mu_{(N-1)D}$ of a given size, and by it measure the «length» of $S_{(N-1)D}$. The box-counting method implies, rather, considering the same *ND* space (*e.g.*, a map in 2D), an *ND* set S_N of points (*e.g.*, any ensemble of coast lines and/or any other feature such as small islands, etc.), and measuring S_N by an *ND* rigid box μ_{ND} of given size (*e.g.*, on a map use a polygon of pre-chosen fixed shape and size). But, unlike when applying the ruler method, such a set S_N does not need to be continuous and it can be defined in any most general way: *e.g.*, in 3D its elements can be either point-like, or line-like, or surface-like, or even continuous 3D distributions of finite size. The fractal properties of either $S_{(N-1)D}$ or S_N deal with the intrinsic repetitivity of their respective features on different spatial scales within the *nD* space. In fact, one definition of fractal was proposed by Mandelbrot in 1986: «*a fractal is a shape made of parts similar to the whole in some way*». In natural reality, an actual system behaves like a fractal only within some given range (*e.g.*, its fractality breaks down on the atomic scale). A fractal property is an abstract concept, much like (in geometry) the concept of point, of line, of surface. Natural reality fits, or not, with such abstract concepts, much like *e.g.*, a measured distribution fits, or not, with the central limit theorem and with the normal distribution. The fractal dimension is a way of setting logical order within some properties that otherwise ought to be simply considered random.

Consider a sequence of $\mathcal{M}$ point-like events (such as *e.g.*, a catastrophe of a given type) that occurred according to a pace typical of some pre-chosen and known cycle *C* of total duration

$\mathcal{T}$ (in terms of the afore-mentioned Imbò rationale for volcanic eruptions). Call Δt_1 and Δt_2 the shortest and the longest time-lag, respectively, in between any two subsequent events within C. Consider an ideal sequence $f(t)$ composed of n_c such identical C, subsequent to each other, every one composed of such $\mathcal{M}$ events: say that $f(t)$ contains altogether $\mathcal{N} = n_c\mathcal{M}$ elements. Consider a box of total time span δ_t and measure by it $f(t)$, *i.e.* completely cover $f(t)$ by an ordered time sequence of identical and contiguous boxes δ_t and count only 1 event every time one such box contains at least one event of $f(t)$. In this way for every given δ_t evaluate a corresponding number $N(\delta_t)$ of counted events. Plot (fig. 5) [log $N(\delta_t)$] *versus* [log δ_t]. When $\delta_t \le \Delta t_1$ it is $N(\delta_t) = \mathcal{N}$ or a constant trend. When $\delta_t > \Delta t_2$ you get $N(\delta_t) = [n_c\mathcal{T}]/\delta_t$ or a decreasing trend. When $\Delta t_1 \le \delta_t \le \Delta t_2$ the trend smoothly changes between constant and decreasing, with a detail that depends on the shape of C. If the original C is composed, *e.g.*, of a sequence of three C's every one having a duration $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$, respectively, let us treat $f(t)$ as being originated by one C alone of duration $\mathcal{T}$ equal to the minimum common multiple of $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$. In such a case, fig. 5 displays a few almost discontinuous, or almost step-like, features in the transition region between constant and decreasing trend, every such feature being associated with either $\mathcal{T}_1$, or $\mathcal{T}_2$, or $\mathcal{T}_3$, respectively. Such a transition region is shown here below to be the most interesting information provided by such a diagram. The equation of the decreasing portion is

$$\log N(\delta_t) = \log n_c + \log \mathcal{T} - \log \delta_t \, . \qquad (6.1)$$

The absolute value of the slope of such a line defines the fractal dimension D, that in the present case is just $D = 1$ (by (6.2) here below). From (6.1) if we know n_c, we can evaluate $\mathcal{T}$, that is associated with the intercept of the line (6.1) with the ordinate axis: hence, we can define a scale on it that directly provides $\mathcal{T}$. Owing to the calorimetric criterion, or to the Imbò algorithm, we know that $\mathcal{T}$ is inversely proportional to the prime energy-supply $\mathcal{E}$ to the ideal $f(t)$. Differently stated, the ordinate axis in fig. 5 can be re-labelled by reversing its

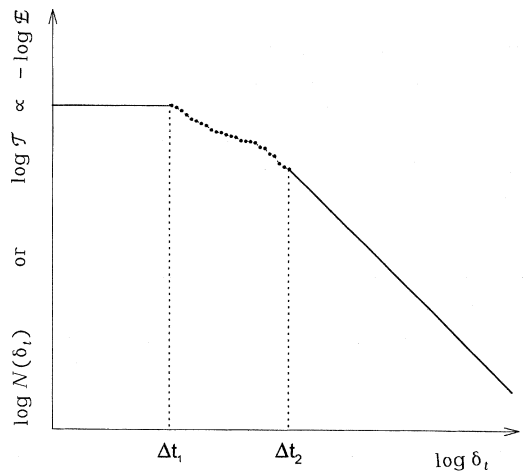

Fig. 5. Plotting [log $N(\delta_t)$] *versus* [log δ_t] resulting in two portions, displaying, respectively, a constant and a decreasing trend. The absolute value of the slope defines the fractal dimension D. It can be shown that in the present case it is always $0 \le D \le 1$. The heuristically most interesting part of the diagram is the transition region between such two portions. See text.

orientation, and by putting an energy scale on it by changing the former units into their inverse values (the energy, however, is thus expressed in some arbitrary units, in the fact that we can only infer its relative time variation).

