92-194

сообщения
объединенного
института
ядерных
исследований
дубна

E18-92-194

V.B.Zlokazov, S.P.Tretyakova

MATHEMATICAL ANALYSIS OF TEMPORAL SPECTRA OF SOIL RADON EMANATION

1992

Physical methods of measurements of the radon concentration in soil, water and ground air in some parts of Ashkhabad seismic area are described in [Ashirov T.A. et al.,1986]. At the time being many earth scientists are certain that there is a temporal connection between radon concentration anomalies and seismic events.

The purpose of the present work is the elaboration of a mathematical method for the analysis of recorded radon emission data and the detection of its information indications, which could be correlated with significant seismic events, and the application of this method to some concrete data, in particular, to radon concentration in the ground air of an area in Kopetdag Peredowoy Razlom (Central Asia).

Since some years ago a number of papers has appeared, covering questions of earthquake prediction by such precursory effects as radon, uranium etc. [Ashirov T.A. et al.,1986, Borrodall S. and Enriekson S.,1980,1980,Varshal G.M.,1985]; however, a problem of creation of an efficient method for such prediction still remains at the beginning of its solution.

The problem is complicated by two circumstances:

1/ the mechanism of radon anomalies is not sufficiently clear, and, therefore, the causal connection between them and, say, earthquakes is not clear either;

2/ the system is a many factor one, and extraction of information indications from this distribution, i.e., its parts and details, which could be matched with strong earthquakes by, at least, correlation connections, is itself a complex mathematical problem.

This paper is intended to overcome somehow the difficulties pointed out. Since building an adequate physical model is not possible now, it seems that the most appropriate way to tackle with such problems is the development of formal methods for decoding information structure of distributions.

It is clear from the general considerations, that information indications are either characteristic details, concentrated within some intervals of argument values, or characteristic details of analogous or gomologous form, repeated along the whole distribution.

1. The decomposition of distributions

A mathematical basis of the formal analysis of the simplest information structure of distributions is the additive decomposition

[Zlokazov V.B.,1985], or the extraction of components from them. Shortly, it can be described as follows.

Problems of the function decomposition can be systematized in a framework of a following formalism. Let us consider an expression

$$f(t) = \sum_{i=1}^{k} f_i(t) + e(t), \tag{1}$$

where f_i and e are functions, mutually linear-independent. The function $e(t)$ is singled out, called the error of the function f, and the maximum of its amplitude at each point is t assumed to be known.

The semantic interpretation of the functions f_i is as follows: they are factors, determining the structure of the summary distribution f and part of them is linked with events of interest by a causal connection.

Under qualitative decomposition a mapping

$$f \to (n,[M_i]),\ i=1,2,\ldots,n$$

is meant, where M_i are disjoint function classes such, that $f_i \in M_i$ for some indices i.

A function class M_i is called the model of component f_i.

Under quantitative decomposition a mapping

$$f \to (n,[g_i(x)]),\ i=1,\ldots,n,$$

is meant, where $g_j \in M_j$.

Classes M_i describe possible types of formal factors, which can determine the structure of a distribution, and items g_i are concrete values of these factors.

One can point out the following modifications, which exhaust the concept of decomposition: a factor decomposition and classification ones (discriminant and clusterization).

All this is the operation which allows us to solve our problem: automatically recognize the additive structure of functions and distributions and extraction components from them, which are important to us in informational or classificational sense.

The next step is the formalization of a concept of a component. This can be done on the basis of the practice of the analysis of real experimental data.

Generalizing a concept of a function component, extracted by a visual analysis of function graphs, one can formalize the following types of components:

1/ almost-concentrated ones, i.e. functions, integrable at a given degree over all the axis, i.e. such that for any ε and a given positive n there exists a region A, $A \in R$, such, that

$$\left| \int_A | f(t) |^n dt - \int_R | f(t) |^n dt \right| < \varepsilon;$$

2/ almost-periodical ones, i.e. such that for any $\varepsilon > 0$ there exists a positive number L such that any interval of length L contains at least one T such that for any t

$$| f(t+T) - f(t) | < \varepsilon;$$

holds;

3/ distributed components, i.e. functions non-integrable over the total axis.

REMARK 1. For almost-concentrated components the region A is called integral (δ,n)-support of the function, and for almost-periodical ones the number T is called the almost-period.

