

УДК 519.946; 519.254

ХАОТИЧЕСКАЯ И СЛУЧАЙНАЯ СОСТАВЛЯЮЩИЕ В ПРИРОДНЫХ ВРЕМЕННЫХ ДАННЫХ

В.Г. БОНДАРЕНКО, К.К. ТРУСКОВСКИЙ

Для временных данных предложен метод определения соотношения детерминированной и стохастической составляющих. Для решения данной задачи выполнен ряд вычислительных экспериментов, использующих имитационное моделирование логистической хаотической последовательности и значений фрактального броуновского движения с различными показателями Харста — H . В полученной аддитивной смеси задается соотношение энергий детерминированной и случайной составляющих. Для больших значений показателя Харста хаотическое слагаемое оказывается более агрессивным: контрольные статистики смеси значимо отличаются от эталонных значений, соответствующих фрактальному броуновскому движению. Для малых значений H (антиперсистентный случай) наблюдается обратный результат. Рассмотренные примеры реальных временных данных описываются антиперсистентной моделью.

ВВЕДЕНИЕ

Анализ временных рядов — последовательных измерений x_1, \dots, x_n некоторой величины произвольной природы — состоит, прежде всего, в построении адекватной математической модели. Если эти данные априорно являются детерминированными, то есть члены последовательности $\{x_k\}$ представляются значениями некоторой динамической системы, то модель определяется уравнением

$$x_k = f(x_{k-1}, \dots, x_{k-p}),$$

описывающим эту систему. Если же временной ряд имеет стохастическую природу, то его модель определяется равенством

$$x_k = X\left(\frac{k}{n}\right),$$

где $X(t)$ — реализация некоторого случайного процесса $\xi(t)$, $t \in [0:1]$.

Реальные временные данные, описывающие природные явления, как правило, содержат и детерминированную (хаотическую), и случайную составляющую. Хаотичность означает крайне нерегулярное поведение последовательности. В первом приближении хаотичность можно определить

свойством системы разводить первично близкие траектории в ограниченной области фазового пространства. В данной работе предложен алгоритм оценивания доли каждой из этих составляющих в наблюдаемых временных рядах.

ПОСТАНОВКА ЗАДАЧИ

Предполагается, что наблюдаемые значения являются аддитивной смесью детерминированной хаотичной и случайной последовательностей:

$$x_k = \alpha u_k + \beta v_k,$$

где u_k — значения некоторой динамической системы, v_k — значения случайного процесса. Последовательности $\{u_k\}$, $\{v_k\}$ нормированы по энергии, то есть

$$\frac{1}{n} \sum u_k^2 = \frac{1}{n} \sum v_k^2 = 1.$$

Тогда величина $a = \frac{\beta}{\alpha}$ определяет долю стохастичности в наблюдаемых данных.

В качестве модели для хаотической составляющей выберем логистическую динамическую систему:

$$u_k = 4u_{k-1}(1 - u_{k-1}), \quad (1)$$

решения которой образуют хаотическую последовательность.

Для модели стохастической составляющей положим

$$v_k = \sigma B_H\left(\frac{k}{n}\right), \quad (2)$$

где $B_H(t)$ — фрактальное броуновское движение — гауссов случайный процесс, определенный соотношениями:

$$E B_H(t) = 0, \quad B_H(0) = 0, \quad E[B_H(t) B_H(s)] = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}).$$

Свойства моделей, определенных равенствами (1) и (2), подробно изучены в работах [1] и [2]. Выбор именно этих моделей объясняется их универсальностью. Так, fBm (fractional Brownian motion — фрактальное броуновское движение) применимо к описанию широкого класса природных явлений [3].

Цель работы — определение соотношения a стохастической и детерминированной составляющих в реальных наблюдаемых данных x_1, \dots, x_n в предположении, что они образуют аддитивную смесь

$$x_k = u_k + a v_k, \quad (3)$$

где последовательности u_k и v_k определены соотношениями (1) и (2).

ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

Приведем некоторые результаты, изложенные в работе [4] о статистической обработке временного ряда v_1, \dots, v_n , $\bar{v} = \frac{1}{n} \sum v_k = 0$, описываемого моделью (2). Параметры модели σ (волатильность) и H (параметр Харста) предполагаются неизвестными. Алгоритм их оценивания состоит в следующем. Рассмотрим статистику

$$Q(H) = \frac{0,8}{R_1} \sqrt{\frac{(S_H^{-1}y, y)}{n}},$$

где вектор $y = (y_1, \dots, y_n)$, $y_k = v_k - v_{k-1}$, $k = 1, \dots, n$, $R_1 = \overline{|y|} = \frac{1}{n} \sum |y_k|$, S — корреляционная матрица вектора y с элементами

$$\rho(y_j, y_k) = \frac{1}{2} (|k - j + 1|^{2H} + |k - j - 1|^{2H} - 2|k - j|^{2H}).$$

Значения Q вычисляются перебором H (в матрице S_H^{-1}) с некоторым шагом: то значение H , при котором $|Q - 1| \rightarrow \min$ объявляется оценкой \hat{H} параметра Харста. Оценка волатильности определяется формулой

$$\hat{\sigma} = 1,25 R_1 n^{\hat{H}}.$$

Эффективность алгоритма оценивания параметра Харста проверена в [4]; значения fBm получены методом имитационного моделирования.

Для произвольного априорно стохастического временного ряда x_1, \dots, x_n , ($\bar{x} = 0$) в работе [5] предложен алгоритм его аппроксимации fBm. Идея алгоритма — представление наблюдаемой траектории $x(t)$ в виде

$$x(t) = \Phi(\sigma B_H(\cdot))(t), \quad x_k = x\left(\frac{k}{n}\right),$$

где Φ — некоторое преобразование в пространстве траекторий. Реализация алгоритма базируется на значении статистики

$$d(y) = \frac{R_1^2}{R_2},$$

где $R_2 = \overline{y^2} = \frac{1}{n} \sum y_k^2$.

Если $d(y)$ значительно отличается от $\frac{2}{\pi} = 0,637$ (случай гауссовых приращений), перейдем к новой последовательности

$$z_k = \text{sgn } y_k |y_k|^{\frac{1}{\lambda}}, \quad \lambda > 0, \tag{4}$$

где λ определяется из уравнения $d = \frac{1}{\sqrt{\pi}} \frac{\Gamma^2\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\lambda + \frac{1}{2}\right)}$, тогда $d(z) = \frac{2}{\pi}$.

Таким образом, исходные данные приведены к виду

$$x_k = \sum_{j=1}^k \operatorname{sgn} y_j |z_j|^\lambda, \quad (5)$$

где z_1, \dots, z_n предполагаются приращениями fBm.

Последнее предположение представляет собой статистическую гипотезу T и нуждается в проверке. Если гипотеза T верна, то формула (4) есть представление исходного временного ряда функционалом от fBm (аппроксимация фрактальным броуновским движением). Алгоритм проверки гипотезы (то есть качества аппроксимации) базируется на предельных теоремах, доказанных в [6] для статистик

$$A_n = \frac{1}{n} \sum \xi_k z_k^3, \quad B_n = \frac{1}{n^{1+H}} \sum \xi_k^2 z_k^3, \quad D_n = \frac{1}{n^{2H}} \sum \xi_k z_k^3,$$

где $\xi_k = \sum_{j=1}^{k-1} z_j$.

Если гипотеза T — z_k являются приращениями fBm — верна, то имеет место сходимость ($n \rightarrow \infty$):

$$A_n \rightarrow -1,5R_2^2 \equiv A, \quad B_n \rightarrow 3R_2^{5/2}\eta, \quad \text{если } H \in \left(0; \frac{1}{2}\right),$$

$$D_n \rightarrow 1,5R_2^2 B^2(1), \quad \text{если } H \in \left(\frac{1}{2}; 1\right),$$

где случайная величина $\eta \sim N\left(0; \frac{1}{2H+2}\right)$.

При проверке гипотезы T для статистик B_n и D_n задается уровень значимости $\alpha = 0,1$ гипотеза принимается, если имеют место неравенства

$$\left|A_n + 1,5R_2^2\right| < 0,3; \quad |B_n| < \beta_1; \quad 0 < D_n < \beta_2, \quad (6)$$

где β_k — квантили предельных распределений, соответствующие

$$\alpha = 0,1, \quad \beta_1 = \frac{4,95R_2^{2,5}}{\sqrt{2H+2}}, \quad \beta_2 = 4,08R_2^2.$$

В работе [7] гипотеза T проверена для реальных временных данных стохастической природы.