Refer now to some actual observed time-series $f_{total}(t)$ of point-like events composed of $\mathcal{N}_{total}$ elements. Consider some pre-chosen, suitable, and fixed number $\nu < \mathcal{N}_{total}$. Select a moving sub-time-series $\phi(t_0)$ composed of $\nu(t_0)$ elements, and associate it with a time instant t_0 corresponding to the central epoch of its entire time-span. Consider a fictitious time-series $f(t_0)$ composed of a sequence of $n_c \sim \mathcal{N}_{total}/\nu$ identical such $\phi(t_0)$'s and apply the box-counting method. Evaluate by means of a plot like fig. 5 its corresponding $\mathcal{T}(t_0)$. Finally, plot $1/\mathcal{T}(t_0)$ *versus* t_0, that, apart from the unknown units, is the time-variation of the average prime energy-supply to the system. By this, let us guess that we inferred the prime energy that feeds the physical system during the time interval spanned by ϕ, and for simplicity we state that such energy is approximately associated with the time instant t_0. Such a procedure, however, does not work, as it is $1/\mathcal{T}(t_0) = 1/v(t_0)$ just by definition, *i.e.* such a procedure is tautological. We have, rather, to refer to $f_{total}(t)$ and to focus on the transition region (in fig. 5) between constant and decreasing trend. Plot *versus* [log δ_t] the derivative $[(-d \log (N(\delta_t))/(d \log \delta_t)]$ that, owing to fig. 5, increases from 0 to 1, although in general it is not monotonic: its spikes give evidence of the eventual periodicities of $f_{total}(t)$. The actual details of such a spectrum depend on the physical separation between different such periodicities within $f_{total}(t)$ itself. Moreover, it can be formally proven (shown below) that, in the present case, it is always $0 \leq D \leq 1$, or the height of every such peak is always ≤ 1.

Note, however, that these same periodicities can also be obtained in a much more straightforward way simply by ARP. Therefore, in such a respect, the box-counting method appears useful only for a matter of logical comparison. In either case, we can plot the variation *versus* t of some apparent dominating periodicity associated with $f_{total}(t)$. However, although in general the most direct application of the Imbò algorithm achieves the same final result in a much simpler way, in special circumstances the Imbò histogram will eventually prove not viable, and the formal fractal approach will therefore be much more effective. For example, in general, a historical data series contains several gaps (the extent, location, and frequency of which is unknown), sometimes even deriving from an unconscious change in event definition adopted by the chronicler (refer to the psychological factor mentioned in Section 2). Therefore, *a priori* nobody can be sure that some algorithm can apply successfully to any given $f(t)$. For example, when ARP was applied to the ~ 25 century series of the Tiber flood in Rome, it gave inconclusive results, while it was successful for the Tanaro valley (Section 3).

Let us prove that, in the present case, it is always $0 \leq D \leq 1$. The line in the Richardson plot, *i.e.* [log $N(\mu)$] *versus* [log μ], is (*e.g.*, Feder, 1988 or Turcotte, 1992)

$$N(\mu) = C\,\mu^{-D}. \qquad (6.2)$$

The parameter D is called fractal dimension. Consider a time series of events, occurring at times $\{t_j\}$ ($j = 1, 2, ..., M$), spanning altogether a total time interval of duration T. The aforementioned analysis amounts: i) to choose a given trial interval μ; ii) to cover the total time interval T by a set of $\sim T/\mu$ contiguous trial intervals, every one of duration μ; and iii) to count the total number $N(\mu)$ of elements of the subset of such trial intervals that contain at least one event. Let us suppose that the events are equally spaced, and call $\Delta t = t_{j+1} - t_j = \text{const} = T/M$ for every j. In such a case, it is (fig. 6a)

$$N(\mu) = M \quad \text{or } D = 0. \text{ for } \mu < \Delta t \qquad (6.3a)$$

$$N(\mu) = T/\mu \quad \text{or } D = 1. \text{ for } \mu \geq \Delta t \qquad (6.3b)$$

$$N(\mu) = 1 \quad \text{or } D = 0. \text{ for } \mu \geq T. \qquad (6.3c)$$

In the most general case, let us call the time lag between any two consecutive events

$$\Delta t_j = t_{j+1} - t_j \quad (j = 1, 2, ..., M-1) \qquad (6.4)$$

and let us define

$$\mu_{min} = \min\{\Delta t_j\} \quad \mu_{Max} = \text{Max}\{\Delta t_j\} \qquad (6.5a,b)$$
$$(j = 1, 2, ..., M-1)$$

and check that (fig. 6b)

$$N(\mu) = M \quad \text{or } D = 0. \text{ for } \mu \leq \mu_{min} \qquad (6.6a)$$

$$N(\mu) = T/\mu \quad \text{or } D = 1. \text{ for } \mu > \mu_{Max} \qquad (6.6b)$$

$$N(\mu) = 1 \quad \text{or } D = 0. \text{ for } \mu > T. \qquad (6.6c)$$

However, in general, for $\mu_{min} < \mu < \mu_{Max}$ it can be either $0 < D < 1$ (fig. 6b), or $D \geq 1$ (fig. 6c). But, let us consider the actual given $\{t_j\}$ data series, and let us re-define a new time series $\{t_j\}$ by means of the same events, but by redistributing them uniformly over the same total time interval T. That is, let us imagine transforming, in some continuous way, fig. 6a into either fig. 6b or c, by means of a suitable progressive re-definition of $\{t_j\}$, starting from a

