УДК 519.72

© 2016 г. **А.В. Лапко**, д-р техн. наук,
В.А. Лапко, д-р техн. наук
(Институт вычислительного моделирования СО РАН, Красноярск,
Сибирский государственный аэрокосмический университет имени
академика М.Ф. Решетнева, Красноярск)

СТАТИСТИЧЕСКИЕ МОДЕЛИ ПРОСТРАНСТВЕННО РАСПРЕДЕЛЕННЫХ ВРЕМЕННЫХ ПРОЦЕССОВ*

Рассматривается методика построения статистических моделей пространственно распределенных временных процессов, часто встречающихся при исследовании экологических систем. Анализируются особенности их синтеза и оптимизации в условиях однородной и неоднородной сред.
Ключевые слова: пространственно распределенные системы, статистические модели, временные процессы, непараметрические модели, коллективы решающих правил.

DOI: 10.22250/isu.2016.50.69-75

Введение

Исследуется пространственно распределенный временной процесс

$$y(x(t), z) = \psi(y(t-1), x(t), z), \qquad (1)$$

динамика выходной переменной $y(x(t), z) \in R^1$ которого в точке с координатами $z \in R^2$ в момент времени t определяется ее предыдущим значением $y(t-1)$ и значениями факторов $x(t) \in R^k$. Вид преобразования $\psi(y(t-1), x(t), z)$ априори неизвестен. Исходную информацию составляют статистические данные $V_j = \left(x(t), y(t), t = \overline{1, n_j}\right)$, $j = \overline{1, n}$ о динамике процесса (1) в n точках пространственных координат $\left(z_1^j, z_2^j\right)$, $j = \overline{1, n}$. Подобные условия часто встречаются, например, при моделировании экологических систем по данным дистанционного зондирования.

С позиций принципов коллективного оценивания и методов непараметрической статистики предлагается «обход» проблемы математического моделирования однородных и неоднородных пространственно распределенных временных

* Работа выполнена в рамках проектной части государственного задания Минобрнауки РФ № 2.914.2014/К и Программы СО РАН IV.35.1.

процессов в условиях априорной неопределенности.

Идея предлагаемого подхода состоит в построении статистических моделей временных процессов

$$y\left(x(t), z^j\right)=\psi\left(y(t-1), x(t), z^j\right)=\varphi_j\left(y(t-1), x(t)\right), \quad j=\overline{1, n}, \qquad (2)$$

соответствующих пространственным координатам z^j, $j=\overline{1, n}$ наблюдения за процессом (1) и последующей их интеграцией в единую решающую функцию с использованием принципов коллективного оценивания.

Непараметрическая модель временного процесса коллективного типа

Для восстановления зависимостей (2) будем применять методику синтеза непараметрических моделей временных процессов коллективного типа, обеспечивающих наиболее полный учет информации обучающих выборок V_j на основе сочетания преимуществ параметрических и локальных аппроксимаций [1 – 3].

Поставим в соответствие каждому наблюдению $\left(x^\tau=x(\tau), y^\tau=y(\tau)\right)\in V_j$ часть исходной выборки $V_j(\tau)=\left(x^t, y^t, t=\overline{\tau+1, n_j}\right)$ и линейный полином $\overline{\varphi}_j^\tau\left(y(t-1), x(t), \alpha_j^\tau\right)$, параметры которого удовлетворяют условиям

$$y^\tau=\overline{\varphi}_j^\tau\left(y(\tau-1), x(\tau), \overline{\alpha}_j^\tau\right), \ \overline{\alpha}_j^\tau=\underset{\alpha_j^\tau}{\operatorname{argmin}}\left(n_j-\tau\right)^{-1}\sum_{\left(y^{i-1}, x^i\right)\in V_j(\tau)}\left(y^i-\overline{\varphi}_j^\tau\left(y^{i-1}, x^i, \alpha_j^\tau\right)\right)^2.$$

Упрощенные аппроксимации $\overline{\varphi}_j^\tau\left(y(t-1), x(t), \overline{\alpha}_j^\tau\right)$ проходят через наблюдения $\left(y^{\tau-1}, x^\tau, y^\tau\right)$ и близки в среднеквадратическом к элементам выборки $V_j(\tau)$, $\tau=\overline{1, n_j-(k+1)}$. Здесь n_j – количество элементов выборки $V_j(\tau)$.

Графическая иллюстрация на примере построения последовательности упрощенных аппроксимаций $\overline{\varphi}_j\left(t, \overline{\alpha}_j\right)$, $j=\overline{1, (n-1)}$ при восстановлении зависимости $\varphi(t)$ приведена на рис. 1.

