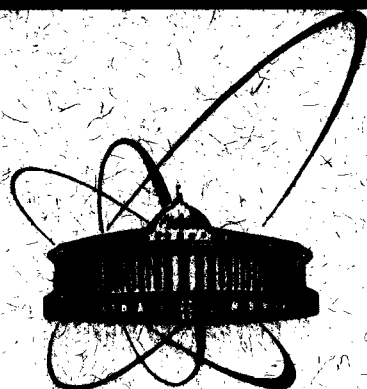


89-659



СООБЩЕНИЯ
ОБЪЕДИНЕННОГО
ИНСТИТУТА
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ
ДУБНА

3681

P11-89-659

В.Б. Злоказов, Д. Ишанкулиев*,
С.П. Трегьякова, Т.А. Аширов*

МАТЕМАТИЧЕСКИЙ ПОДХОД
К АНАЛИЗУ ВРЕМЕННЫХ СПЕКТРОВ
ПОЧВЕННОГО РАДОНА

* Институт сейсмологии АН ТССР, Ашхабад.

1989

Физические методы регистрации изменения концентрации радона в подземных водах и почвенном воздухе в некоторых участках Ашхабадской сейсмоактивной зоны описаны в работе^{1/}. В настоящее время общепринятой является точка зрения, согласно которой существует временная связь между аномалиями изменений концентрации радона и сейсмическими событиями. Цель настоящей работы — разработка математической методики анализа регистрируемых распределений радонных эманаций и выявления в них информационных признаков, связанных с крупными сейсмическими событиями, и применение этой методики к некоторым конкретным данным, в частности к измерениям концентрации радона в атмосфере подпочв участка зоны Передового разлома Копетдага.

В настоящее время имеется ряд публикаций, посвященных вопросам прогнозирования землетрясений по ряду предвестников (радон, уран и т.д.)^{1, 2, 5/}, но проблема создания эффективной методики такого прогнозирования находится, безусловно, в самой начальной стадии своего решения. Ситуация осложнена очень существенно двумя обстоятельствами:

1) механизм появления аномальных концентраций радона пока недостаточно ясен, и, следовательно, неясна причинно-следственная связь между ними и сейсмическими событиями;

2) система является исключительно многофакторной, и выделение в общем зарегистрированном распределении радоновых эманаций информационных признаков, то есть частей и деталей этого распределения, допускающих хотя бы корреляционную увязку с сейсмическими событиями, само по себе является очень сложной математической задачей.

Данная работа и представляет собой попытку как-то преодолеть указанные трудности. С учетом того факта, что построение адекватной физической модели затруднительно, упор был сделан на развитие идеологии формального анализа распределений и формальной методологии расшифровки их информационной структуры.

1. ДЕКОМПОЗИЦИЯ РАСПРЕДЕЛЕНИЙ

Математической основой формального анализа простейшей информационной структуры распределений является аддитивная декомпозиция^{13/}. Суть ее может быть изложена следующим образом. Задачи де-

композиции функций могут быть систематизированы в рамках следующего формализма. Пусть

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

где f_i и e — функции, взаимно линейно независимые. Функция e выделяется особо, называется погрешностью функции f , и максимум ее модуля в каждой точке t предполагается известным.

Содержательная интерпретация функций f_i состоит в том, что они являются факторами, определяющими структуру суммарного распределения f , и часть из них связана с событиями, представляющими интерес, причинной связью.

Под качественной декомпозицией распределения f будем понимать отображение

$$f \rightarrow (n, \{M_i\}), i = 1, 2, \dots, n,$$

где M_i — непересекающиеся классы функций из некоторого множества функциональных классов, такие, что $f_i \in M_i$ для некоторых индексов i (в частности, множество таких индексов может быть и пустым).

Под количественной декомпозицией f мы будем понимать отображение

$$f \rightarrow (n, g_i), i = 1, 2, \dots, n,$$

где $g_i \in M_i$.

Содержательная интерпретация классов M_i состоит в том, что с помощью их элементов g_i можно получить представление суммарной функции f , а сами g_i могут рассматриваться как факторы, часть которых связана с сейсмическими событиями связью корреляционного типа.