В табл. 1 представлены значения оценки \hat{H} статистик A_n, B_n, D_n для генерируемого fBm, предельные значения A_n и значения квантилей.

Таблица 1. Сравнительные значения статистик для генерируемого fBm $10 B_H(t)$, $n = 2000$

H	0,1	0,2	0,3	0,4	0,6	0,7	0,8	0,9
\hat{H}	0,1	0,2	0,3	0,4	0,6	0,7	0,8	0,9
A_n	-794	-37,3	-1,18	-0,051	-	-	-	-
A	-789	-35	-1,52	-0,068	-	-	-	-
B_n	-1100	-94	0,33	-0,02	-	-	-	-
β_1	8407	160	3,46	0,622				
D_n	-	-	-	-	0,00034	0	0	0
β_2	-	-	-	-	0,044	0,010	0,0022	0,0004

Табл. 1 иллюстрирует «идеальный» случай, критерии согласия для гипотезы T выполнены для всех контрольных статистик. Данные таблицы используются для сравнения с аналогичными статистиками при анализе временных рядов, содержащих хаотическую составляющую.

ОПРЕДЕЛЕНИЕ КАЧЕСТВА АППРОКСИМАЦИИ СМЕСИ

Аппроксимация временных данных x_1, \dots, x_n по формулам (4), (5) признается удовлетворительной, если контрольные статистики удовлетворяют неравенствам (6). Применение алгоритма аппроксимации к логистической последовательности (1) дает следующие результаты (табл. 2).

Таблица 2. Результаты аппроксимации для логистической последовательности ($n = 2000$)

x_0	0,1	0,2	0,3	0,4	0,6	0,7	0,8	0,9
\hat{H}	0,25	0,15	0,25	0,15	0,15	0,4	0,15	0,25
A_n	1,20	0,62	1,38	0,60	0,62	-0,043	0,56	1,26
B_n	-3,65	-1,90	-4,51	-1,81	-1,90	-0,03	-1,63	-3,93
$A = \lim A_n$	-0,068	-0,077	-0,071	-0,080	-0,080	-0,085	-0,079	-0,068
β_1	0,065	0,080	0,070	0,082	0,082	0,082	0,082	0,065

Из табл. 2 следует, что оценка \hat{H} зависит от начального условия x_0 , во всех случаях аппроксимация соответствует антиперсистентному процессу. Значения контрольных статистик далеки от предельных теоретических значений ($A_n > 0$, $|B_n| > \beta_1$), то есть аппроксимация неудовлетворительна. Исследуем качество аппроксимации для аддитивной смеси (3), увеличивая долю a стохастической составляющей.

Данные табл. 3 свидетельствуют об «агрессивности» хаотической составляющей по отношению к стохастической для $H_{fBm} \geq 0,2$ (для $a = 1; a = 2$) и для $H_{fBm} \geq 0,3$ если $a = 10$. Для указанных значений H ге-

нерируемого f_{Bm} характер смеси определяет логистическая последовательность, то есть значения контрольных характеристик далеки от предельных (неравенства (6) не выполнены), что указывает на невозможность аппроксимации. Для $H_{f_{Bm}} = 0,1$ наблюдается обратная картина — отклонение статистик от предельных значений такое же, как для «чистого» f_{Bm} (табл. 1), то есть смесь допускает качественную аппроксимацию при сопоставимых энергиях ее составляющих. Для $H_{f_{Bm}} = 0,2$ удовлетворительная аппроксимация возможна лишь для $a = 10$.

Таким образом, персистентность ($\hat{H} > 0,5$) исследуемого ряда при выполнении неравенства $D_n < \beta_2$ означает его стохастическую природу: антиперсистентность ($\hat{H} = 0,1 - 0,2$; $A_n \approx A$, $|B_n| < \beta_1$) допускает наличие хаотической компоненты.

