УДК 519.24.001.57

А.В.Глушков, *д.ф.-м.н.*

Одесский государственный экологический университет

АНАЛИЗ И ПРОГНОЗ АНТРОПОГЕННОГО ВЛИЯНИЯ НА ВОЗДУШНЫЙ БАССЕЙН ПРОМЫШЛЕННОГО ГОРОДА НА ОСНОВЕ МЕТОДОВ ТЕОРИИ ХАОСА: МАТЕМАТИЧЕСКИЕ ОСНОВЫ

С целью развития теоретических основ общего аппарата анализа и прогноза влияния антропогенной нагрузки на состояние атмосферы промышленного города и разработки новой схемы моделирования свойств полей концентраций загрязняющих воздушный бассейн веществ на основе методов теории хаоса в статье выполнен анализ тестов на наличие хаоса в системе (воздушный бассейн промышленного города) и изложена усовершенствованная методика восстановление фазового пространства.

Ключевые слова: *воздушный бассейн промышленного города, экологическое состояние, временные ряды концентраций, загрязняющие вещества, анализ и прогноз, методы теории хаоса*

Введение. Проблема анализа и прогноза влияния антропогенной нагрузки на состояние атмосферы промышленного города и разработка адекватных схем моделирования свойств полей концентраций загрязняющих воздушный бассейн веществ относится к числу наиболее важных и фундаментальных проблем современной прикладной экологии и урбоэкологии. Большая часть моделей, использующихся в настоящее время для оценки состояния (а так же, прогноза) уровня загрязнения атмосферы, является или детерминистическими моделями, или основана на простых статистических регрессиях. Успешность этих моделей, однако, ограничивается как их неспособностью описать нелинейные характеристики загрязняющих веществ, так и недостаточным пониманием вовлеченных физических и химических процессов. В современной теории прогнозов временной ряд изменения какой-либо динамической характеристики сложной системы рассматривается как реализация случайного процесса, когда случайность является результатом сложного движения с многими независимыми степенями свободы. Альтернативой случайности является хаос, который имеет место даже в очень простых детерминистических системах. Хотя использование методов теории хаоса устанавливает определенное фундаментальное ограничение на долгосрочный прогноз (см., напр., [1-4]), тем не менее, как было показано в целой серии работ (см., напр., [4-16]), данные методы могут быть использованы для кратко- или средне-срочного прогноза. В предыдущих наших работах было показано (см. [1-3]), что для временных изменений концентраций двуокиси азота (NO_2) и сернистого ангидрида (SO_2) на двух постах Гданьского региона имеет место низкоразмерный хаос, что позволяет применить для них метод нелинейного прогноза. Такая методология с успехом использовался при анализе многих гидрометеорологических характеристик (см., напр., [11,14,17,18]).

Целью нашей работы является изложение и развитие теоретических основ общего аппарата анализа и прогноза влияния антропогенной нагрузки на состояние атмосферы промышленного города и разработка новой количественной схемы моделирования свойств полей концентраций загрязняющих воздушный бассейн веществ на основе методов теории хаоса. Общий анализ и прогноз включают в себя исследование временных рядов концентраций загрязняющих веществ, восстановление спектра размерностей Ляпунова и на его основе расчет размерности Калана-Йорка, энтропии Колмогорова и далее разработку усовершенствованной методики построения модели краткосрочного прогноза экологического состояния атмосферы промышленного города. В данной статье выполнен анализ тестов на наличие хаоса в системе (воздушный бассейн промышленного города) и изложена усовершенствованная методика восстановления фазового пространства.