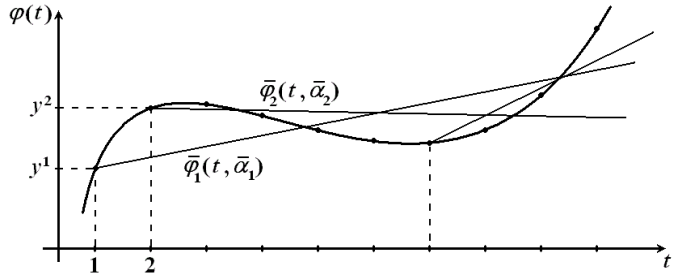

Рис. 1. Иллюстрация формирования последовательности упрощенных аппроксимаций временного процесса $\varphi(t)$.

Исходную информацию в этом случае составляют статистические данные $V=\left(t, y^j=\varphi_j(t), t=\overline{1, n}\right)$. Количество упрощенных аппроксимаций равно $n-1$.

На основе полученной системы «опорных» функций $\overline{\varphi}_j^\tau\left(y(t-1), x(t), \overline{\alpha}_j^\tau\right)$, $\tau=\overline{1, n_j-(k+1)}$ построим модель зависимости $y\left(x(t), z^j\right)=\varphi_j\left(y(t-1), x(t)\right)$ в ви-

де непараметрической оценки условного математического ожидания [4]

$$\bar{\bar{y}}_j = \bar{\bar{\varphi}}_j\left(y(t-1), x(t)\right) = \sum_{i=1}^{n_j-(k+1)} \bar{\varphi}_j^i\left(y(t-1), x(t), \bar{\alpha}_j^i\right) \beta_i\left(y(t-1), x(t)\right), \quad j = \overline{1, n}, \quad (3)$$

где $\beta_i\left(y(t-1), x(t)\right) = \dfrac{\Phi\left(\dfrac{y(t-1)-y^i}{c_0}\right) \prod\limits_{v=1}^{k} \Phi\left(\dfrac{x_v(t)-x_v^i}{c_v}\right)}{\sum\limits_{i=1}^{n_j-(k+1)} \Phi\left(\dfrac{y(t-1)-y^i}{c_0}\right) \prod\limits_{v=1}^{k} \Phi\left(\dfrac{x_v(t)-x_v^i}{c_v}\right)}$.

Ядерные функции $\Phi(\cdot)$, удовлетворяющие условиям положительности, симметричности и нормированности [5, 6]:

$$\Phi(u) = \Phi(-u), \quad 0 \le \Phi(u) < \infty,$$

$$\int \Phi(u)du = 1, \quad \int u^2\, \Phi(u)du = 1,$$

$$\int u^m\, \Phi(u)du < \infty, \quad 0 \le m < \infty.$$

Положительные коэффициенты размытости $c_0 = c_0(n_j)$, $c_v = c_v(n_j)$, $v = \overline{1, k}$ ядерных функций с ростом n_j стремятся к нулю. Здесь бесконечные пределы интегрирования опускаются.

Непараметрическая модель коллективного типа (3) характеризуется высоким уровнем помехозащищенности, что обеспечивается двойным сглаживанием при формировании системы упрощенных аппроксимаций $\bar{\varphi}_j^{\tau}\left(y(t-1), x(t), \bar{\alpha}_j^{\tau}\right)$, $\tau = \overline{1, n_j-(k+1)}$ и путем их усреднения в соответствии с процедурой (3).

В условиях рис. 1 непараметрическая модель (3) в пункте наблюдения z_j запишется в виде $\bar{\bar{y}}_j = \bar{\bar{\varphi}}_j(t) = \sum\limits_{i=1}^{n_j-1} \bar{\varphi}_j^i\left(t, \bar{\alpha}_j^i\right) \beta_i(t)$, $\beta_i(t) = \dfrac{\Phi\left(\dfrac{t-i}{c}\right)}{\sum\limits_{\tau=1}^{n_j-1} \Phi\left(\dfrac{t-\tau}{c}\right)}$.