REMARK 2. Finite functions are typical instances of almost-concentrated functions. Their supports for any δ and n are the regions of their definition; as instances of almost-periodical functions one can take linear combinations of continuous periodical functions with incommensurate periods, or simply a linear combination of continuous periodical functions.

An arbitrary function can contain any combination of above-mentioned components.

These fundamental components represent the most universal typification of elements of functions of an additive structure, by which distributions, containing experimental information, are usually described.

A multiplicative structure of functions can be reduced to the additive one with the help of logarithmic transformation.

Since for almost-concentrated components moments are defined:

$$mi = \int_A (t-ti)^i f(t)dt,$$

this gives us a possibility to describe such functions using the following characteristics:

1/ m_0, or, function area, at $i = 0$;

2/ t_c, or a center of a function; this is a root of equation $m_1 = 0$ at $i=1$;

3/ m_2/m_0, or width of a function; at $i = 2$ and $t_2 = t_1$.

Let us denote these characteristics by letters

S, P, W.

Almost-periodical functions are characterized by the following quantities:

1/ Amplitude, or the maximum value of a function;

2/ Almost-period (or simply period).

We denote them by letters:

A, T.

Now we can formulate our problem as follows. Let be given:

$f(t)$ = a distribution of recorded radon yields along a time interval (t_0,t_1);

$z(t)$ = function of strong seismic events, not equal to zero at some points t_i $(t_0 < t_i < t_1)$, and equal to zero elsewhere;

$u(t)$ = a continuous function of distributed type;

It is required:

1/ extract almost-periodical components from the distribution $f(t)$;

2/ extract almost-concentrated components $p_k(t)$ from $f(t)$;

3/ determine correlations between $z(t_i)$ and components $p_k(t)$.

4/ determine correlations between $z(t)$ and function $y(t)$ or their parts without extracting these parts.

It can be that a structure of $f(t)$ is not explicit. In such a case one needs to find a transformation Q such that the function $d(t) = Q*f(t)$ will have a more pronounced structure and then operations 1/-3/ should be performed over $d(t)$.

The structure-enhancing transformations are.

1/ First of all one can use such simple function transformations as convolution or linear filters of smoothing type

$$g(x) = \int K(x,t)f(t)dt$$

with corresponding weight function $K(x,t)$;

or nonlinear filters, such as, e.g., autocorrelation:

$$g(x) = \int f(t)\ f(t+x)\ dt.$$

However, it should be noted that smoothing, based on principles of a frequency filtration, is not relevant for our case, because in physical curves information has, as a rule, an amplitude-frequency coding.

In this case we can recommend filters of variational types, which possess important peak-preserving properties, described in [Zlokazov V.B.,1985].

A smoothed function $g(t)$ is obtained from $f(t)$ by minimizing

$$\sum_t w(t)g''(t)^2 + \sum_t (f(t) - g(t))^2,$$

where t is a discrete variable, g'' is the 2 derivative (difference) of the function $g(t)$, and $w(t)$ is weight function, equal to

$$w(t) = \frac{1}{(1 + g_0'(t)^2)}.$$

Here g_0' is the 1st derivative (difference) of some a priori estimate of $g(t)$.

This filter efficiently suppresses the noise component, but preserves peaks, both weak and strong ones.

2/ The information can be contained in the dynamics of changes of $f(t)$ and then the derivative $f'(t)$ or its modulus should be analyzed.

$$g(t) = \hat{R} \cdot f(t),$$

where $\hat{R}$ is a discrete analog of the differentiation operator.

The discrete derivatives can be computed by one of the following formulae

a/ $g' = \mathrm{MAX}(-f',0)$

b/ $g' = \mathrm{MAX}(f',0)$

c/ $g' = |\ f'\ |;$

Every formula produces its own patterns.

3/ The following transformation is peak-amplifying

$$g(t) = \frac{g''(t)}{(1 + g'(t)^2)^{.5}}$$

and, in addition, it suppresses a continuous component, which, as a rule, does not contain information.

Method, used for search for periodicities, both harmonic and anharmonic ones, is described in [Zlokazov V.B.,1990] in detail.

An ideal instance of a concentrated function is a peak - non-negative function with one maximum, not equal to zero within a bounded interval $[t_1,t_2]$.

Surely, an arbitrary distribution can have concentrated components of the most different types. However, for an automatic search and analysis of them it is expedient to apply some transformation to the distribution, which transfer these components into peaks. These are various combinations of smoothing and differentiating filters, mentioned above.