Общая методика анализа хаотических систем: Тест на наличие хаоса в системе. Отметим сразу, что методика работы с хаотическим временным рядом может быть сведена к нескольким процедурам: идентификация хаотического режима, реконструкция фазового пространства, классификация временных рядов, построение модели прогноза. С целью выявления наличия хаоса в системе с шумом обычно рекомендуется использовать тест, введенный Готтводом и Мелбеном [6]. Он имеет преимущества перед традиционными методами идентификации хаоса, которые будут описаны ниже. Во-первых, этот тест применяется непосредственно к данным временного ряда и не требует восстановления фазового пространства. Более того, несущественна и природа динамической системы, лежащей в основе временного ряда. Во-вторых, это бинарный тест, дающий, в принципе, 0 или 1, т.е. вполне однозначный результат. Рассмотрим дискретный временной ряд (скажем, концентраций загрязняющих воздушный бассейн города веществ) $s(n)$ с n =1, 2, 3, … Выберем вещественную константу c (на практике выбирается некое количество значений c). Для каждого значения c, определим для всех n

$$p(n) = \sum_{j=1}^{n} s(j)\cos(jc) \cdot \tag{1}$$

Далее, определим среднеквадратический сдвиг

$$M(n) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} [p(j+n) - p(j)]^2 \cdot \tag{2}$$

Если динамика системы является регулярной (периодической или квазипериодической), то с вероятностью единица $M(n)$ является ограниченной функцией от n. Однако если динамика хаотична (в довольно нежестком смысле), то с вероятностью единица $M(n) = V(n) + O(1)$ для некоторых $V > 0$ [6]. Далее можно определить скорость асимптотического роста среднеквадратического сдвига с помощью выражения

$$K = \lim_{n\to\infty} \frac{\log M(n)}{\log n} \cdot \tag{3}$$

Тогда, $K = 0$ является признаком регулярной динамики, а $K = 1$ – хаотической. Далее, предположим, что тестируется временной ряд с конечным числом данных $s(n)$, $1 \le n \le N$. Определим

$$M(n) = \frac{1}{N-n} \sum_{j=1}^{N-n} [p(j+n) - p(j)]^2 \cdot \tag{4}$$

Если $n \ll N$, то вполне допустимо, что отношение $M(n)$ к n будет аналогичным указанному выше. Чтобы предотвратить логарифмирование отрицательных значений, можно использовать соотношение $\log(M(n) + 1)$ к $\log n$ для $1 \le n \le N_1$ с некоторым выбранным N_1, $1 \ll N_1 \ll N$. На практике достаточно выбрать $N_1 = N/10$. Тогда, значение K, близкое к 0, определяет регулярную динамику, а близкое к 1 – хаотическую. В принципе рекомендуется выбрать случайным образом некоторое количество (100) значений констант c и для каждой рассчитать значение K, а в качестве конечного результата взять медиану всех рассчитанных K.

Общая методика анализа хаотических систем: Восстановление фазового пространства. Если динамическая система является нелинейной, применение преобразования Фурье не даст, скорее всего, какого-либо удовлетворительного результата, как в случае линейной системы. Связано это с тем, что процессы, приводящие к хаотическому режиму, являются фундаментально многомерными. Поэтому необходимо восстанавливать фазовое пространство системы, как можно лучше используя информацию, содержащуюся в векторе значений динамической характеристики, скажем, $s(n)$. Этот процесс реконструкции приведет к некоему набору d-мерных векторов $\mathbf{y}(n)$, которые заменят наблюденные скалярные данные, и заключается в сочетании динамических концепций о нелинейных системах, как о генераторах информации, и геометрических представлений о том, как обнаружить аттрактор при помощи координат, определенных на основе их информационно-теоретического содержания. В работе Пакарда и др. [7] представлены соображения об использовании координат с временной задержкой для реконструкции фазового пространства наблюденной динамической системы и, как следствие, производной по времени от s. В действительности нет необходимости иметь производные, чтобы сформировать систему координат, в которой захвачена структура орбит в фазовом пространстве, а можно напрямую использовать запаздывающие переменные $s(n+\tau)$, где τ – некоторое целое число, которое нужно определить. Используя совокупность временных задержек для создания вектора в d-мерном пространстве

$$\mathbf{y}(n) = [s(n), s(n+\tau), s(n+2\tau), \ldots, s(n+(d-1)\tau)], \tag{5}$$

можно получить требуемые координаты. В нелинейной системе $s(n+j\tau)$ – некоторая (неизвестная) нелинейная совокупность реальных физических переменных. Отметим, что такой метод вполне удовлетворительно работает даже в классической задаче об аттракторе Лоренца.

