

СИНТЕЗ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ МОДЕЛЕЙ В ЗАДАЧАХ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА

© Мантула Е., 2013

Дослідження спрямовано на синтез прогнозуючих моделей для нестационарного часового ряду. На відміну від загальноприйнятих підходів до розв'язання різноманітних задач екологічного моніторингу, що використовують різні структури моделей і методів оцінки параметрів, запропоновано метод адаптивної процедури прогнозу в поєднанні з низькою обчислювальною складністю, що дає змогу синтезувати кращі оцінки щодо заданого критерію.

Ключові слова: нестационарний часовий ряд, прогнозуючі моделі, комбінування прогнозів.

Исследования были направлены на синтез моделей для нестационарного временного ряда. В отличие от общепринятых подходов для решения различных задач экологического мониторинга, которые используют различные структуры моделей и методов оценки параметров, был предложен метод адаптивной процедуры прогноза в сочетании с низкой вычислительной сложностью, который позволяет синтезировать наилучшие оценки с точки зрения заданного критерия.

Ключевые слова: нестационарный временной ряд, прогнозирующие модели, комбинирование прогнозов

Введение

В задачах экологического мониторинга, связанных с контролем движения воздушных масс и стоком рек, уровней загрязнения воды и воздуха и т.п., достаточно часто возникает необходимость построения прогнозов различных показателей, характеризующих состояние окружающей среды [3]. Использование традиционных подходов к прогнозированию в таком случае резко ограничивается существенной нестационарностью контролируемых процессов, содержащих как нерегулярные тренды, так и внезапные скачки, стохастичностью и хаотичностью, ограниченностью априорной информации (только объем обучающих выборок) и т.п., что не позволяет синтезировать универсальную прогнозирующую модель, а вынуждает в каждом конкретном случае подбирать наилучший вариант от простейших регрессионных уравнений до искусственных нейронных сетей [4–7]. В связи с этим представляется целесообразным использование комплекса прогнозирующих моделей, отличающихся структурой, методами оценивания параметров, используемыми априорными предложениями и т.п., которые параллельно решали бы задачу экологического прогнозирования. При этом такая система должна содержать блок принятия решений, который, анализируя полученные прогнозы, на каждом шаге синтезировал бы наилучшее с точки зрения принятого критерия прогнозное значение анализируемого показателя, описывающего ту или иную характеристику экологической системы.

Синтез прогнозирующих моделей

В задачах экологического мониторинга достаточно эффективными показали себя прогнозирующие модели [4, 7], основанные на подходе Дж. Бокса – Г. Дженкинса [8–10], исполь-

зующем те или иные формы дискретных разностных уравнений, описывающих так называемый формирующий фильтр, имеющий в общем случае вид:

$$A(z^{-1})y(k) = \frac{B(z^{-1})}{F(z^{-1})}x(k) + \frac{B(z^{-1})}{F(z^{-1})}\xi(k), \quad (1)$$

где $y(k)$ $x(k)$ $\xi(k)$ – прогнозируемая экзогенная (входная) и ненаблюдаемая стохастическая компонента соответственно, при этом относительно $\xi(k)$ предполагается, что она имеет нулевое математическое ожидание и ограниченную дисперсию; $k = 1, 2, \dots$ – дискретное время;

$$\begin{aligned} A(z^{-1}) &= 1 + a_1z^{-1} + a_2z^{-2} + \dots + a_{n_A}z^{-n_A}, \\ B(z^{-1}) &= b_1z^{-1} + b_2z^{-2} + \dots + b_{n_B}z^{-n_B}, \\ C(z^{-1}) &= 1 + c_1z^{-1} + c_2z^{-2} + \dots + c_{n_C}z^{-n_C}, \\ D(z^{-1}) &= 1 + d_1z^{-1} + d_2z^{-2} + \dots + d_{n_D}z^{-n_D}, \\ F(z^{-1}) &= 1 + f_1z^{-1} + f_2z^{-2} + \dots + f_{n_F}z^{-n_F} \end{aligned}$$