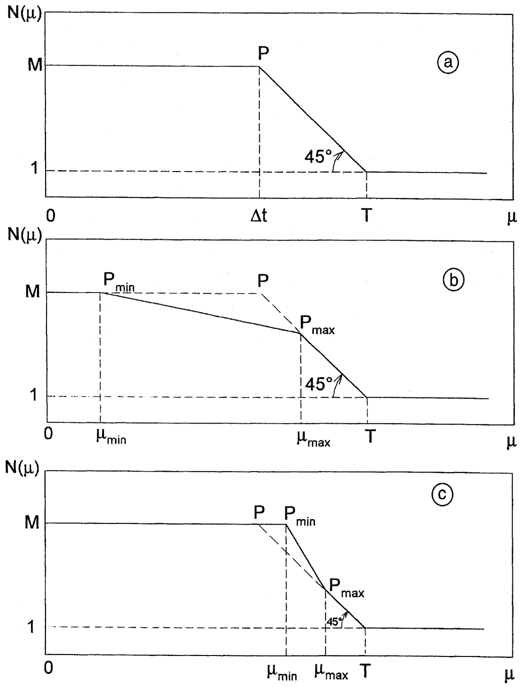

Fig. 6a-c. Richardson's plot for the counts $N(\mu)$ of a series of events, occurring at times $\{t_j\}$, *versus* the sampling «box» of duration μ. a) Case of uniform relative spacing of the times $\{t_j\}$. b) General case with $0 < D < 1$. c) General case with $D > 1$, that can never occur.

uniform distribution, towards the final actual given $\{t_j\}$. We note that any re-definition of the distribution of $\{t_j\}$, that implies a deviation from the uniform distribution, can only produce a decrease of the straight line between points M and P in fig. 6a. Moreover, it can be clearly realized that the point P_{min} can never be to the right of point P, as the minimum Δt_j can never exceed the $\Delta t = T/M$ of fig. 6a (in fact, if a variate gets some larger value than its own mean, we must also find some value less than the mean itself, hence the minimum value can never exceed the mean). Therefore, fig. 6c will never occur. As far as fig. 6b is concerned, either the plot between P_{min} and P_{Max} is an approximately straight line, or not. In the first case, we conventionally say that the phenomenon «behaves like a fractal», and that the fractal dimension must be $0 \leq D \leq 1$. When the trend is not linear, we say that the phenomenon «does not behave like a fractal».

Some implications of this algorithm can be clarified by a means of a simple example. Imagine that a large round apple-pie has to be shared among a large number M of children. The azimuth ϕ, that defines the slices of the pie, is to be likened to the time t of the time series above, whose total time interval T is to be likened to the total round surface of the apple-pie. An equity criterion is to give a slice of one and the same size to every child. In this case, the application of the algorithm above gives fig. 6a. Instead, suppose applying a different criterion, *e.g.,* by considering the age of every child, whether he is more or less tall or fat, etc. Therefore, suppose a slice of a different size is given to every child. In such a case, the algorithm above gives a figure similar to fig. 6b, where, however, the trend between P_{min} and P_{Max} can be either linear or not, depending on the criterion chosen for deciding the sizes of the different slices. For example, suppose that a few children are sick (say a very small number $m \ll M$ of them), so that due to their health they get only an almost negligible slice of pie. In such a case, P_{min} in fig. 6b gets very close to point M, although this does not imply that the trend between P_{min} and P_{Max} is (even only approximately) linear. Hence, the phenomenon may, or may not, «behave like a fractal». However, if one neglects the specific case of the $m \ll M$ sick children that altogether get a total amount $\varepsilon \ll T$ of apple-pie, it is possible that the total remaining amount $T - \varepsilon \sim T$ of pie, being distributed among $M - m \sim M$ children, still represents a phenomenon that «behaves like a fractal». For example, suppose that, apart the m sick children, the remaining pie is distributed according to the equity criterion as above. In such a case, fig. 6a-c is only slightly modified, in the fact that the plateau between M and P very soon decays to a lower plateau of level M-m, while point P is only very slightly displaced towards the bottom-right, *i.e.* from co-ordinates $(\Delta t, M)$ to $(\Delta' t, M - m)$ where $\Delta' t = (T - \varepsilon)/(M - m)$, while the fractal dimension still is $D = 1$. That is, a few sick children do not affect the information that we can obtain from fig. 6a-c relating to the basic rationale applied when dividing the apple-pie.

Moreover, suppose that, in addition to the case of the m sick children, we have additional constraints, by which the entire set of children is divided into several subsets, every one of which has a slice of a given size. Figure 6a progressively transforms as follows. Let us describe fig. 6a as a plateau between M and P, followed by an inclined plane with a 45° slope (that, in fact, is not a plane, as it is a staircase, as $N(\mu)$ is an integer function, hence it is step-wise and not continuous). Every subset of children, that have a slice different from other subsets', determines a fraction of the staircase having its own slope. Hence, the former staircase of fig. 6a has only one leap of 45° slope, while in the modified actual case it eventually divides into a staircase composed of several leaps, each with its own different slope. But, as noted above, the staircase can only be «eroded» such as from fig. 6a towards fig. 6b, and never «accreted» such as from fig. 6a towards fig. 6c. Moreover, the erosion determines either one, or a few, leaps between P_{min} and P_{Max}, distinguished from the very last 45° leap for $\mu > P_{Max}$. Whenever the newly defined staircase composed of several such leaps can be approximately described in terms of one unique leap between P_{min} and P_{Max}, it is possible to define an average slope and a fractal dimension D, although, in general, the composition of the new staircase can eventually deviate from an approximately straight line.