Рассмотрим далее вопрос о выборе временной задержки, используя теорему вложения Мане-Тейкенса [10,11]. Положение о том, что в принципе, допустима любая временная задержка, не всегда применимо в конкретном случае. Если τ слишком мало, то координаты $s(n+j\tau)$ и $s(n+(j+1)\tau)$ будут так близки друг к другу, что их нельзя будет различить одна от другой. Если τ слишком велико, то $s(n+j\tau)$ и $s(n+(j+1)\tau)$ будут полностью независимы друг от друга в статистическом смысле и проекция орбит на аттрактор будут направлены по двум совершенно несвязанным направлениям. Поэтому необходимо выбрать некое промежуточное положение между двумя упомянутыми случаями. Первым подходом для выбора τ может быть расчет линейной автокорреляционной функции

$$C_L(\delta) = \frac{\dfrac{1}{N}\sum\limits_{m=1}^{N}[s(m+\delta)-\bar{s}][s(m)-\bar{s}]}{\dfrac{1}{N}\sum\limits_{m=1}^{N}[s(m)-\bar{s}]^2}, \tag{6}$$

где

$$\bar{s} = \frac{1}{N}\sum_{m=1}^{N} s(m),$$

и нахождение временной задержки, на которой $C_L(\delta)$ первый раз переходит через некое значение, что дает вполне допустимое, в принципе, значение τ, при котором $s(n+j\tau)$ и $s(n+(j+1)\tau)$ линейно независимы. Однако линейная независимость двух переменных не означает, что эти переменные независимы нелинейно, так как нелинейная взаимосвязь между ними может быть больше линейной. Поэтому предпочтительнее применять подход, в котором используется нелинейное понятие независимости, а именно, среднюю взаимную информацию. Коротко, концепцию взаимной информации можно описать следующим образом. Пусть имеется две системы A и B с результатами измерений a_i и b_k. Количество битов, которое можно узнать об измерениях a_i по измерениям b_k, задается параметрами теории информации [12], как

$$I_{AB}(a_i,b_k) = \log_2\left(\frac{P_{AB}(a_i,b_k)}{P_A(a_i)P_B(b_k)}\right), \tag{7}$$

где – $P_{AB}(a_i, b_k)$ общая плотность распределения вероятности, а $P_A(a_i)$ и $P_B(b_k)$ – предельные плотности распределения вероятности для систем A и B, соответственно. Взаимная информация является неотрицательной, симметричной и равна нулю только тогда, когда системы независимы. Средняя взаимная информация между измерениями любых значений a_i системы A и b_k системы B – это средняя величина по всем возможным измерениям от $I_{AB}(a_i, b_k)$, т.е.

$$I_{AB}(\tau) = \sum_{a_i,b_k} P_{AB}(a_i,b_k) I_{AB}(a_i,b_k)\,. \tag{8}$$

Чтобы применить последнее определение в контексте некой физической системы, заменим систему A измерениями $s(n)$, а систему B – измерениями через некоторую временную задержку $s(n+\tau)$. Тогда среднее количество информации о $s(n+\tau)$, которое имеется во время измерений $s(n)$, будет

$$I(\tau) = \sum_{n=1}^{N} P(s(n), s(n+\tau)) \log_2\left(\frac{P(s(n), s(n+\tau))}{P(s(n))P(s(n+\tau))}\right). \tag{9}$$

Теперь, как и в случае с автокорреляционной функцией, нужно определить ту величину $I(\tau)$, которая бы подходила для выбора временной задержки τ. Если τ слишком мало, измерения $s(n)$ и $s(n+\tau)$ содержат так много информации друг о друге, что нет необходимости делать оба измерения. Если τ велико, то $I(\tau)$ будет стремиться к нулю, и ничто не будет связывать $s(n)$ и $s(n+\tau)$, а это также не является лучшим выбором. Фрейзер и Суинни [13] рекомендовали (и показали на примерах) использовать в качестве τ значение, при котором $I(\tau)$ достигает своего первого локального минимума.