– некоторые полиномы от оператора сдвига назад z^{-1} ; a_i, b_i, c_i, d_i, f_i – параметры формирующего фильтра, подлежащие определению. Несложно заметить, что в общем случае модель (1) содержит $n_A + n_B + n_C + n_D + n_F$ параметров и когда это число достаточно велико, может потребоваться большой объем исходных данных для их оценивания. В связи с этим на практике распространение получили более простые формы прогнозирующих моделей, среди которых в первую очередь можно отметить так называемую ARX-модель. Следует подчеркнуть, что основные ограничения связаны с неравенством $n_A + n_B < N$. И, кроме того, параметры мониторинга процесса должны быть стационарными при наблюдаемом интервале, что вряд ли возможно для большинства задач мониторинга природной среды. Получаем

$$y(k) = -a_1y(k-1) - \dots - a_{n_A}y(k-n_A) + b_1x(k-1) + \dots + a_{n_B}x(k-n_B) + \xi(k) \quad (2)$$

В дальнейшем будем рассматривать сокращенную форму $y(k-1) = -x_1(k), \dots, y(k-n_A) = -x_{n_A}(k)$, $x(k-1) = x_{n_A+1}(k), \dots, x(k-n_B) = x_{n_A+n_B}(k)$, $\varphi(k) = (x_1(k), \dots, x_{n_A}(k), x_{n_A+1}(k), \dots, x_{n_A+n_B}(k))^T$, $w = (a_1, \dots, a_{n_A}, b_1, \dots, b_{n_B})^T$.

Можно переписать (2) в компактной форме

$$y(k) = \sum_{i=1}^{n_A+n_B} w_i \varphi_i(k) + \xi(k) = w^T \varphi(k) + \xi(k) \quad (3)$$

Несложно заметить, что неизвестные параметры модели (3) w_i могут быть определены с помощью стандартного регрессионного анализа [9], при этом, если имеется выборка, содержащая N наблюдений, исходные параметры могут быть определены в виде

$$w(N) = \sum_{k=1}^N (\varphi(k)\varphi^T(k))^{-1} \sum_{k=1}^N y(k)\varphi(k) \quad (4)$$

Для $k \geq N$ может быть построен прогноз на $(k+1)$ -й момент времени в виде

$$\hat{y}(k+1) = w^T(N)\varphi(k+1),$$

при этом, как несложно заметить, все компоненты вектора $\varphi(k+1)$ в общем случае уже известны.

В связи с этим вместо (4) целесообразно использование его рекуррентной модификации [11].

$$\begin{cases} w(k) = w(k-1) + \frac{P(k-1)(y(k) - w^T(k-1)\varphi(k))}{1 + \varphi^T(k)P(k-1)\varphi(k)}\varphi(k), \\ P(k) = P(k-1) - \frac{P(k-1)\varphi(k)\varphi^T(k)P(k-1)}{1 + \varphi^T(k)P(k-1)\varphi(k)} \end{cases}$$

либо (для нестационарных задач) его экспонента взвешенного варианта [12]

$$\begin{cases} w(k) = w(k-1) + \frac{P(k-1)(y(k) - w^T(k-1)\varphi(k))}{\alpha + \varphi^T(k)P(k-1)\varphi(k)}\varphi(k), \\ P(k) = \frac{1}{\alpha}(P(k-1) - \frac{P(k-1)\varphi(k)\varphi^T(k)P(k-1)}{\alpha + \varphi^T(k)P(k-1)\varphi(k)}), 0 < \alpha < 1 \end{cases} \quad (5)$$

где $w(0)$ – произвольно $P(0) = \delta E$, $\delta \gg 0$, E is $(n_A + n_B) \times (n_A + n_B)$ – единичная матрица, отвечающая за устаревшую информацию.