Таблица 3. Результаты аппроксимации смеси ($a = 1$; $a = 2$; $a = 10$; $n = 2000$)

$H_{f_{Bm}}$		\hat{H}	A_n	B_n	D_n	$A = \lim A_n$	β_1	β_2
0,1	$a = 1$	0,1	-2,78	-1,14	-1215	-2,95	7,77	5,72
	$a = 2$	0,1	-19,4	-11,6	-8497	-19,4	82,0	14,7
	$a = 10$	0,1	-8289	-21038	-3623570	-8239	157803	302
0,2	$a = 1$	0,15	10,0	-83	2046	-0,84	1,57	3,05
	$a = 2$	0,15	6,83	-39,3	1397	-1,26	2,6	3,7
	$a = 10$	0,2	-48,8	-182	-4663	-45,5	227	22,5
0,3	$a = 1$	0,15	11,4	-104	2330	-0,8	1,49	2,98
	$a = 2$	0,15	10,7	-88,2	2194	-1,0	2,0	3,35
	$a = 10$	0,25	-8,23	8,53	-368	-14,4	52,8	12,6
0,4	$a = 1$	0,15	11,8	-113	2413	-0,75	1,37	2,88
	$a = 2$	0,15	11,9	-111	2430	-0,81	1,52	3,0
	$a = 10$	0,3	1,37	-4,11	28,6	-3,14	7,73	5,90
0,5	$a = 1$	0,15	11,85	-118	2424	-0,72	1,31	2,84
	$a = 2$	0,15	12,1	-123	2480	-0,74	1,31	2,90
	$a = 10$	0,3	12,1	-39,0	253	-1,34	2,66	3,85
0,6	$a = 1$	0,15	11,0	-101	2265	-0,72	1,31	2,83
	$a = 2$	0,15	10,7	-93	2180	-0,72	1,31	2,83
	$a = 10$	0,15	7,27	-38,0	1487	-0,78	1,43	2,93
0,7	$a = 1$	0,15	11,9	-115	2432	-0,72	1,31	2,83
	$a = 2$	0,15	12,3	-122	2517	-0,72	1,31	2,83
	$a = 10$	0,2	15,6	-120	1487	-0,75	1,35	2,89
0,8	$a = 1$	0,15	10,95	-101	2239	-0,72	1,31	2,83
	$a = 2$	0,15	10,4	-92,5	2132	-0,72	1,31	2,83
	$a = 10$	0,15	6,32	-38,0	1293	-0,73	1,32	2,84
0,9	$a = 1$	0,15	11,0	-101	2253	-0,72	1,31	2,83
	$a = 2$	0,15	10,6	-93,2	2158			
	$a = 10$	0,15	6,82	-40,0	1395	-0,74	1,35	2,87

ПРИМЕРЫ АППРОКСИМАЦИИ РЕАЛЬНЫХ ВРЕМЕННЫХ ДАННЫХ

В качестве примеров рассмотрим следующие реальные данные.

1. Концентрация планктона (мг/л) и хлорофилла (мкг/л) в озере Нарочь — всего 132 наблюдений, полученных усреднением многолетних измерений [8].

2. Скорость ветра в штате Аризона за 01.01.2001–01.12.2013 всего 156 наблюдений [9].

3. Осцилляция волн в Северной Атлантике — 10.1980–10.2014 всего 409 наблюдений [10].

Графическое изображение данных приведено на рис. 1, 2, 3, 4.

Применим к этим данным метод аппроксимации согласно описанному выше алгоритму. Значения оценок d , λ , \hat{H} , контрольных статистик A_n , B_n и параметров A и β_2 представлены в табл. 4.