Method, used for the search of peaks, is described in [Serebryannikov M.G. & Pervozvansky A.A.,1965]. The work of its algorithm gives us estimates of peak characteristics, which in our case are above-mentioned function moments.

Let a seismic event of interest be characterized by a pair of numbers: amplitude A and time t, and let be given a set of pairs $[A_i,t_i]$, i=1,2,...,L and the function f(t), whose information structure having been analyzed, the results were: the function contains:
1/ a set of periodical components $[w_j(t),B_j,T_j]$, j=1,2,..,m;
2 a set of concentrated components - peaks $[p_k(t),S_k,P_k,W_k]$, k=1,..,n;
and a distributed component b(t), which is of no interest to us.
Here T_j are periods of components w_j, and S_k,P_k,W_k are areas, positions and halfwidths of peaks.

Our final problems then are formulated as follows:

1/ test for common periodicities of A(t) and f(t);
2/ test for dependences between pairs $[A_i,t_i]$ and threes $[S_k,P_k,W_k]$, in particular, for presence of a common time shift between $[t_i]$ and $[P_k]$, i.e. determine whether there exists τ with a small error δ such that for any t_i there is a P_i, satisfying the relation

$$|P_i - t_i| < \tau \pm \delta.$$

3/ test whether there exist dependences between parts of z(t) and y(t) without extracting these parts.

The first two problems can be solved by such techniques as those of cluster-analysis.
The third problem is a complex problem. To solve it an approach was used called a generalized fitting. It consists in the following.
We have a problem: functions

$$y(x) = y_1(x) + y_2(x) + e_1(x),$$
$$f(x) = f_1(x) + f_2(x) + e_2(x)$$

being given, where y_1,y_2,f_1,f_2 are components and e_1,e_2 are errors, one needs to find out whether component, say, f_1 is connected with the component, say, y by a simply recognizable dependence.

A method for constructing such dependences can be formed using the method of approximation described in [Zlokazov V.B.,Comp.Phys.Comm.,1989 ,v.54]. Namely, if $f_1(x)$ is a result of some continuous transformation of a function $y_1(x)$, then we can define $f_1(x)$ as the rough model of the function $y_1(x)$, and approximate the latter by the former as follows:

$$y_1 = [\ Q_k(x) f_1(P_n(x))\ ], \quad \text{where}$$

Q_k and P_n are polynomials of degrees k and n, respectively, or other simple approximators.
In the simplex case we can write:

$$y_1 = a \cdot f_1\left(\frac{x-p}{c\cdot x+w}\right) , \qquad (2)$$

or Q_0 = a, and P_1 is ratio of linear polynomials. The parameters a,p,w,c have a clear physical meaning: amplitude, shift, and broadening.
The next step is: how one can make the approximation (2) without having explicitly extracted components of f_1 and y_1:
The approach used was as follows.
We use f and y instead f_1 and y_1:

$$y = a \cdot f\left(\frac{x-p}{c\cdot x+w}\right) ,$$

But then our goal is: find metrics ρ such that the approximation y by f in this metrics is given by an element which in fact is y_1 as function of f_1 of the type (2). One can suggest such a metrics.
Let P denote the parameter vector and be given a vector P_0, which is a priori estimates of P. Let's take a quadratic expression

$$\sum_x w(x,P_0)[\ y(x) - f(x,P)\ ]^2 , \qquad (8)$$

where

$$w(x,P_0) = \begin{cases} 1/||e(x)||^2, & \text{if } |h(x)| < c \\ (1+\beta)/[||e(x)||^2((h(x)/c)^2+\beta)], & \text{otherwise.} \end{cases} \qquad (9)$$

Here $h(x) = y(x) - f(x,P_0)$, c and β are given constants.
One can point out some practically important particular cases of filtration, when approximation in such a metrics indeed provides fitting of only $y_1(x)$ by $f_1(x)$:

a/ a distributed function from its sum with almost-concentrated, if the union of the integral supports of the latter does not cover the total region, at which the former is defined;

b/ almost-concentrated from its sum with other almost-concentrated, if the union of the integral supports of the latter does not cover the support of the former.