Summarizing, such a fractal algorithm is a way of computing some kind of an index (the fractal dimension D) that is reminiscent, in some way, of the criterion applied when cutting the apple-pie.

Similarly to such a simple example, when applying this algorithm to any given data series, D is reminiscent, in some way, of the rationale that determines the pace of the events within the data series. Summarizing, our concern is about the trend of $N(\mu)$ within the Richardson plot, but only within the interval $\mu_{min} < \mu < \mu_{Max}$. Whenever such a trend is (approximately) linear, we conventionally say that the phenomenon «behaves like a fractal», and we define D in terms of its slope.

One possibility is searching, *e.g.,* by means of a Monte Carlo method, for the distribution of the D values corresponding to a random input data series. In some respect, this is of little help. We can apply, rather, such a fractal analysis to several different data series, and consider directly whether their respective D's are identical or not. However, when carrying out actual applications, it was realized (Wenjie Dong, within an investigation in cooperation with the author and with A. Poma, in preparation) that the definition of μ_{min} according to (6.5a) was not convenient, due to several facts related to the uneven nature of the database. For practical purposes the definition (6.5a) had to be changed into

$$\mu_{min} = L/M \tag{6.7}$$

(where L is the total time span of the series $\{t_j\}$), that amounts to investigating, within fig. 6b, only the slope of some suitable right portion of the segment $[P_{min}, P_{Max}]$.

When comparing the D of several data series, the implicit rationale is that all such phenomena, independent of their eventual mutual time-delays (that can be null or not, and either constant, or even varying, on different circumstances, depending *e.g.,* on different calorimetric roles played by different energy reservoirs as per Section 4), should reveal, in any case, the same D, provided that the pace of their respective occurrence is controlled by one and the same prime mover. This, however, is a necessary, although not sufficient, condition. Such a rationale can appear either reasonable or not, depending on the kind of natural event that is considered. For example, a volcanic eruption is an incontrovertible physical event, measured in some well defined way. In contrast, a climate temperature series, or a sea level change, is displayed as a curve that must be intrinsically smoothed over a large set of indirect data points, every one of which in general can hardly be characterized by its own error bar. Therefore, even the same definition of their respective relative minima or maxima depends on some arbitrary choice, when we must decide how to smooth a curve over such scattered data. This means that, when comparing, *e.g.,* the D

either of volcanic, or of the climate-temperature, or of the sea-level data series, we implicitly (and unavoidably) assume, as a working hypothesis, that the relevant trend of either the temperature or the sea level is physically controlled by the pace of vulcanism, in addition to, and independent of, eventual other effects (such as the Milankovitch modulation) that introduce some additional variability. Hence, we implicitly, arbitrarily, and tentatively, presume that any such additional variability is less important than the effect due to endogenous energy. Moreover, we assume all this, in addition to the afore-mentioned drawbacks concerned with the different error-bars of the data sets.

A concern is about the formal error-bar of every D, when assessing whether any two D's of two different data series ought to be considered equal or not. For such a purpose, let us consider the distribution of every estimate of D. Namely, call $\{t_j^{(k)}\}$ ($j = 1, 2, \ldots, M^{(k)}$) the series of the dates for the k-th variate, and define the set $\{D_i^{(k)}\}$ being $i = 1, 2, \ldots, \binom{M^{(k)}}{2}$ where every $D_i^{(k)}$ is one estimate of D resulting from consideration of any one given couple of points drawn on the Richardson plot for the k-th variate. Then, one can evaluate for every k both the average $D^{(k)}$ of such a set, and its r.m.s. deviation $\Delta D^{(k)}$. Moreover, one can evaluate the probability that any two distributions $\{D_i^{(k)}\}$ and $\{D_i^{(h)}\}$ $\{k \neq h\}$ have significantly different mean values, or not. Such a probability can be estimated in terms of either one of two different Student's t-tests: the first one applies when the two distributions are believed to have the same variance, and the second one when the two distributions are supposed to have significantly different variances (*e.g.*, use the subroutines TTEST and TUTEST by Press *et al.*, 1986).

The box counting method is a simple algorithm that can be generalized in several different ways. In terms of the symbols used in (6.2), D is the slope of the line in the Richardson plot [log $N(\mu)$] *versus* $\log \mu$, or it can be defined as the limit of such a slope for $\mu \to 0$, provided that one considers only the physical values for μ. In the case of a function $f(t)$ of time t (*i.e.* this is not a point-like process), the following method was proposed by Grassberger and Procaccia (1983) (see also de Santis *et al.*, 1997). Suppose that the d.o.f.'s of the physical system are m (this is a trial parameter called embedded dimension). Consider the set

$$\{f_i\} \equiv \{f(t_i), f(t_i + \tau), \ldots, f(t_i + (m-1)\tau)\} \quad (6.8)$$

where t_i ($i = 1, 2, \ldots, N$) are the time instants of the available records of $f(t)$, and τ is a conventional and given time increment. Such a set $\{f_i\}$ can be representative of the m supposed d.o.f.'s. Consider an mD space, where every point is defined by one such set $\{f_i\}$. Apply as follows some suitable generalisation of the box counting method within such an mD space. Consider an mD box, of linear size r, and let us count the points falling inside it, or their fraction with respect to the total number of the available measured points. Grassberger and Procaccia (ibidem) thus defined the correlation dimension $D_c(m, r)$ as follows

$$D_c(m, r) = \lim_{(m, r \to \infty, 0)} \frac{\log C_m(r)}{|\log r|} \quad (6.9a)$$

$$C_m(r) = \frac{1}{N(N-1)} \sum_{\substack{i,j=1 \\ (i \neq j)}}^{N} \Theta(r - |f_i - f_j|) \quad (6.9b)$$

$$\Theta(u) = (1 + \text{sign}\, u)/2 \quad (6.9c)$$

where Θ is the Heaviside function.