Поскольку при малых значениях параметра α алгоритм (5) может быть неустойчив, более эффективным может оказаться использование адаптивных алгоритмов идентификации [13, 14], таких как алгоритм Качмаша

$$w(k) = w(k-1) + \frac{y(k) - w^T(k-1)\varphi(k)}{\varphi^T(k)\varphi(k)}\varphi(k). \quad (6)$$

его аддитивной и мультипликативной модификации

$$w(k) = w(k-1) + \frac{y(k) - w^T(k-1)\varphi(k)}{\beta + \varphi^T(k)\varphi(k)}\varphi(k), \quad \beta > 0, \quad (7)$$

$$w(k) = w(k-1) + \gamma \frac{y(k) - w^T(k-1)\varphi(k)}{\varphi^T(k)\varphi(k)}\varphi(k), \quad 0 < \gamma < 2 \quad (8)$$

процедуры стохастической аппроксимации

$$\begin{cases} w(k) = w(k-1) + \eta(k)(y(k) - w^T(k-1)\varphi(k)), \\ \eta(k) > 0, \quad \sum_{k=1}^{\infty} \eta(k) = \infty, \quad \sum_{k=1}^{\infty} \eta^2(k) < \infty \end{cases} \quad (9)$$

и множество других.

Специфика задач, связанных с обработкой данных экологического мониторинга, такова, что невозможно сказать заранее, какой из методов окажется наилучшим. В связи с этим допустим, что, если в нашем распоряжении имеется m_1 потенциальных методов идентификации, то нужно быть готовым к тому, что возможна ситуация, когда все m_1 процедур необходимо будет применить.

Следует отметить также, что достаточно сложной проблемой также является определение порядков прогнозирующей модели n_A и n_B . Методология Бокса–Дженкинса подразумевает проведение анализа автокорреляционных и взаимокорреляционных функций анализируемых процессов, однако для этого, во-первых, имеющиеся данные о временных рядах должны быть достаточно репрезентативными, а, во-вторых, сами эти ряды должны быть стационарными. В связи с этим n_A и n_B , как правило, определяются эмпирически, в связи с чем в распоряжении исследователя может оказаться m_2 моделей различных порядков. В сочетании с различными алгоритмами идентификации эти модели формируют $m = m_1 m_2$ различных прогнозирующих структур, каждая из которых в каждый конкретный момент текущего k может оказаться как наилучшей, так и наихудшей. При этом возникает дополнительная задача: как синтезировать наилучший прогноз $\hat{y}_j(k)$ на основе уже имеющихся $\hat{y}_j(k)$, построенных на основе различных моделей и алгоритмов идентификации.

Адаптивное комбинирование прогнозов

Объединить m различных прогнозов $\hat{y}_j(k), j = \overline{1, m}$ одной временной последовательности $y(k)$ можно с помощью того или иного метода комбинирования [15–20] на основе использования линейной формы

$$\hat{y}^*(k) = \sum_{j=1}^m c_j(k) \hat{y}_j(k), \quad (10)$$

с учетом ограничений на несмещенность $\hat{y}(k)$ в виде

$$\sum_{j=1}^m c_j(k) = 1,$$

или

$$c^T(k)I = 1 \quad (11)$$

где $c(k) = (c_1(k), c_2(k), \dots, c_m(k))^T$, I – это $(m \times 1)$ – вектор, состоящий из единиц.

В качестве оценки качества прогнозирования каждой из моделей $\hat{y}_j(k)$ на интервале от 1 до N можно использовать дисперсию

$$\sigma_j^2(N) = \frac{1}{N} \sum_{k=1}^N (y(k) - \hat{y}_j(k))^2.$$

при этом понятно, что чем меньше эта дисперсия, тем лучше соответствующая прогнозирующая модель. Тогда несложно видеть, что в простейшем случае коэффициенты формы (10) могут быть определены с помощью простого выражения

$$c_j(N) = \frac{\sigma_j^{-2}(N)}{\sum_{l=1}^m \sigma_l^{-2}(N)}. \quad (12)$$

В адаптивном варианте вместо (12) удобнее использовать соотношения

$$\sigma_j^2(k) = \sigma_j^2(k-1) + \frac{1}{k} ((y(k) - \hat{y}_j(k))^2 - \sigma_j^2(k-1)) = \frac{k-1}{k} \sigma_j^2(k-1) + \frac{1}{k} ((y(k) - \hat{y}_j(k))^2)$$

для стационарных процессов и

$$\sigma_j^2(k) = \alpha \sigma_j^2(k-1) + (1-\alpha) ((y(k) - \hat{y}_j(k))^2), \quad 0 < \alpha < 1$$

для нестационарных. Здесь α имеет тот же смысл, что и в (5).