Minimization of (6) and (7) is implemented with the help of an iterational procedure. Using it one can either fix the initial $w(x,P_0)$ or change it at every step, counting values of parameters at the previous step for their a priori estimates.
The coefficient c characterizes a priori estimate of the region of possible values of quantities h(x), and if $|h| \le c$ holds for all the x, the estimates (2) and (3) are usual L.S.-estimates; for those x for which $|h| > c$ holds norms of e(x) are recalculated approximately by $h(x)/c$ times.
The coefficient β is for the control: it regulates the extent of the effect of the norm recalculation: at β=0 its effect is maximum, whereas at $\beta = \infty$ recalculation vanishes at all, and the minimum of (2) and (3) is given by the conventional L.S.-estimates, irrespectively of the values of h(x).
One can show that a number of conditions being satisfied the procedure will give the required solution with an admissible accuracy [Zlokazov V.B.,Comp.Phys.Comm.,1989,v.54].

2. Application of the method

The described method was applied to the analysis of radon emission and atmospheric data from Ashkhabad region within a radius of 600 km. Earthquakes of the class above 10.5 were taken as seismic events. Such amounted to 21. Let E denote the set of their t_i

11 145 192 259 306 497 744 760 761 762 768 775 786 802
822 824 837 908 912 981 1054.

The following distributions were analyzed for the possible connection

of their structure with given event:

1/ R(t) = time distribution of daily records of soil air radon concentration;

2/ A(t) = time distribution of daily records of atmospheric pressure;

3/ S(t) = simulated data, which were sum of components:

a/ anharmonic periodicities

$$(f(t)=A\cdot\max(\sin(kt),0)^4)$$

with periods 365, 29, 7;

b/ harmonic periodicity with the period 90;

c/ peak-like functions, a constant one, and random noise.

Peaks were distant from events tested by 6-7 days (with a certain dispersion); 4 peaks had no relations to the events.

The length of the distribution was approximately equal to 1500, $0 < t < 1509$. The data was incomplete, the information for some intervals being missing. These were constructed artificially by interpolation method; while peak search and correlation analysis hypotheses were tested for their significance.

The results were as follows.
Details of the periodicity analysis method are given in [Zlokazov V.B., Comp.Phys.Comm.,1989,v.54]. Here we adduce only final conclusions.

Periodicities

1/ Radon data

a/ The annual periodicity (378 ± 347) is seen, though not very distinctly, and all its anharmonic "tails" are matched by the corresponding periods in the table. This proves that this periodicity has a distinct anharmonic character.

b/ The periodicity (30 ± 1) is very plausible and its anharmonic "tails" have the corresponding matches in the table; this also proves its anharmonic character.

c/ Rather plausible is a hypothesis about the periodicity (23 ± 1).

2/ Atmospheric pressure

a/ The annual periodicity 365 ± 104 is seen very well;

b/ Very plausible are hypotheses about the anharmonic periodicities 98 ± 20, 29 ± 2.

3/ Simulated data.
All periodicities are seen very well.

Peak-like components

Then the distributions, their 1st derivatives and quasicurvatures were analyzed for concentrated (peak-like) components with amplitudes, superior to some given level. This level was chosen each time so, that some almost equal number of peaks was found, and, thus, the probability that a peak is informational was equal for all the cases. Minimal characteristic time shifts between seismic events and peaks were sought for. The following peaks were detected.

1/ In R(t):

31peak, Sens = 1.00 Res = 1.00

87	110	118	145	184	202	226	235	265	289
331	349	357	384	410	422	518	523	534	567
579	593	674	769	789	804	818	864	880	895
936	944	996	1090						

The comparison with the set E showed:

a/ Positive time shift is in the interval (0 - 94), and the most probable value is equal to 49;
clusters detected are: 30 ± 16 and 91 ± 10.

b/ Negative time shift is in the interval (1 - 76), and the most probable value is equal to 38;
cluster detected is: 39 ± 12.

In R(t)' there were detected following peaks:

36 peaks Sens = 3.25 Res = 1.00

38	67	82	114	145	166	182	196	202	224
237	299	318	348	355	379	390	470	521	566
577	591	623	635	724	747	763	787	802	816
834	848	876	895	979	995	1022	1050	1095	

One can see:

a/ Positive time shift is in the interval (0 - 31), and the most probable value is equal to 15;
cluster detected is: 27 ± 9.

b/ Negative time shift is in the interval (1 - 71), and the most probable value is equal to 39;
cluster detected is: 31 ± 18.