Such an algorithm, however, cannot be applied to a point-like process, because the ordinate $f(t)$ of the record is essential. Nevertheless, the physical idea is the same. This is just an example of the variety of different fractal algorithms that can be envisaged and that can be more or less directly reckoned to each other. The crucial logical key is that the fractal analysis investigates the intrinsic regularity of the pace of occurrence of a given kind of event, eventually by means of a robust algorithm in order to get rid of the drawbacks of the incompleteness of the historical data series.

7. Conclusions

Catastrophe management can rely either on a deterministic approach, *i.e.* by attempting to work out some more or less detailed model by which we can recognize some crucial precursor, that ought to be hopefully capable of providing with an error-free forecast with a time- and space-range suited for the practical needs by mankind. In general, however, such a target often appears beyond our reach, due to the several d.o.f.'s and to our poor understanding of several basic mechanisms that govern phenomena.

It appears therefore convenient to rely, as far as possible, on the historical information, capable *per se* of providing some useful hints for forecasting eventual more or less regular return-times of some unwanted conditions, by which the occurrence of a catastrophe becomes comparatively more probable. Such a viewpoint can be defined as a pragmatic approach. A time series of catastrophes, however, can only be approximately likened to a point-like process, due to the concurrence of both natural and anthropic causes: a natural phenomenon is considered a catastrophe only provided that it causes damage and/or casualities, and such an effect is associated with the detector/mankind, depending on its use of territory. Since this has changed during history, the detector of the event also implicitly changed in time: thus, such a time series is necessarily uneven.

It is possible to take advantage of the algorithms of point-like processes only provided that some suitable warnings are considered. In general, four basic viewpoints, or *a priori* assumptions, or rationales, can be applied: i) assuming that phenomena are periodical; ii) assuming that an event occurs only whenever some suitable (although eventually unknown) energy reservoir is filled up to some threshold (calorimetric criterion); iii) assuming that an event occurs whenever the physical system experiences some abrupt change of its boundary conditions; and iv) whenever, owing to insufficient observational information, no such rationale can be applied, the box counting method (or some more involved equivalent algorithm of fractal analysis) can provide interesting hints on the possible cause-and-effect relations among different observational time series.

All such algorithms have some intrinsic analogies and mutual logical relations. They have been here highlighted, while in the literature all of them were used on an intuitive basis, to be separately and independently considered in every case study. Their combined application can eventually result heuristically useful. For example, for clarity, let us suppose we are dealing with a continuous well-behaved $f(t)$, such as either a temperature, or the atmospheric pressure, or else, measured at some given site. Let us smooth it in order to reject both the daily and seasonal variations and, eventually, even some other «short-period» variations (such as the 11-year cycle in fig. 4, top and intermediate diagrams). Consider either some resulting function (such as *e.g.,* the temperature curve in fig. 1), or *e.g.,* consider the yearly excursion of the temperature (or else), or, differently stated, the difference between its maximum and minimum during some given moving time interval, etc. *Tout court,* in some way let us state that we obtained from a former $f(t)$ some new function $g(t)$ of any kind. One can apply some fractal algorithm (such as the Grassberger and Procaccia method mentioned in Section 6). Otherwise, define by means of $f(t)$ or $g(t)$ a new time-series $p(t)$ of point-like events, *e.g.,* in terms of the maxima or minima of $g(t)$. Otherwise, let us deal directly with any $p(t)$ of point-like events of any kind. Note that the definition of $p(t)$ can even suffer from gaps (even having some unknown distribution and consistency), and with uneven error-bars *versus* t, etc. Apply to $p(t)$ the calorimetric criterion, or the Imbò algorithm, or the box-counting method (much like it has been done either in fig. 1 for sunspot maxima, or in the Imbò histograms for Etna and Vesuvius for inferring the bottom diagram of fig. 4). In terms of physical interpretation, this is an effective way of recognizing phenomena that can be reasonably guessed to be modulated by the same prime energy-source. Such an inference is independent of all logical uncertainties and approximations associated with any more detailed physical or mathematical modelling. For example, essentially this is the common logical backbone of the algorithms applied to the two observational databases used to draw fig. 1.

A few previous applications of such an approach were apparently useful in providing either a successful prevision of a great flood in Northern Italy, in 1994, or the forecast of the next expected explosive (sub-Plinian) eruption of Vesuvius. The success of such approaches, however, depends on the quality of the available database, while the skill in data handling only depends on the capability of the algorithms in obtaining from the available observations «all» of what is actually and objectively contained within them, without adding anything else related to the arbitrariness by the decisions and implicit assumptions, that are eventually and/or even unconsciously fed in by the researcher during data handling. Therefore, in the ultimate analysis, the wealth and quality of the historical information appears crucial in determining the physical significance of the results (concerning a general discussion of the problems related to historical data, see Gregori and Gregori, 1998a).