Декомпозиция будет оптимальной в смысле какого-либо критерия групповой близости, если элементы g_i наиболее близки в смысле этого критерия к f_i .

Можно указать следующие модификации, исчерпывающие понятие декомпозиции: факторная декомпозиция и классификационная (дискриминантная и кластеризационная).

1. Задача факторного декомпозиционного анализа

Пусть функции $g_j, j=1,2,\dots,l, l \leq n$, таковы, что (1) допускает представление

$$f(t) = \sum_{j=1}^l g_j(f_1, \dots, f_n) + e(t).$$

Декомпозиция f на g_j является одновременно и факторизацией данной функции.

2. Задача декомпозиционной классификации.

Дискриминантный подход

Пусть задано (1) и пусть задана совокупность функциональных классов (S_j) . Нахождение функциональных элементов $g_j, g_j \in S_j, j=1, 2, \dots, m$, таких, что имеет место

$$f(t) = \sum_{j=1}^m g_j(t) + e(t),$$

соответствует количественной фазе задачи декомпозиционной классификации.

3. Задача декомпозиционной классификации. Кластерный подход

Пусть по-прежнему имеет место (1) и пусть задан критерий $r(S_i, S_j)$ — расстояние между произвольными классами функций S_i и S_j . Введем такое определение: для заданного ϵ будем считать ϵ -различимыми классы S_i и S_j , если $r(S_i, S_j) \geq \epsilon$. С помощью этого определения можно сформулировать следующие две разновидности задачи декомпозиционной кластерной классификации.

Пусть задана некоторая широкая совокупность классов функций B .

1. Для заданного ϵ найти минимальную совокупность ϵ -различимых классов функций $S_j \in B, j = 1, 2, \dots, l$, таких, что

$$f(t) = \sum_{j=1}^l g_j(t) + e(t),$$

где $g_j(t) \in S_j$.

2. Для заданного n найти n максимально различных классов функций $S_j \in B$ так, что

$$f(t) = \sum_{j=1}^n g_j(t) + e(t),$$

где $g_j(t) \in S_j$, и $(S_j) = \text{ARG MAX}_{i,k} \text{MIN}_{i,k} r(S_i, S_k)$.

Декомпозиция и является операцией, позволяющей решить поставленную нами задачу: автоматическое распознавание аддитивной структуры функций и распределений: выделение в них компонент, важных для нас в информационном или классификационном смысле.

Если обобщить понятие функциональной компоненты, выделяемой при визуальном анализе графиков функций, то можно формализовать следующие типы компонент:

1) почти сосредоточенные, то есть функции, интегрируемые в заданной степени на всей прямой, иначе, такие, что для любого ϵ и заданного положительного n существует конечная область A , $A \in R$, такая, что

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

2) почти периодические функции, то есть такие, что для любого $\epsilon > 0$ существует положительное число L такое, что в любом интервале длины L найдется хотя бы одно T такое, что при любом t имеет место

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

3) распределенные компоненты, то есть неинтегрируемые на всей оси функции. В произвольной функции могут присутствовать любые комбинации компонент перечисленного типа. Для почти сосредоточенных компонент область A будем называть интегральным носителем функции. Так как для таких функций имеют смысл моменты

$$m_i = \int_A (t - t_i)^i f(t) dt,$$

то это дает нам возможность весьма полно описывать такие функции с помощью следующих характеристик:

- 1) m_0 , или площадь функции, при $i = 0$;
- 2) t_1 , или центр функции; это корень уравнения $m = 0$ при $i = 1$;
- 3) m_2/m_0 , или ширина функции; при $i = 2$ и $t_2 = t_1$.

Обозначим эти характеристики буквами

S, P, W .

Почти периодические функции характеризуются следующими величинами:

- 1) амплитуда, или максимальное значение функции;
- 2) почти период (или просто период).

Их будем обозначать буквами

A, T .

Эти фундаментальные компоненты представляют собой наиболее универсальную типизацию элементов функций аддитивной структуры, которыми обычно описываются распределения, информационные с точки зрения экспериментальных наук. Следовательно, наша задача будет состоять в следующем.