Далее рассмотрим вопрос выбора размерности вложения. Целью определения размерности вложения является восстановление настолько большого Евклидова пространства R^d, чтобы весь ряд точек размерности d_A мог быть развернут без какой-либо неопределенности. Согласно положениям теоремы вложения, важно иметь такую размерность d_E, чтобы она была больше d_A, тогда как выбор $d_E < d_A$ неприемлем в любом случае. Другими словами, не нужно искать размерность d_A, а можно взять какую-ту заведомо большую размерность d_E. Например, в случае низкоразмерного хаоса можно априори задать размерность вложения 10 или даже 15, что вполне приемлемо с математической точки зрения. Однако при этом возникают две проблемы. Во-первых, большинство данных требуют перебора вариантов при компьютерных расчетах в пространстве R^d, и машинное время экспоненциально возрастает в зависимости от размерности вложения. Во-вторых, в присутствии шума или других высокочастотных компонентов во временных рядах, в «дополнительные» размерности, не обусловленные динамикой системы, попадают исключительно высокочастотные составляющие сигнала. Ясно, что необходимо отыскать именно размерность d_A.

Естественно, в настоящее время имеется несколько методик, позволяющих восстановить размерность аттрактора. По нашему мнению [1], наиболее эффективными следует считать две. Первая методика, называемая методом корреляционной размерности, является одной из наиболее широко используемых при исследовании наличия хаоса во временных рядах. Этот метод использует корреляционный интеграл функции $C(r)$ для того, чтобы найти различия между хаотическими и стохастическими системами. Для расчета корреляционного интеграла наиболее часто используется алгоритм Грассбергера-Прокаччиа [15], в соответствии с которым

$$C(r) = \lim_{N \to \infty} \frac{2}{N(n-1)} \sum_{\substack{i,j \\ (1 \le i < j \le N)}} H\left(r - \| \mathbf{y}_i - \mathbf{y}_j \|\right), \qquad (10)$$

где H – ступенчатая функция Хевисайда, $H(u) = 1$ для $u > 0$ и $H(u) = 0$ для $u \le 0$;
r – радиус сферы с центром в $\mathbf{y}_i$ или $\mathbf{y}_j$;
N – длина временного ряда.

Если временной ряд характеризуется аттрактором, то корреляционный интеграл $C(r)$ соотносится с радиусом r посредством выражения

$$d_2 = \lim_{\substack{r \to 0 \\ N \to \infty}} \frac{\log C(r)}{\log r}, \qquad (11)$$

где d_2 – корреляционная размерность, которую можно определить как наклон линии в координатах $\log C(r)$ и $\log r$ посредством среднеквадратического подбора прямой линии в некотором диапазоне r, называемом диапазоном масштабирования. Если корреляционная размерность достигает насыщения на некотором значении размерности вложения, то динамика системы в целом рассматривается как хаотическая. Значение корреляционной размерности, при котором она достигает насыщения, определяется как корреляционная размерность аттрактора (d_A). Ближайшее целое число большее, чем d_2, дает оптимальную (необходимую) размерность вложения d_E для реконструкции фазового пространства или количество переменных, необходимых для моделирования динамики системы. С другой стороны, если корреляционная размерность неограниченно увеличивается с ростом размерности вложения, динамика системы рассматривается как стохастическая. Хотя изложенный метод корреляционной размерности является достаточно эффективным, тем не менее, вполне возможны ситуации (см., напр.,[1-3,15]), когда его применение приводит к не совсем корректным результатам. Поэтому для получения надежных результатов метод корреляционной размерности следует совместить, напр., с алгоритмом ложных ближайших соседних точек. Этот алгоритм также позволяет определить размерность вложения для реконструкции фазового пространства, а также проверить результаты, полученные по методу корреляционной размерности. Его суть заключается в поиске ответа на ключевой вопрос теоремы вложения: когда будет исключено ложное пересечение орбиты самой себя вследствие проецирования аттрактора в пространство слишком низкой размерности. Другими словами, нужно определить, когда точки в размерности d являются ближайшими соседями самих себя. Если исследовать эту проблему в размерности $d = 1$, затем – $d = 2$ и т.д., пока не останется ложных ближайших соседних точек, то можно установить, основываясь только на геометрическом рассмотрении аттрактора, значение необходимой размерности вложения $d_E = d_N$. В размерности d каждый вектор

$$\mathbf{y}(k) = [s(k), s(k+\tau), s(k+2\tau), \ldots, s(k+(d-1)\tau)] \tag{12}$$