Acknowledgements

The author is indebted: to Dr. E.R. Cook for providing with a transparency showing their plot of the Ural-pines tree-ring index; Prof. Paul E. Damon for the reference of fig. 3; Dr. Angelo de Santis for stimulating discussions, and in particular for reminding me about the Grassberger and Procaccia method; Dr. Wenjie Dong for several long and stimulating discussions on related methodological items; Prof. A. Palumbo, for several discussions and in particular for reminding me about his applications of the calorimetric criterion to several different types of natural catastrophes; Prof. Angelo Poma for discussions on fractal analysis; and Prof. D. Wagenbach for a stimulating discussion and for showing their plot related to their Monte Rosa ice-core.

REFERENCES

BANZON, V., M. COLACINO, G. DE FRANCESCHI, L. DIODATO, G.P. GREGORI, M.P. PAVESE and R. SANTOLERI (1990): Tiber floods, anomalous climatic events in the upper Po valley, and volcanic activity, in *Schröder* (1990), 73-79.

BANZON, V., G. DE FRANCESCHI and G.P. GREGORI (1992): The mathematical handling and analysis of non-homogeneous and incomplete multivariate historical data series, in *Frenzel* et al. (1992), 137-150.

BANZON, V., G.P. GREGORI, R. LEONARDI, G. DE FRANCESCHI and W.C. WANG (1992a): An analysis of the 510 year data series of the dryness/wetness index for 120 regions of China, in *Schröder and Legrand* (1992), 223-238.

BRADLEY, R.S., and P.D. JONES (Editors) (1992): *Climate Since A.D. 1500* (Routledge, London and New York), pp. 679.

BRILLINGER, D.R. (1975). *Time Series: Data Analysis and Theory* (Holt, Reinhart and Wenton, New York).

BUDYKO, M.I., A.B. RONOV and A.L. YANSHIN (1985): *Istorya Atmosferi* (in Russian). Gidrometeoizdat. English translation, *History of the Earth's Atmosphere,* published in 1987 by Springer-Verlag, Berlin, etc., pp. 139.

BULLARD, F. M. (1984): *Volcanoes of the Earth,* II revised edition (University of Texas Press, Austin), pp. 629.

COOK, E.R. and K.R. BRIFFA (1995): IUGG General Assembly, Boulder, Co., invited talk (abstract 286).

DAMON, P.E. and A.N. PERISTYKH (1998): Prediction of the contribution of solar activity to 21st century global warming, in *Ninth Symposium on Global Change Studies*, 78th American Meteorological Society Annual Meeting, 11-16 January 1998, Phoenix, Arizona. Preprint.

DE SANTIS, A., G. DE FRANCESCHI and L. PERRONE (1997): Spectral and fractal analyses of geomagnetic and riometric Antarctic observations and a multidimensional index of activity, *J. Atmos. Solar-Terr. Phys*, **59** (9), 1073-1085.

DIAZ, H.F. and V. MARKGRAF (Editors) (1992): *El Niño. Historical and Paleoclimatic Aspects of the Southern Oscillation* (Cambridge University Press), pp. 476.

FAIRBRIDGE, R.W. (1996): Spectra of solar-terrestrial proxies (auroras, tree rings, varves, ice cores) and a proposed astrochronology of the last 15000 years, in *Schröder and Colacino (1996)*, 89-112.

FEDER, J. (1988): *Fractal Analysis* (Plenum Press. New York and London), pp. 283.

FENG, S. and M.C. TANG (1997): The solar activity and air temperature fluctuation in recent 2500 years (in Chinese with English abstract), to appear in *Quaternary Science*, No. 1.

FERRUCCI, F., G.P. GREGORI, R. GUZZI, B. PAVESI and G. RONDINELLI (1996): Natural risk monitoring from space. [Invited]. *ESA Symposium Proceedings on «Space Station Utilization», ESOC, Darmstadt, Germany, 30 Sept.-2 Oct. 1996 (SP-385, December 1996)*, 599-608.

FERRUCCI, F., G.P. GREGORI, R. GUZZI, B. PAVESI and G. RONDINELLI (1997): Risk and catastrophe management by satellite. Feasibility, costs, timing, and priorities. [Invited]. Presented at the *Space Congress '97*, Bremen, 14-17 May 1997 (in press).

FRENZEL, B., C. PFISTER and B. GLÄSER (Editors) (1992): *European Climate Reconstructed from Documentary Data: Methods and Results*, Proceedings of the ESF Meeting on held in Mainz (Germany) in March 1-3, 1990, *Paläoklimaforschung – Palaeoclimate Research*, 7, pp. 265 (European Science Foundation, Strasbourg, and Akademie der Wissenschaften und der Literatur, Mainz, Gustav Fischer Verlag, Stuttgart, etc).

FRIIS-CHRISTENSEN, E. and K. LASSEN (1991): Length of the solar cycle: an indicator of solar activity closely associated with climate, *Science*, **254**, 698-700.

FRIIS-CHRISTENSEN, E. and K. LASSEN (1995): Variability of the solar cycle length during the past five centuries and the apparent association with terrestrial climate, *J. Atmos. Terr. Phys.*, **57**, 835-845.

GRASSBERGER, P. and I. PROCACCIA (1983): Characterization of strange attractors, *Phys. Rev. Lett.*, **50**, 346-349.

GREGORI, G.P. (1990): A few mathematical procedures for the analysis of incomplete historical data series, in *Schröder (1990)*, 80-127.