Пусть заданы:

$f = f(t)$ — распределение зарегистрированных радонных эманаций вдоль временного промежутка (t_0, t_1) ;

$z = z(t)$ — функция сильных сейсмических событий, не равная в некоторых точках $t_i (t_0 < t_i < t_1)$ нулю и равная нулю в остальных.

Требуется:

1) выделить в данном распределении $f(t)$ почти периодические компоненты;

2) выделить в распределении $f(t)$ почти сосредоточенные компоненты $p_k(t)$;

3) установить соответствие корреляционного типа между величинами $z(t_i)$ и компонентами $p_k(t)$.

Может оказаться так, что структура у функции $f(t)$ явно не проявляется. В таком случае следует найти преобразование Q такое, что у функции $d(t) = Q \cdot f(t)$ структурные особенности выражены более ярко и тогда операции 1-3 следует проделать над функцией $d(f)$.

Для усиления структурных особенностей анализируемой функции можно применять следующие преобразования:

1. В первую очередь, разумеется, такие широко используемые простые преобразования функций, как конволюция или нелинейные фильтры сглаживающего типа

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

с соответствующей весовой функцией $K(x, t)$; или нелинейные фильтры, такие как, например, построение автокорреляционной функции:

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

Однако следует заметить, что сглаживание, основанное на принципах частотной фильтрации, не подходит для нашего случая, так как в физических кривых информация имеет, как правило, амплитудно-частотную кодировку.

Были использованы фильтры вариационного типа, обладающие замечательными пикосохраняющими свойствами, описанные в работе¹³. Сглаженная функция $g(t)$ получается из $f(t)$ минимизацией выражения

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

где t — дискретная переменная, g'' — 2-я производная (разность) функции $g(t)$, а $w(t)$ — весовая функция, равная

$$w(t) = \frac{1}{(1 + g'_0(t)^2)}.$$

Здесь g'_0 — 1-я производная (разность) некоторой априорной оценки $g(t)$. Этот фильтр эффективно подавляет шумовую компоненту, но сохраняет пики, как слабые, так и сильные.

2. Информация может оказаться заключенной в динамике изменения функции $f(t)$ и тогда анализировать следует производную $f'(t)$ или ее модуль

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

где \hat{R} — дискретный аналог оператора дифференцирования. При этом можно дискретную производную вычислять по следующим формулам:

$$g' = \text{MAX}(-f', 0), \quad g' = \text{MAX}(f', 0), \quad g' = |f'|.$$

Каждая формула порождает собственные образы. Желательно выбирать ту, которая дает образы, как можно более близкие к пикам.

3. Следующее преобразование является непосредственно пикоусиливающим:

$$g(t) = \frac{g''(t)}{\sqrt{(1 + g'(t)^2)}},$$

и оно дополнительно подавляет неинформационную пологую компоненту. Метод, использованный для поиска периодичностей, как гармонических, так и ангармонических, подробно описан в работе¹⁶. Идеальным примером сосредоточенной функции является пик — неотрицательная функция с одним максимумом, не равная нулю только на ограниченном

интервале $[t_1, t_2]$. Правда, произвольное распределение может иметь сосредоточенные компоненты самого разного вида. Но для их автоматического поиска и определения целесообразно применение к распределению различных преобразований, трансформирующих эти компоненты в пики. Это всевозможные комбинации сглаживающих и дифференцирующих фильтров, упомянутые выше. Метод, использованный для поиска пиков, описан в работе¹⁴.

Пусть интересующее нас сейсмическое событие характеризуется парой чисел: амплитудой A и временем t , и пусть задано множество пар $[A_i, t_i]$, $i = 1, \dots, L$, и функция $f(t)$; информационную структуру мы проанализировали и установили, что она содержит:

- 1) набор периодических компонент $[w_j(t), B_j, T_j]$, $j = 1, \dots, m$;
- 2) набор сосредоточенных компонент — пиков $[p_k(t), S_k, P_k, W_k]$, $k = 1, 2, \dots, n$.