При этом вместо (12) естественно используются выражения

$$c_j(k) = \frac{\sigma_j^{-2}(k)}{\sum_{l=1}^m \sigma_l^{-2}(k)}. \quad (13)$$

Поскольку выражения (12) и (13) имеют сугубо интуитивный характер, естественно рассмотреть эту задачу с более формальной точки зрения.

Для этого введем критерий идентификации вида

$$J = \sum_{k=1}^N (y(k) - c^T \hat{y}(k))^2 = \sum_{k=1}^N (y(k) - \hat{y}^*(k))^2 \quad (14)$$

(здесь $\hat{y}(k) = (\hat{y}_1(k), \hat{y}_2(k), \dots, \hat{y}_m(k))^T$), на основе которого с учетом ограничений (11) можно записать функцию Лагранжа

$$L = \|Y - \hat{Y}c\|^2 + \lambda(c^T I - 1)$$

где $Y = (y(1), y(2), \dots, y(N))^T$, $\hat{Y}_j = (\hat{y}_j(1), \hat{y}_j(2), \dots, \hat{y}_j(N))^T$, $\hat{Y} = (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_m)$ - $N \times m$ матрица.

Решая систему уравнений Куна-Таккера, получаем оптимальные решения в виде [15]

$$c(N) = \frac{R^{-1}(N)I}{I^T R^{-1}(N)I} \quad (15)$$

где $R(N) = \hat{Y}^T \hat{Y}$.

В [16, 21] была введена рекуррентная модификация этой процедуры (аналогично (3) и (4)), имеющая вид

$$\begin{cases} c(k) = c^*(k) - R^{-1}(k) \frac{I^T c^*(k) - 1}{I^T R^{-1}(k)I} I, \\ c^*(k) = c^*(k-1) + \frac{R^{-1}(k-1)(y(k) - c^{*T}(k-1)\hat{y}(k-1))}{1 + \hat{y}^T(k)R^{-1}(k-1)\hat{y}(k)} \hat{y}(k), \\ R^{-1}(k) = R^{-1}(k-1) - \frac{R^{-1}(k-1)\hat{y}(k)\hat{y}^T(k)R^{-1}(k-1)}{1 + \hat{y}^T(k)R^{-1}(k-1)\hat{y}(k)} \end{cases} \quad (16)$$

где $c^*(k)$ – оценка рекуррентного метода наименьших квадратов, получаемая с помощью последних двух соотношений (16).

Заметим также, что положительным значениям параметров $c_j(k)$ можно придать смысл вероятностей использования отдельных прогнозов $\hat{y}_j(k)$ при построении результирующего прогноза $\hat{y}^*(k)$. При этом можно ввести в рассмотрение текущую информативность j -й прогнозирующей модели в виде

$$I_j(k) = \frac{1}{c_j(k)} \ln \frac{1}{c_j(k)}.$$

Анализируя полученные соотношения, можно заметить, что выражение (15) следует использовать совместно с (3), а (16) – с (14). Что же касается одношаговых адаптивных алгоритмов идентификации типа (6)–(9) и им подобных – соответствующие адаптивные процедуры комбинирования прогнозов неизвестны.