GREGORI, G.P. (1991): Anthropic control of the natural long-term climatic variation: proposal for a discussion on greenhouse effect controlled by geothermal heat flow, in *Sindoni and Wong (1991)*, 353-360.

GREGORI, G.P. (1993): The next eruption of Somma-Vesuvius, in *Schröder (1993)*, 191-213.

GREGORI, G.P. (1993a): Geo-electromagnetism and geodynamics: «corona discharge» from volcanic and geothermal areas, *Phys. Earth Planet. Inter.*, **77**, 39-63.

GREGORI, G.P. (1994): Geomagnetism, volcanoes, global climate change, and predictability. A progress report, *Ann. Geofis.*, **37** (suppl. n. 5), 1329-1340.

GREGORI, G. P. (1995): Remote sensing of volcanoes and their role in the global climate change, *Adv. Space Res.*, **15** (11), 17-26.

GREGORI, G.P. (1995a): La ricerca di ciclicità e l'algoritmo di Imbò – Applicazione alle serie storiche delle eruzioni dell'Etna e del Vesuvio, in *Piccione* et al. *(1995)*, 729-753.

GREGORI, G.P. (1996): Satellites, volcanoes, and global change, *Earth Space Rev.*, **5** (1), 17-26.

GREGORI, G.P. (1996a): The next eruption of Somma-Vesuvius, in *Piccione and Antonelli (1996)*, **2**, 399-468.

GREGORI, G.P. (1997): Historical data and global change. Case studies, in *Schröder (1997)*, 183-210.

GREGORI, G.P. (1998): The magnetosphere of the Earth. A theory of magnetospheric substorms and of geomagnetic storms, in *Schröder (1998a)*, 68-106.

GREGORI, G.P. and L.G. GREGORI (1998): Restoro d'Arezzo and Earth's sciences in the Middle Age, in *Schröder (1998)*, 315-350.

GREGORI, G.P. and L.G. GREGORI (1998a): Solar-terrestrial relations. A reminder, in *Acta Geod. Geophys. Hung.*, **33** (2-4), 391-459.

GREGORI, G.P., R. SANTOLERI, M.P. PAVESE, M. COLACINO, E. FIORENTINO and G. DE FRANCESCHI (1988): The analysis of point-like data series, in *Schröder (1988)*, 146-211.

GREGORI, G.P., V. BANZON, R. LEONARDI and G. DE FRANCESCHI (1992): Geomagnetic activity *versus* volcanic cycles, and their forecasting. Applications to Etna and Vesuvius, in *Schröder and Legrand (1992)*, 188-222.

GREGORI, G.P., V. BANZON, M. COLACINO, G. DE FRANCESCHI, L. DIODATO, M.P. PAVESE and R. SANTOLERI (1994): Floods, climatic anomalies, and explosive vulcanism: a four-decade cycle?, in *Schröder and Colacino (1994)*, 109-133.

GREGORI, G.P., V. BANZON and R. LEONARDI (1994a): The cycles of volcanoes, and the global synchronism of the time variation of their heat source, in *Schröder and Colacino (1994)*, 152-191.

GUERZONI, S. and R. CHESTER (Editors) (1996): *The Impact of Desert Dust Across the Mediterranean* (Kluwer Academic Publishers, Dordrecht, etc.), pp. 389.

IMBÒ, G. (1928): Variazioni cicliche nella successione dei periodi di riposo etnei, *Bull. Volcanol.* (15/18), 80-88.

JACOBS, J.A. (1987): *Geomagnetism* (Academic Press, Harcourt Brace Jovanovich, Publ., London, etc.), vol. 2, pp. 579.

KELLY, P.M. and T.M.L. WIGLEY (1992): Solar cycle length, greenhouse forcing and global climate, *Nature, Lond.*, **360**, 328-330.

LANDSCHEIDT, T. (1997): Klimavorhersage mit astronomischen Mitteln?, in *Schröder (1997)*, 137-155.

PALUMBO, A. (1997): Chaos hides and generates order: an application to forecasting the next eruption of Vesuvius, *J. Volcan. Geotherm. Res.*, **79**, 139-148.

PALUMBO, A. (1997a): Long-term forecasting of large volcanic eruptions, *J. Volcan. Geotherm. Res.*, **78**, 179-183.

PALUMBO, A. (1997b): A new methodological contribution to forecast extreme damaging events: an application to extreme marine floodings in Venice, *J. Coast. Res.*, **13**, 890-894.

PALUMBO, A. (1998): Long-term forecasting of extreme eruptions of Etna, *J. Volcan. Geotherm. Res.* (in press).

PALUMBO, A. and F. PALUMBO (1998): Storm-surges and floodings in Trieste, Northern Adriatic Sea, *J. Coast. Res.* (in press).

PAVESE, M.P. and G.P. GREGORI (1985): An analysis of six centuries (XII through XVII A.D.) of climatic records from the upper Po valley, in *Schröder (1985)*, 185-220.

PAVESE, M.P., V. BANZON, M. COLACINO, G.P. GREGORI and M. PASQUA (1992): Three historical data series on floods and anomalous climatic events in Italy, in *Bradley and Jones (1992)*, 155-170.

PEITGEN, H.-O., H. JÜRGENS and D. SAUPE (1992): *Chaos and Fractals. New Frontiers of Science* (Springer-Verlag, Inc., New York, etc.), pp. 984.

PHILANDER, S.G. (1990): *El Niño, La Niña, and the Southern Oscillation* (Academic Press Inc., Harcourt Brace Jovanovich, Publ., San Diego, etc.), pp. 289.