Оптимизация непараметрической модели (3) по параметрам размытости $c_0 = c\,\bar{\sigma}_{t-1}$, $c_v = c\,\bar{\sigma}_v$ ($\bar{\sigma}_{t-1}$, $\bar{\sigma}_v$, $v = \overline{1, k}$ – оценки среднеквадратических отклонений $y(t-1)$ и $x_v(t), v = \overline{1, k}$) осуществляется из условия минимума эмпирического критерия

$$\bar{\bar{W}}_j(c) = \frac{1}{n_j} \sum_{i1=1}^{n_j} \left(y^{i1} - \bar{\bar{\varphi}}_j\left(y(i1-1), x(i1)\right)\right)^2, \quad (4)$$

отражающего меру близости между экспериментальными данными y^{i1} и результатами их оценивания по модели (3). Выбор параметра c в непараметрической модели (3) осуществляется в режиме «скользящего экзамена». Ситуация $\left(y^{i1-1}, x^{i1}, y^{i1}\right)$, представляемая на контроль, исключается из процесса обучения в процедуре (3), что выполняется при $i \ne i1$.

Статистическая модель однородного пространственно распределенного временного процесса

Для построения статистической модели процесса (1) в однородных условиях воспользуемся оператором условного математического ожидания в пространстве $z=(z_1, z_2)$, непараметрическая оценка которого имеет вид

$$\bar{y}(x(t), z)=\bar{\psi}(y(t-1), x(t), z)=\frac{\sum_{j=1}^{n}\bar{\bar{\varphi}}_j(y(t-1), x(t))\,\beta^j(z, z^j)}{\sum_{j=1}^{n}\beta^j(z, z^j)}, \quad (5)$$

$$\beta^j(z, z^j)=\prod_{v=1}^{2}\Phi\left(\frac{z_v-z_v^j}{c_v}\right) \quad (6)$$

– двумерная ядерная функция в пространстве $z=(z_1, z_2)$.

Для повышения аппроксимационных свойств непараметрической модели (5) целесообразно ввести в структуру ядерной функции (6) показатели эффективности $\bar{\bar{W}}_j$, $j=\overline{1,n}$ статистики (3), т.е.

$$\beta^j(z, z^j)=\prod_{v=1}^{2}\Phi\left(\frac{z_v-z_v^j}{c_v}\right)\Phi\left(\frac{0-\bar{\bar{W}}_j}{c}\right). \quad (7)$$

Дополнительная ядерная функция определяется в виде

$$\Phi\left(\frac{0-\bar{\bar{W}}_j}{c}\right)=\begin{cases}0, \text{если} & \bar{\bar{W}}_j>c,\\ \frac{1}{2}, \text{если} & \bar{\bar{W}}_j\le c.\end{cases}$$

Ядерная функция (7) позволяет компенсировать влияние «неудачных» моделей (3) в условиях $(y(t-1), x(t), z)$, тем самым повышается устойчивость и помехозащищенность процедуры формирования решений (5).

Оптимизация обобщенной модели однородного пространственно распределенного временного процесса (5) по коэффициентам размытости ядерных функций (6), (7) осуществляется из условия минимума статистического критерия

$$\bar{W}(c)=n^{-1}\sum_{j=1}^{n}n_j^{-1}\sum_{t=1}^{n_j}\left(\bar{\bar{\varphi}}_j(y(t-1), x(t))-\bar{\psi}(y(t-1), x(t), z^j)\right)^2, \quad (8)$$

который определяет оценку среднеквадратической меры близости между частными моделями $\bar{\bar{\varphi}}_j(\cdot)$ при различных условиях $(y(t-1), x(t), z^j)$, $t=\overline{1,n_j}$, $j=\overline{1,n}$ и их непараметрическим коллективом $\bar{\psi}(y(t-1), x(t), z)$ (5).

Анализ статистической модели (5). Обозначим параметры в линейной аппроксимации $\bar{\varphi}_j^i(y(t-1), x(t), \bar{\alpha}_j^i)$, соответствующие переменным $y(t-1)$, $x_v(t)$, $v=\overline{1,k}$ через $\bar{\tilde{\alpha}}_j^i$. Пусть свободный член этой модели принимает значение $\tilde{\alpha}_j^i$. Тогда из условий формирования модели $\bar{\varphi}_j^i(y(t-1), x(t), \bar{\alpha}_j^i)$ следует, что

$$\tilde{\alpha}_j^i = y^i - \tilde{\bar{\varphi}}_j^i\left(y(t-1), x(t), \tilde{\bar{\alpha}}_j^i\right),$$

где $\tilde{\bar{\varphi}}_j^i\left(y(t-1), x(t), \tilde{\bar{\alpha}}_j^i\right)$ определяет линейную аппроксимацию $\bar{\varphi}_j^i\left(y(t-1), x(t), \bar{\alpha}_j^i\right)$, без учета свободного члена $\tilde{\alpha}_j^i$. В принятых обозначениях статистическая модель временного процесса коллективного типа (3) запишется в виде

$$\tilde{\bar{\varphi}}_j\left(y(t-1), x(t)\right) = \sum_{i=1}^{n_j-(k+1)} y^i \beta_i\left(y(t-1), x(t)\right) -$$

$$- \sum_{i=1}^{n_j-(k+1)} \tilde{\bar{\varphi}}_j^i\left(y(t-1), x(t), \tilde{\bar{\alpha}}_j^i\right)\beta_i\left(y(t-1), x(t)\right), \quad j = \overline{1, n}. \qquad (9)$$