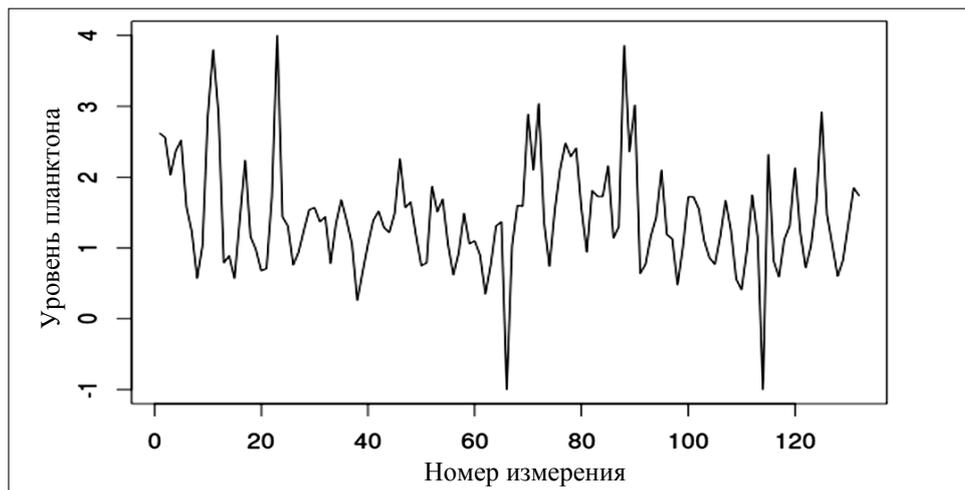


Рис. 1. Концентрация планктона

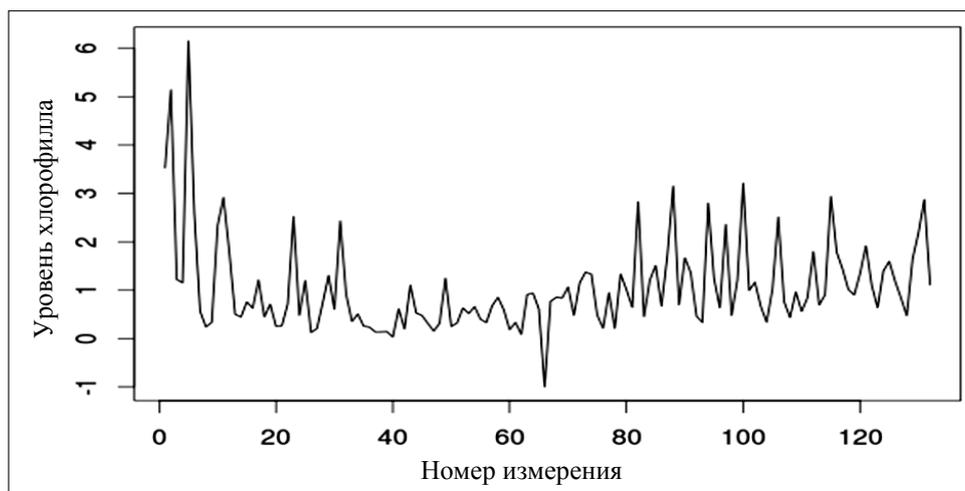


Рис. 2. Концентрация хлорофилла

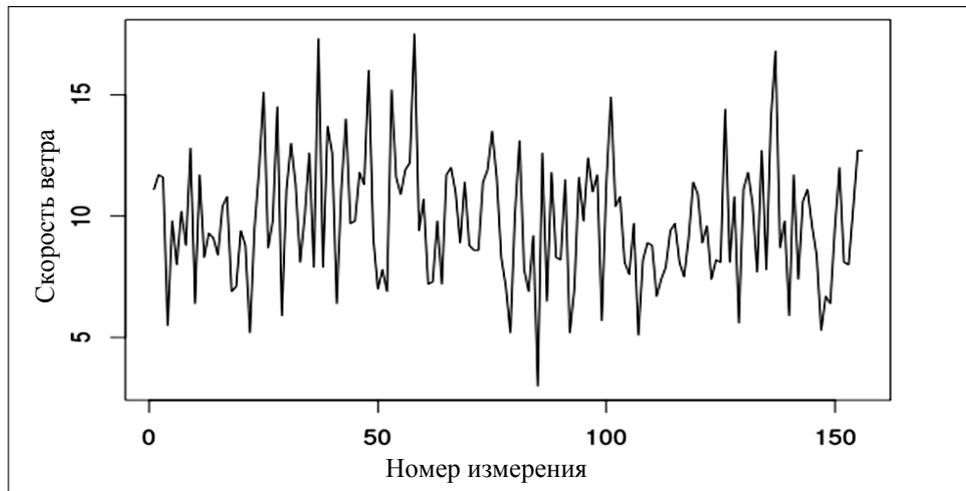


Рис. 3. Скорость ветра

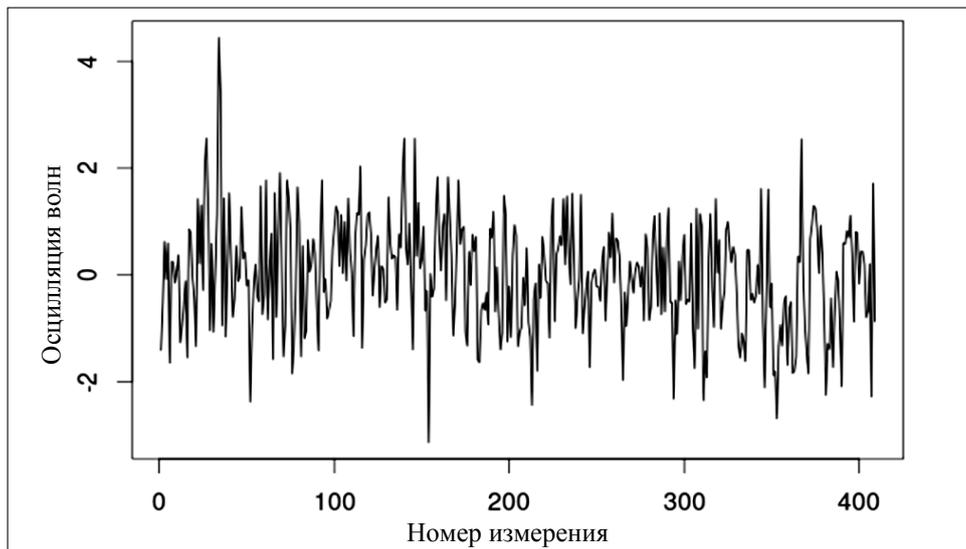


Рис. 4. Осциляция волн

Таблица 4. Параметры аппроксимации реальных данных

Характеристики	Планктон	Хлорофилл	Скорость ветра	Осциляция волн в Атлантике
D	0,55	0,48	0,63	0,61
λ	1,2	1,44	1	1,07
\hat{H}	0,3	0,1	0,4	0,1
A_n	-0,41	-1,26	-274	-3,44
$A = \lim A_n$	-0,48	-1,17	-267	-3,57
B_n	0,19	1,2	-123	-6,38
β_B	0,74	2,46	1921	9,88

Из табл. 4 следует антиперсистентность всех исследуемых данных. Контрольные статистики всех примеров удовлетворяют неравенствам (6), то есть аппроксимация приведенных временных рядов удовлетворительна. Возможно наличие детерминированной составляющей для концентрации хлорофилла и осцилляции океанских волн.

ВЫВОДЫ

Рассмотрена следующая задача анализа временных рядов: определить соотношение детерминизма (как правило, хаотической природы) и стохастичности наблюдаемой временной последовательности. Известные методы решения такой задачи основаны на вычислении доминантного показателя Ляпунова и требуют большого объема выборки. В работе предложен новый метод вычисления искомого соотношения, использующий идею аппроксимации временного ряда фрактальным броуновским движением. Соотношение хаотичность/стохастичность в исследуемом временном ряде определяется значениями статистик, соответствующих этой аппроксимации.