Введем критерий идентификации

$$J^A = (y(k) - c^T(k)\hat{y}(k))^2 = (y(k) - \hat{y}^*(k))^2$$

и с учетом ограничений (11) Лагранжа

$$L^A = (y(k) - c^T(k)\hat{y}(k))^2 + \lambda(c(k)I - 1).$$

Решая далее систему уравнений Куна-Таккера

$$\begin{cases} \frac{\partial L^A}{\partial \lambda} = c^T(k)I - 1 = 0, \\ \bar{V}_c L^A = -2(y(k) - c^T(k)\hat{y}(k))\hat{y}(k) + \lambda I = 0 \end{cases}$$

получаем

$$c(k) = \frac{\hat{y}(k)y(k)}{\|\hat{y}(k)\|^2} + \lambda \frac{\hat{y}(k)\hat{y}^T(k)I}{2\|\hat{y}(k)\|^4} \quad (17)$$

а с учетом того, что

$$I^T c(k) = I^T \frac{\hat{y}(k)y(k)}{\|\hat{y}(k)\|^2} + I^T \frac{\hat{y}(k)\hat{y}^T(k)I}{2\|\hat{y}(k)\|^4} \lambda = 1$$

находим

$$\lambda = 2 \frac{\|\hat{y}(k)\|^4 - I^T \hat{y}(k)y(k)\|\hat{y}(k)\|^2}{I^T \hat{y}(k)\hat{y}^T(k)I}. \quad (18)$$

Подставляя далее (16) в (15), окончательно получаем простое выражение

$$c(k) = \frac{\hat{y}(k)}{I^T \hat{y}(k)} \quad (19)$$

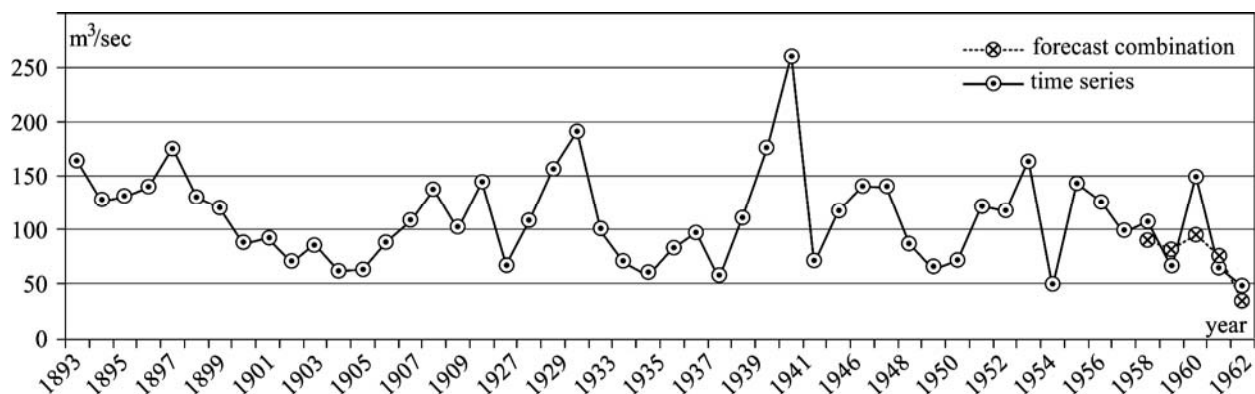
или

$$c_j(k) = \frac{\hat{y}_j(k)}{\sum_{l=1}^m \hat{y}_l(k)}.$$

Экспериментальные исследования

Эксперимент проводился на реальных данных, описывающих сток реки Северский Донец с 1893 по 1962 год (не учитывая 1930, 1931 и 1942, 1943 годов) [1]. В качестве обучающей выборки было выбрано 45 точек, а проверочной – 5 точек. Для прогнозирования использовались AR-модели второго и третьего порядка, а для настройки их параметров – рекуррентный метод наименьших квадратов в (4) и аддитивная модификация алгоритма Качмаша (7) с различными значениями параметра. В качестве оценки точности прогнозов использовалась относительная ошибка прогноза за 5 лет Δ . Результаты прогнозирования, полученные с помощью этих моделей, сравнивали с результатами, полученными с помощью метода группового учета аргументов [1]. Точность результатов, приведенных в [1], лежит в интервале $0,32 < \Delta < 0,65$. Точность полученных нами прогнозов лежит в интервале $0,2 < \Delta < 0,64$, при этом результаты комбинирования позволили получить устойчивый результат на уровне 0,4.