Тогда статистическая модель (5) преобразуется к форме

$$\bar{\psi}\left(y(t-1), x(t), z\right) = \sum_{j=1}^{n}\sum_{i=1}^{n_j-(k+1)} y^i \beta_i\left(y(t-1), x(t)\right)\bar{\beta}^j\left(z, z^j\right) -$$

$$- \sum_{j=1}^{n}\sum_{i=1}^{n_j-(k+1)} \tilde{\bar{\varphi}}_j^i\left(y(t-1), x(t), \tilde{\bar{\alpha}}_j^i\right)\beta_i\left(y(t-1), x(t)\right)\bar{\beta}^j\left(z, z^j\right), \qquad (10)$$

где

$$\bar{\beta}^j\left(z, z^j\right) = \frac{\beta^j\left(z, z^j\right)}{\sum_{v=1}^{n}\beta^v\left(z, z^v\right)}.$$

Нетрудно заметить, что первое слагаемое выражения (10) отражает информацию, содержащуюся в точках исходных статистических данных, а второе – взаимосвязи между временными наблюдениями в конкретных пространственных координатах.

Таким образом, в одном решающем правиле сочетаются возможности параметрических и локальных методов аппроксимации. Данный подход позволяет в условиях априорной неопределенности максимально учитывать исходную статистическую информацию.

Синтез статистической модели в неоднородной среде с «непрозрачными» границами

Будем полагать наличие в пространстве координат z областей Ω_λ, $\lambda = \overline{1, M}$, в пределах которых вид зависимости (1) не зависит от z, т.е.

$$y\left(y(t-1), x(t), z\right) = \psi_\lambda\left(y(t-1), x(t)\right) \quad \forall \quad z \in \Omega_\lambda, \ \lambda = \overline{1, M}. \qquad (11)$$

Модель подобного неоднородного пространственно распределенного временного процесса (9) имеет двухуровневую структуру (рис. 2).

На первом уровне структуры модели в соответствии с решающим правилом $m(z)$ определяется принадлежность ситуации $\left(y(t-1), x(t), z\right)$ к конкретной области однородности Ω_λ изучаемого процесса. На втором уровне на основе статистической модели $\bar{\psi}_\lambda\left(y(t-1), x(t), z\right)$ оценивается значение выходной переменной $\bar{y}\left(y(t-1), x(t), z\right)$.

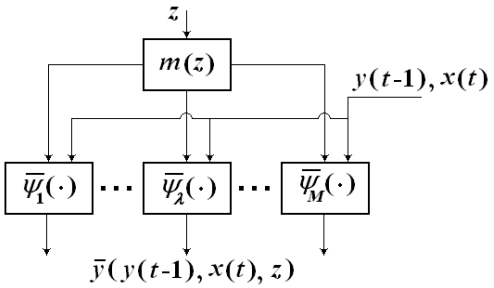

Рис. 2. Структура модели неоднородного пространственно распределенного временного процесса.

При построении статистической модели $\bar{\psi}_\lambda\left(y(t-1), x(t), z\right)$ используется методика, изложенная при описании непараметрических моделей однородных пространственно распределенных временных процессов. Для этого исходная информация $V_j = \left(x(t), y(t), t = \overline{1, n_j}\right),\ \ j = \overline{1, n}$ предварительно разбивается на M групп $V(\lambda), \lambda = \overline{1, M}$ наблюдений переменных временного процесса.

Пусть $I(\lambda)$ – множество номеров элементов исходной выборки из области однородности Ω_λ. Тогда статистическая модель $\psi_\lambda\left(y(t-1), x(t)\right)$ имеет вид

$$\bar{\psi}_\lambda\left(y(t-1), x(t), z\right) = \frac{\sum\limits_{j \in I(\lambda)} \bar{\bar{\varphi}}_j\left(y(t-1), x(t)\right) \beta^j\left(z, z^j\right)}{\sum\limits_{j \in I(\lambda)} \beta^j\left(z, z^j\right)},\ \lambda = \overline{1, M}. \qquad (12)$$

В условиях сложных границ между областями однородности рекомендуется использовать непараметрические алгоритмы распознавания образов $\bar{m}(z)$ [7].