ЛИТЕРАТУРА

1. Кузнецов С.П. Динамический хаос. — М.: Физматлит, 2006. — 296 с.
2. Mishura Y. Stochastic Calculus for Fractional Brownian Motion and Related Processes. Lecture Notes in Mathematics. V. 1929 // Springer-Verlag, 2008. — 392 p.
3. Beran J. Statistics for Long-Memory Processes / Beran J. — Chapman and Hall. — 1995. — 315 p.
4. Бондаренко В.В. Итерационный алгоритм оценивания параметров фрактального броуновского движения // Проблемы управления и информатики. — 2012. — № 4. — С. 28–33.
5. Бондаренко В.В. Аппроксимация временного ряда степенной функцией фрактального броуновского движения // Проблемы управления и информатики. — 2013. — № 3. — С. 113–116.
6. Nourdin I., Nualart D., Tudor C. Central and non-central limit theorems for weighted power variations of fractional Brownian motion // Ann. Institute H. Poincaré Probab Statist. — 2010 — **46**, № 4. — P. 1055–1079.
7. Бондаренко В.В. Проверка качества аппроксимации временного ряда фрактальным броуновским движением // Проблемы управления и информатики. — 2014. — № 3. — С. 102–108.
8. Данные Научно-практического центра АН Республики Беларусь. — <http://biobel.bas-net.by/zoo>.
9. University of Arizona Tree Ring Data: <http://rda.ucar.edu/#!lfd?nb=y&b=topic&v=Paleoclimate>.
10. Climate Prediction Center(USA). — <http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/pna.shtml>.

Поступила 21.12.2014