Здесь T_j — периоды компонент w_j , а S_k, P_k, W_k — площади, положения и полуширины пиков.

Наша дальнейшая задача качественно формулируется так:

- 1) установить наличие общих периодичностей у $A(t)$ и $f(t)$;
- 2) установить зависимости между парами $[A_i, t_i]$ и тройками $[S_k, P_k, W_k]$, в частности наличие общего временного сдвига между $[t_i]$ и $[P_k]$, то есть определить, существует ли величина τ с небольшим разбросом δ , такая, что для любого t_i найдутся P_i , удовлетворяющие условию

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

2. ПРИМЕНЕНИЕ МЕТОДА

Описанная методика была применена к конкретным данным. В качестве сейсмических событий были взяты землетрясения класса не ниже 10,5 в области радиусом 600 км от Ашхабада. Таких оказалось 21. Обозначим через E множество их t_i :

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

Следующие распределения анализировались с целью определения возможной связи их структуры с данными событиями:

- 1) $R(t)$ — временное распределение суточных регистраций концентрации радона в атмосфере подпочв участка зоны Передового разлома Копетдага;

2) $A(t)$ — временное распределение суточных регистраций атмосферного давления в том же районе;

3) $S(t)$ — симулированные тестовые данные.

Симулировалась сумма следующих компонент:

а) ангармонических периодичностей ($f(t) = A \cdot \max(\sin(kt), 0)^4$) с периодами 365, 29, 7;

б) гармонической периодичности с периодом 90;

в) пикообразных функций, константы и случайного шума.

Пики отстояли от проверяемых сейсмических событий на 6-7 дней (с некоторым разбросом); 4 пика были добавлены вне связи с событиями.

Длина распределений приблизительно равна 1500 дней, $0 < t < 1509$. Эти данные неполные, в некоторых временных интервалах измерения не проводились. Недостающие числа были построены искусственно с помощью интерполяции, при поиске пиков и анализе корреляций проверялись и гипотезы о возможном нахождении пиков в данных интервалах.

Итоги анализа были следующие; детали анализа периодичностей для всех трех распределений подробно освещены в '6'. Здесь приводим лишь выводы.

1. Выводы о периодичностях распределения радона:

а) видна, хотя и очень нечетко, годовая периодичность (378 ± 347), и для всех ее ангармонических "хвостов" есть соответствующие кандидаты в таблице, что доказывает резко ангармонический характер этой периодичности;

б) явно видна тоже ангармоническая периодичность (30 ± 1);

в) можно предположить, хотя и с меньшей убедительностью, что самостоятельной является и периодичность (23 ± 1);

г) периодичности (16 ± 1) и (6 ± 1) являются, по-видимому, все же ангармоническими "хвостами" периодичностей 378 и 30.

2. Выводы о периодичностях распределения атмосферного давления:

а) хорошо просматривается годовая ангармоническая периодичность 365 ± 104 ;

б) весьма правдоподобно предположение о наличии ангармонических периодичностей 98 ± 20 , 29 ± 2 ; значительно менее убедительно предположение о периоде 20 ± 2 ; об остальных периодах сказать ничего нельзя.

3. Выводы о периодичностях симулированного распределения:

все периодичности хорошо видны, и их характер тоже поддается определению.

Затем в распределениях, их первых производных и квазикривизнах были выявлены сосредоточенные (пикообразные) компоненты с амплитудой не ниже некоторого условного уровня. Этот уровень выбирался

каждый раз так, чтобы выявлялось некоторое почти одинаковое количество пиков, и, таким образом, вероятность пика быть информационным была одинаковой для всех случаев. Искались минимальные характерные временные сдвиги между сейсмическими событиями и пиками.

1. В $R(t)$ были обнаружены следующие пики:

31 пик, 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						

Их сравнение с множеством E дало следующую картину:

а) опережение заключено в интервале (0 — 94), а наиболее вероятное значение равно 49; обнаружены кластеры: 30 ± 16 и 91 ± 10 ;

б) запаздывание заключено в интервале (1 — 76), а наиболее вероятное значение равно 38; обнаружен кластер: 39 ± 12 .