In the quasicurvatures R(t) there were following peaks detected:

37 peaks Sens = 5.25 Res = 1.00

37	47	132	145	202	217	230	238	316	349
376	394	413	418	429	452	468	489	502	509
520	544	621	651	677	731	755	762	776	784
798	823	832	895	927	932	977	992	1000	1042

One can see:

a/ Positive time shift is in the interval (0 - 68), and the most probable value is equal to 29;
clusters detected are: 9 ± 4 and 28 ± 3.

b/ Negative time shift is in the interval (0 - 58), and the most probable value is equal to 26;
cluster detected is: 21 ± 13.

2/ A(t) was analyzed to detect components (continuous or discrete), connected with R(t) by a shift relation.

a/ A filtration regression [Zlokazov V.B.,1985] of R(t) on A(t) was carried out; the quality of filtration fitting, characterized by the ratio of chi-square and number of degrees of freedom, being equal approximately to 4, the time shift of R(t) to A(t) was approximately equal to 3 (delay).

b/ The set of peaks of A(t)' was compared with the set of peaks of R(t)'; The shift interval was equal to 0 - 170; only 9 out of 55 belong to a range (0 - 5), which is comparable with estimates 3, obtained from the continuous analysis.

3/ The purpose of the simulated data analysis was intended to test the method, and it showed a good quality of the results. All the peaks with amplitude, superior to the 0.5 of the background level, were found, and the distances between them and seismic events did not exceed the confidence interval.

7. Conclusion

The results of the ERTQU analysis of daily record of the radon concentration for the period of 1984 - 1986 allowed us to make the following conclusions:

1/ In the oscillation of this concentration annual and monthly periodicities can be observed with a significant amplitude, and these periodicities are obviously anharmonic.

2/ The 1st derivative and quasi-curvature are richer than initial data of local events, which can be correlated with strong seismic events; the 1st derivative is more informative for prediction purposes.

3/ Results of the analysis confirmed that the method is efficient enough.

References

1. Ashirov T.A. et al.
JINR 18-86-452, Dubna, 1986.
2. Borrodall S., Enriekson S.
Geophysics, 1980, v.45, No 3, (part 1).
Geophysics, 1980, v.45, No 11 (part 2).
3. Zlokazov V.B.
Mathematical methods for the analysis of experimental spectra and spectrum-like distribution. FECHAYA, v.16,5,p.1126-1163.
4. Serebryannikov M.G., Pervozvansky A.A.
The detection of hidden periodicities. Moscow, Nauka, 1965.
5. Hydro-geochemical predictors of earthquakes.
Varshal G.M. Moscow, Nauka, 1985.
6. Zlokazov V.B.
Computer Physics Communications,1989,v.54. 371-379.
7. Zlokazov V.B.
Computer Physics Communications,1990,v.59,p.217-25.

Received by Publishing Department
on May 5, 1992.

Злоказов В.Б., Третьякова С.П.

E18-92-194

Математический анализ временных спектров подпочвенного радона

Описан математический подход по выявлению информационных признаков в распределениях регистрируемых данных и корреляционной увязке их с событиями, представляющими интерес. Основой анализа таких распределений является их декомпозиция — разложение на информационные компоненты. Постулированы следующие типы компонент произвольного распределения: 1) сосредоточенные; 2) периодические; 3) распределенные. Описана математическая методика выявления этих компонент, оценки их характеристик и кластеризационные методы установления корреляционной связи между ними и заданными событиями. Созданная методика и реализующая ее фортранная программа ERTQU была применена при анализе временных распределений концентрации радона в воде глубинных скважин участка зоны Передового разлома Копетдага.

Работа выполнена в Лаборатории ядерных реакций ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1992

Zlokazov V.B., Tretyakova S.P.

E18-92-194

Mathematical Analysis of Temporal Spectra of Soil Radon Emanation

An approach is described for the detection of information indications in experimental distributions and estimation of their correlation with events of interest. The basis of the analysis of such distributions is their decomposition into informational components. Three types of components are postulated: 1) concentrated, 2) periodical, 3) distributed. The mathematical method is described for the detection of these components, estimation of their characteristics and cluster methods of determination of a correlation between them and the events of interest. The method and a corresponding program, called ERTQU, was applied to the analysis of soil radon concentration and atmospheric data.

The investigation has been performed at the Laboratory of Nuclear Reactions, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1992