имеет ближайший соседний вектор $\mathbf{y}^{NN}(k)$. Евклидово расстояние в размерности d между векторами $\mathbf{y}(k)$ и $\mathbf{y}^{NN}(k)$ назовем $R_d(k)$

$$\begin{gathered} R_d^2(k) = [s(k) - s^{NN}(k)]^2 + [s(k+\tau) - s^{NN}(k+\tau)]^2 + \\ \ldots + [s(k+\tau(d-1)) - s^{NN}(k+\tau(d-1))]^2. \end{gathered} \tag{13}$$

По-видимому, $R_d(k)$ мало для случая большого количества данных N и имеет порядок $1/N^{1/d}$. В размерности $d+1$ это расстояние изменяется вследствие $(d+1)$-х координат $s(k+d\tau)$ и $s^{NN}(k+d\tau)$, а выражение для него имеет вид

$$R_{d+1}^2(k) = R_d^2(k) + [s(k+d\tau) - s^{NN}(k+d\tau)]^2. \tag{14}$$

Переходя от размерности d к размерности $d+1$, можно определить некую пороговую величину R_T, до которой соседние точки являются ложными. Тогда, если

$$\frac{|s(k+d\tau) - s^{NN}(k+d\tau)|}{R_d(k)} > R_T, \tag{15}$$

ближайшие соседние точки в момент времени k являются ложными. Как было показано Кеннелом и др. [16], для значений R_T в диапазоне $10 \le R_T \le 50$ число ложных соседних точек, определяемых этим критерием, постоянно. На практике определяется процентное содержание ложных ближайших соседних точек, и в качестве d_N берется то значение, при котором это процентное содержание почти равно нулю.

Выводы. В данной статье, преследуя цель развития теоретических основ усовершенствованного общего аппарата анализа и прогноза влияния антропогенной нагрузки на состояние атмосферы промышленного города, и разработку новой схемы моделирования свойств полей концентраций загрязняющих воздушный бассейн веществ на основе методов теории хаоса, мы изложили наиболее эффективные методики тестирования наличия хаоса в динамической системе (применительно к воздушному бассейну промышленного города) и восстановления фазового пространства системы. Эти аспекты являются ключевыми для последующего рассмотрения задач классификации временных рядов загрязняющих веществ в атмосфере промышленного города и построения моделей кратко-и средне-срочного прогноза влияния антропогенной нагрузки на экологическое состояние системы.

Список литературы

1. *Бунякова Ю.Я, Глушков А.В.* Анализ и прогноз влияния антропогенных факторов на воздушный бассейн промышленного города.- Одесса: Экология, 2010.-256с.
2. *Глушков А.В., Хохлов В.Н.,Сербов Н.Г, Бунякова Ю.Я, Балан А.К., Баланюк Е.П.* Низкоразмерный хаос во временных рядах концентраций загрязняющих веществ в атмосфере и гидросфере// Вестник Одесского гос. экологического ун-та.-2007.-N4.-С.337-348.
3. *Glushkov A.V., Khokhlov V.N., Prepelitsa G.P., Tsenenko I.A.* Temporal variability of the atmosphere ozone content: Effect of North-Atlantic oscillation// Optics of atmosphere and ocean.-2004.-Vol.14.-P.219-223.
4. *Sivakumar B.* Chaos theory in geophysics: past, present and future // Chaos, Solitons & Fractals.-2004.-Vol.19.-P.441-462.
5. *Chelani A.B.* Predicting chaotic time series of PM10 concentration using artificial neural net-