В $R(t)'$ были обнаружены следующие пики:

36 пик, 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	

При этом:

а) опережение заключено в интервале (0 — 31), а наиболее вероятное значение равно 15; обнаружен кластер: 27 ± 9 ;

б) запаздывание заключено в интервале (1 — 71), а наиболее вероятное значение равно 39; обнаружен кластер: 31 ± 18 .

В квазикривизнах от $R(t)$ были обнаружены следующие пики:

37 пик, 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

При этом:

а) опережение заключено в интервале (0 — 68), а наиболее вероятное значение равно 29; обнаружены кластеры: 9 ± 4 и 28 ± 3 ;

б) запаздывание заключено в интервале (0 — 58), а наиболее вероятное значение равно 26; обнаружен кластер: 21 ± 13 .

2. Анализ $A(t)$ велся с целью обнаружения компонент (непрерывных или дискретных), связанных с $R(t)$ сдвиговыми соотношениями.

а) была осуществлена фильтрационная регрессия $R(t)$ по $A(t)$. При качестве фильтрационной подгонки, характеризующемся χ^2 на степень свободы, равным приблизительно 4, временной сдвиг $R(t)$ относительно $A(t)$ равен приблизительно 3 дням (запаздывание);

б) в качестве множества событий были взяты пики у $A(t)$ и был проведен описанный выше анализ сравнения с пиками $R(t)$. Интервал сдвигов равен 0 — 170; при этом только 9 из 55 сдвигов заключены в промежутке (0 — 5 дней), сравнимом с оценкой (3 дня), полученной при непрерывном анализе.

3. Анализ пиков в симулированных данных ставил своей целью проверку работоспособности метода, и он дал требуемое качество результатов. Все информационные пики с амплитудой, сравнимой с 0,5 от уровня фона, были найдены, и расстояния от них до сейсмических событий не выходят из доверительного интервала.

3. ЗАКЛЮЧЕНИЕ

Изучение результатов применения программы ERTQU к ежедневным регистрациям концентрации почвенного радона зоны Передового разлома Копетдага в течение 1984-1986 гг. позволяет сделать следующие выводы:

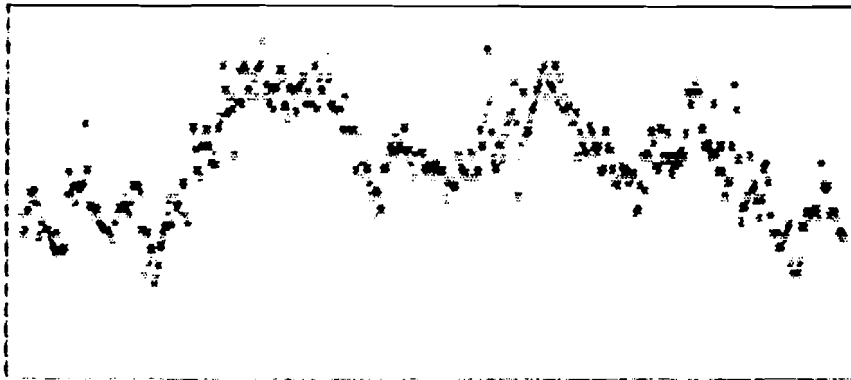


Рис. 1. Радоновое распределение $R(t)$ (компрессия с коэффициентом 4).

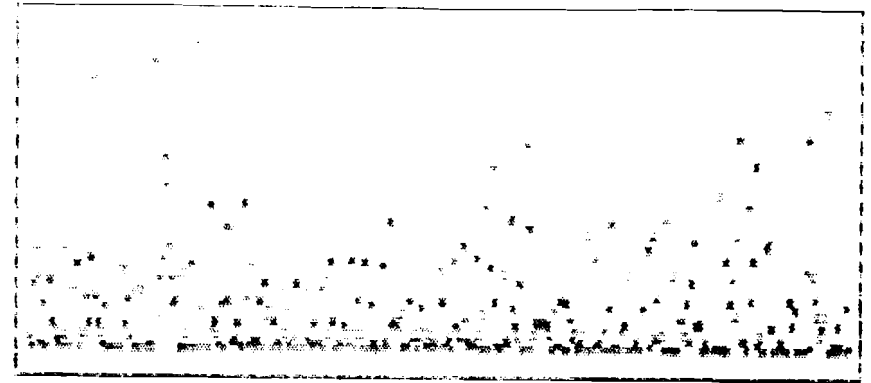


Рис. 2. Нормированная производная от $R(t)$.