Построение статистической модели в неоднородной среде с частично «прозрачными» границами

Рассмотрим условия неоднородной среды, когда допускается взаимное влияние смежных процессов (9). Будем считать, что $M = 2$, т.е. имеются две однородные области Ω_1, Ω_2 в пространстве z. Для конкретных условий $\left(z, z^j\right)$ введем функцию

$$h\left(\frac{z_v - z_v^j}{c_v(\gamma)}\right) = \begin{cases} 1, \textit{если} \quad \left|z_v - z_v^j\right| \le c_v(1)\ \forall\ \left(z_v, v = 1, 2\right) \in \Omega_1\ u\ \left(z_v^j, v = 1, 2\right) \in \Omega_1 \\ r_2, \textit{если} \quad \left|z_v - z_v^j\right| \le c_v(1)\ \forall\ \left(z_v, v = 1, 2\right) \in \Omega_1\ u\ \left(z_v^j, v = 1, 2\right) \in \Omega_2 \\ 0, \textit{если} \quad \left|z_v - z_v^j\right| > c_v(1)\ \forall\ \left(z_v, v = 1, 2\right) \in \Omega_1\ u\ \left(z_v^j, v = 1, 2\right) \in \Omega_1 \\ \qquad \textit{либо}\ \left(z_v^j, v = 1, 2\right) \in \Omega_2, \end{cases}$$

которая определяет степень влияния временного процесса $\varphi_j\left(y(t-1), x(t)\right)$ из области однородности Ω_2 на формирование решения в контрольной ситуации $\left(y(t-1), x(t), z\right)$ из области $z \in \Omega_1$.

Тогда непараметрическая модель пространственно распределенного вре-

менного процесса в Ω_1 с частично «прозрачными» границами запишется в виде статистики

$$\tilde{\psi}_1\left(y(t-1),\, x(t),\, z\right) = \frac{\sum\limits_{j=1}^{n} \overline{\overline{\varphi}}_j\left(y(t-1),\, x(t)\right) \beta^j\left(z,\, z^j\right) \prod\limits_{v=1}^{2} h\left(\dfrac{z_v - z_v^j}{c_v(1)}\right)}{\sum\limits_{j=1}^{n} \beta^j\left(z,\, z^j\right) \prod\limits_{v=1}^{2} h\left(\dfrac{z_v - z_v^j}{c_v(1)}\right)}. \qquad (13)$$

Оптимальные параметры $c_v(1)$, $v = 1, 2$ и $r_2 \le 1$ непараметрической модели находятся из условия минимума критерия типа (8).

Заключение

С позиций принципов коллективного оценивания и методов непараметрической статистики предлагаются модели пространственно распределенных временных процессов в условиях априорной неопределенности, когда исходная информация представлена наблюдениями временной динамики параметров изучаемой системы из некоторого множества точек пространственных координат. Рассмотрены особенности синтеза непараметрических моделей исследуемого класса систем в неоднородных средах, которые представляются в виде коллектива решающих правил с учетом их взаимодействия. Предлагаемые модели учитывают не только информацию, содержащуюся в элементах исходных статистических данных, но и взаимосвязи между ними. Полученные результаты могут быть использованы при исследовании пространственно распределенных экологических систем в условиях априорной неопределенности.

ЛИТЕРАТУРА

1. *Лапко А.В., Лапко В.А.* Обобщенная непараметрическая регрессия и ее свойства // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2013. – №3 (24). – С.95-103.
2. *Лапко В.А.* Синтез и анализ непараметрических моделей коллективного типа // Автометрия. – 2001. – №6. – С.98-106.
3. *Лапко В.А.* Непараметрические модели временных зависимостей, основанные на методе двойного коллективного оценивания // Автометрия. – 2002. – №1. – С. 42-50.
4. *Хардле В.* Прикладная непараметрическая регрессия. – М.: Мир, 1993.
5. *Parzen E.* On estimation of a probability density function and mode // Ann. Math. Statistic. – 1962. – Vol.33. – P. 1065-1076.
6. *Епанечников В.А.* Непараметрическая оценка многомерной плотности вероятности // Теория вероятности и ее применения. – 1969. – Т.14, №1. – С. 156-161.
7. *Лапко А.В.,. Лапко В.А, Соколов М.И., Ченцов С.В.* Непараметрические. системы классификации. – Новосибирск: Наука, 2000.

E-mail:
Лапко Александр Васильевич – lapko@icm.krasn.ru;
Лапко Василий Александрович – lapko@icm.krasn.ru.