work // Int. J. Environ. Stud.-2005.-Vol.62.-P.181-191.
6. *Gottwald G.A., Melbourne I.* A new test for chaos in deterministic systems// Proc. Roy. Soc. London. Ser. A. Mathemat. Phys. Sci.- 2004.-Vol.460.-P.603-611.
7. *Packard N.H., Crutchfield J.P., Farmer J.D., Shaw R.S.* Geometry from a time series// Phys. Rev. Lett. -1980.-Vol.45.-P.712-716.
8. *Sauer T., Yorke J., Casdagli M.* Embedology// J. Stat. Phys.-1991.-Vol.65.-P.579-616.
9. *Abarbanel H.D.I., Brown R., Sidorowich J.J., Tsimring L.Sh.* The analysis of observed chaotic data in physical systems // Rev. Mod. Phys.-1993.-Vol.65.- P.1331-1392.
10. *Mañé R.* On the dimensions of the compact invariant sets of certain non-linear maps// Dynamical systems and turbulence, Warwick 1980. Lecture Notes in Mathematics No. 898 / D.A. Rand, L.S. Young (Eds.). – Berlin: Springer, 1981.-P.230-242.
11. *Takens F.* Detecting strange attractors in turbulence // Dynamical systems and turbulence, Warwick 1980. Lecture Notes in Mathematics No. 898 / D.A. Rand, L.S. Young (Eds.). – Berlin: Springer, 1981.-P.366-381.
12. *Gallager R.G.* Information theory and reliable communication.- NY: Wiley, 1968.- 608p.
13. *Fraser A.M., Swinney H.* Independent coordinates for strange attractors from mutual information// Phys. Rev. A.-1986.-Vol.33.-P.1134-1140.
14. *Schreiber T.* Interdisciplinary application of nonlinear time series methods // Physics Rep.-1999.-Vol.308.-P.1-64.
15. *Grassberger P., Procaccia I.* Measuring the strangeness of strange attractors// Physica D. – 1983.-Vol.9.-P.189-208.
16. *Kennel M., Brown R., Abarbanel H.* Determining embedding dimension for phase-space reconstruction using a geometrical construction// Phys. Rev. A.-1992.-Vol.45.-P.3403-3411.
17. *Glushkov A.V., Loboda N.S., Khokhlov V.N.* Using meteorological data for reconstruction of annual runoff series over an ungauged area: Empirical orthogonal functions approach to Moldova-Southwest Ukraine region//Atmospheric Research (Elsevier).-2005.-Vol.77.-P.100-113.
18. *Glushkov A.V., Loboda N.S., Khokhlov V.N., Lovett L.* Using non-decimated wavelet decomposition to analyze time variations of North Atlantic Oscillation, eddy kinetic energy, and Ukrainian precipitation // Journal of Hydrology (Elsevier).-2006.-Vol.322.-P.14-24.

Аналіз і прогноз антропогенного впливу на повітряний басейн промислового міста на основі методів теорії хаосу: Математичні основи. Глушков О.В.

З метою розвитку теоретичних основ загального апарату аналізу та прогнозу впливу антропогенного навантаження на стан атмосфери промислового міста і розробки нової схеми моделювання властивостей полів концентрацій забруднюючих повітряний басейн речовин на основі методів теорії хаосу, виконано аналіз тестів на наявність хаосу в системі (повітряний басейн промислового міста) і викладено удосконалену методику відновлення фазового простору.

***Ключові слова**: повітряний басейн промислового міста, екологічний стан, часові ряди концентрацій, забруднюючі речовини, аналіз і прогноз, методи теорії хаосу*

Analysis and forecast of anthropogenic impact on air basin of industrial city on basis of a chaos theory methods: Mathematical foundations. Glushkov A.V.

In order to develop the theoretical foundations of the total approach to analysis and prediction of the influence of anthropogenic impact on the atmosphere of the industrial city and development of a new scheme of modelling the properties of fields of the polluting substances concentrations in the air basin by means of a chaos theory methods we present an analysis of the most effective tests on the presence of chaos in the system (air basin of industrial city) and improved method for reconstruction of the phase space.

***Keywords**: air basin of the industrial city, the ecological state of the, time series of concentrations, pollutants, analysis and prediction methods of the theory of chaos*