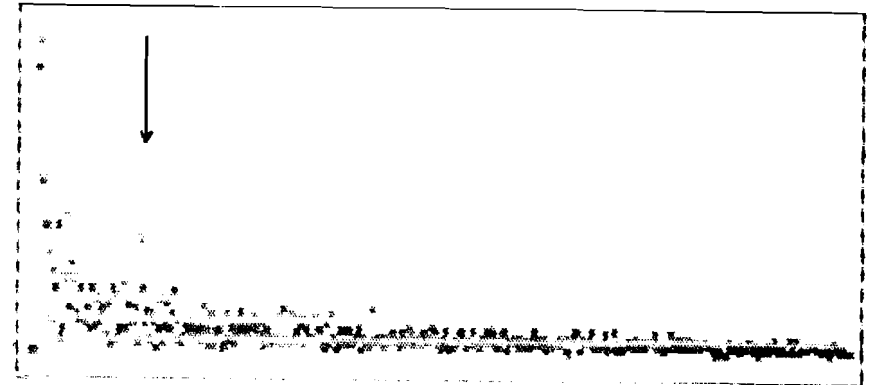


Рис. 3. Модуль преобразования Фурье $R(t)$ (первые 3 ординаты удалены, так как их значения достигают 800, тогда как остальные не превосходят 49). Стрелкой обозначена периодичность в 30 дней.

1) в колебаниях этой концентрации заметны годовая и месячная периодичности со значимой амплитудой, носящие явно ангармонический характер;

2) 1-я производная и квазикривизна гораздо богаче, чем исходные данные, локальными "событиями", допускающими корреляционную увязку с сильными землетрясениями; при этом большую информативность с точки зрения прогнозирования обнаружила 1-я производная;

3) результаты анализа фактического материала (см. рис. 1-3) показали работоспособность методики.

ЛИТЕРАТУРА

1. Аширов Т.А. и др. — ОИЯИ, 18-86-452, Дубна, 1986.
2. Borrodall S., Enriekson S. — Geophysics, 1980, v.45, No.3 (part 1); Geophysics, 1980, v.45, No.11 (part 2).
3. Злоказов В.Б. — ЭЧАЯ, 1985, т.16, вып.5, с.1126.
4. Серебрянников М.Г., Первозванский А.А. — Выявление скрытых периодичностей. М.: Наука, 1965.
5. Гидро-геохимические предвестники землетрясений (под ред. Г.М.Варшала). М.: Наука, 1985.
6. Злоказов В.Б. — ОИЯИ, P11-89-378, Дубна, 1989.

Рукопись поступила в издательский отдел
18 сентября 1989 года.

Злоказов В.Б. и др.

P11-89-659

Математический подход к анализу
временных спектров почвенного радона

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

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1989

Перевод авторов

Zloказov V.B. et al.

P11-89-659

Mathematical Method for the Analysis
of Time Spectra of Soil Radon

The mathematical approach is described, intended to a detection of informational indications in the distributions of recorded data and a correlation matching them to the events, which are of interest. The basis of the analysis of such distributions is their decomposition into information components. The following types of components of arbitrary distributions are postulated: 1) concentrated components; 2) periodical ones; 3) distributed ones. The mathematical method for the detection of these components and for the estimation of their characteristics, and the clusterization methods for the recognition of the correlation link between them and the given events are described. The method and implementation in Fortran program ERTQU were applied to the analysis of the time distributions of the concentration of radon in the air of subsols of a region in the Peradovoj Razlom of Kopetdag.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research, Dubna 1989