Как видно, оценки точности МГУА-моделей и использованных адаптивных моделей имеют один порядок, однако, с вычислительной точки зрения, сочетание адаптивного подхода с адаптивным комбинированием (19) резко упрощает процесс построения прогнозов.



Сток реки Северский Донец

Заключение

Рассмотрена задача предсказания нестационарных временных рядов на основе системы моделей, отличающихся как структурой, так и методами оценивания параметров. Введена численно простая адаптивная процедура комбинирования прогнозов, позволяющая синтезировать наилучшую с точки зрения принятого критерия оценку. Результаты подтвердили свою эффективность при решении задачи экологического мониторинга.

1. Иваненко А.Г. Долгосрочное прогнозирование и управление сложными системами. – К.: Техника, 1975. – 312 с.
2. Nalimov, V.V. Analysis of the fundamentals of ecological prediction. *Man and Nature*. – 1983. – № 8. – 31–47.
3. Zanetti P. 1990 Air Pollution Modelling. – N.Y. Van Nostrand Reinhold, 1990-302p.
4. Abou-Loukh S., Sinha A.K. Self-tuning predictors with application to river flow prediction. // *Int. J. Syst. Sci.* 1986. – №4, P. 655-668.
5. Perez P.; Trier A.; Reyes J. Prediction of PM_{2.5} concentrations several hours in advance using neural networks in Santiago, Chile // *Atmospheric Environment*. – 2000. – 34. – p.1189-1196.
6. Corani Y. Air quality prediction in Milan: feed ward neural networks, pruned neural network and lazy learning // *Ecological Modeling*. – 2005. – 185. – P.513–529.
7. Lira T.S.; Barrozo M.A.S.; Assis A.J.. Air quality prediction in Uberlandia, Brazil, using linear models and Neural Networks // *17th European Symposium on Computer Aided Process Engineering* 17. – Elsevier B.V., 2007 – P.1–6.
8. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Вып.1. – М.: Мир, 1974. – 406 с.
9. Кашиян Р.Л., Рао А.Р. Построение динамических стохастических моделей по экспериментальным данным. – М.: Наука, 1983. – 383 с.
10. Nelles O. *Nonlinear Systems Identification* – Berlin: Springer, 2001. – 785 p.
11. Грон Д. Методы идентификации систем – М.: Мир, 1979. – 302 с.
12. Перельман И.И. Оперативная идентификация объектов управления. – М.: Энергоиздат, 1982. – 272 с.
13. Райбман Н.С., Чадеев В.М. Построение моделей процессов производства – М.: Энергия, 1975. – 376 с.
14. Льюнг Л. Идентификация систем. Теория для пользователя. – М.: Наука, 1991. – 432 с.
15. Dickinson J.P. Some comments on the combination of forecast // *Oper.Res.Quart.* – 1975. – 26. – №2. – P.205–210.
16. Бодянский Е.В., Плисс И.П., Соловьева Т.В. Адаптивное обобщенное прогнозирование многомерных случайных последовательностей // *Доклады АН УССР*. – 1989. – Сер. А. – № 9. – С. 73–75.
17. Hansen L.K., Salamon P. Neural networks Ensembles // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 1990. – 12. – P.993–1000.
18. Kowalczyk L. Competitive identification for self-tuning control // *Automatica*. – 1992–28. - 2. - P.193–201.
19. Sharkey A.J.C. On combining artificial neural nets // *Connect.Sci.* – 1996. – 8. – P.299-313.
20. Naftaly U., Intrator N., Horn D. Optimal Ensemble Averaging of Neural Networks // *Network: Comput. Neural Syst.* – 1997. – 8. – 283–296.
21. Бодянский Е.В., Воробьев С.А. Рекуррентная нейронная сеть для обнаружения изменений свойств нелинейных стохастических последовательностей // *Автоматика и телемеханика*. – 2000. – № 7. – С.55–67.