PICCIONE, V. and C. ANTONELLI (Editors) (1996): *Atti del 4° Convegno del Progetto Strategico Clima, Ambiente e Territorio nel Mezzogiorno,* Proceedings of a meeting held at Lecce, November 11-14, 1991 (CNR, Roma), pp. 644.

PICCIONE, V., C. ANTONELLI and A. GUERRINI (Editors) (1995): *Atti del Convegno del Progetto Strategico*

Clima, Ambiente e Territorio nel Mezzogiorno, Proceedings of a meeting held at Potenza in 1990 (CNR, Roma), pp. 795.

PRESS, W.H., B.P. FLANNERY, S.A. TEUKOLSKY and W.T. VETTERLING (1986): *Numerical Recipes. The Art of Scientific Computing* (Cambridge University Press), pp. 818.

RONOV, A.B. (1982): The Earth's sedimentary shell, *Int. Geol. Rev.*, **24**, (11/12), 1313-1363, 1365-1388. Reprinted in 1983 by the Am. Geol. Inst., *AGI Reprint Series* V, pp. 80.

SCHRÖDER, W. (Editor) (1985): *Historical Events and People in Geosciences* (Peter Lang, Frankfurt am Main, etc.) pp. 220.

SCHRÖDER, W. (Editor) (1988): *Past, Present and Future Trends in Geophysical Research* (Interdivisional Commission on History of IAGA, Bremen-Roennebeck), pp. 342.

SCHRÖDER, W. (Editor) (1990): *Advances in Geosciences* (Interdivisional Commission on History of IAGA, Bremen-Roennebeck), pp. 358.

SCHRÖDER, W. (Editor) (1993): *The Earth and the Universe, Festschrift in Honour of Prof. Hans-Jürgen Treder* (Bremen, Science Edition), pp. 497.

SCHRÖDER, W. (Editor) (1997): *Physics and Geophysics with Special Historical Case Studies – A Festschrift in Honour of K. H. Wiederkehr* (Science Edition, Bremen-Roennebeck), pp. 409.

SCHRÖDER, W. (Editor) (1998): *Geomagnetism and Aeronomy (with Special Case Studies)* (Interdivisional Commission on History of IAGA and Historical Commission on History of the German Geophysical Society. Science Edition, Bremen-Roennebeck), pp. 355.

SCHRÖDER, W. (Editor) (1998a): *From Newton to Einstein. Festschrift in Honour of the 70th Birthday of Prof. Hans-Jürgen Treder* (Bremen, Science Edition), pp. 489.

SCHRÖDER, W. and M. COLACINO (Editors) (1994): *Geophysics: Past Achievements and Future Challenges* (Interdivisional Commission on History of IAGA, Bremen-Roennebeck), pp. 291.

SCHRÖDER, W. and M. COLACINO (Editors) (1996): *Global Change and History of Geophysics* (Interdivisional Commission on History of IAGA, Bremen-Roennebeck), pp. 291.

SCHRÖDER, W. and J.P. LEGRAND (Editors) (1992): *Solar Terrestrial Variability and Global Change* (Interdivisional Commission on History of IAGA, Bremen-Roennebeck), pp. 243.

SIMKIN, T. and L. SIEBERT (1994): *Volcanoes of the World* (Geoscience Press, Tucson, Arizona), pp. 349.

SIMKIN, T., L. SIEBERT, L. MCCLELLAND, D. BRIDGE, C. NEWHALL and J.H. LATTER (1981): *Volcanoes of the World – A Regional Directory, Gazetteer, and Chronology of Volcanism During the Last 10000 Years* (Hutchinson Ross Publ. Co., Stroutsburg, Pennsylvania), pp. 233.

SINDONI, E. and A.Y. WONG (Editors) (1991): *Controlled Active Global Experiments (CAGE),* Proceedings of the International School of Plasma Physics «Piero Caldirola», Varenna (Como-Italy), September 5-12, 1989 (Editrice Compositori, Bologna), pp. 371.

STEPHENSON, F.R. and A.W. WOLFENDALE (Editors) (1988): *Secular Solar and Geomagnetic Variations in the Last 10000 Years* (Kluwer Academic Publ., Dordrecht, etc.), pp. 510.

STUIVER, M. and P.D. QUAY (1980): Changes in atmospheric Carbon-14 attributed to a variable Sun, *Science*, **207** (4426), 11-29.

STUIVER, M. and B. BECKER (1986): High precision decadal calibration of the radiocarbon time scale, A.D. 1950-2500 B.C., *Radiocarbon*, 863-911.

STUIVER, M. and T.F. BRAZIUNAS (1988): The solar component of the atmospheric ^{14}C record, in *Stephenson and Wolfendale (1988)*, 245-266.

STUIVER, M. and T.F. BRAZIUNAS (1989): Atmospheric ^{14}C and century-scale solar oscillations, *Nature, Lond.*, **338**, 405-408.

TOMADIN, L., D. WAGENBACH and V. LANDUZZI (1996): Mineralogy and source of high altitude glacial deposits in the Western Alps: clay minerals as Saharan dust tracers, in *Guerzoni and Chester (1996)*, 223-232.

TUCKEY, J.W. (1977): *Exploratory Data Analysis* (Addison-Wesley Publ. Co., Reading, Massachusetts, etc.) pp. 688.

TURCOTTE, D.L. (1992): *Fractals and chaos in geology and geophysics* (Cambridge Univ. Press.), pp. 221.