UDC 628.5.66

# CREATION OF MATHEMATICAL MODELS OF COMPLEX PROCESSES BASED ON PRINCIPLES OF GENETIC ALGORITHMS

M. Gorbiychuk[1], Doctor of Technical Sciences, Professor
M. Shufnarovych,[2] assistant
Ivano-Frankivsk National Technical University of Oil and Gas, Ukraine[1]
Ivano-Frankivsk State Medical University, Ukraine[2]

A method of creation of mathematical models of complex processes based on principles of genetic algorithms is developed. This method allows synthesizing mathematical models of any level without preliminary selection of the number from selection rows. The method will be useful in improvement of oscillatory processes predictions accuracy, such as the level of the Dniester river.

**Keywords:** synthesis of mathematical models, genetic algorithm, selection criterion, population, chromosome.

Conference participant, National championship in scientific analytics

УДК 628.5.66

# ПОСТРОЕНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ СЛОЖНЫХ ПРОЦЕССОВ НА ПРИНЦИПАХ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Горбийчук М.И[1]., д-р техн. наук, проф.
Шуфнарович М.А[2]., ассистент
Ивано-Франковский национальный технический университет нефти и газа, Украина[1]
Ивано-Франковский национальный медицинский университет, Украина[2]

Разработан метод построения математических моделей сложных процессов на принципах генетических алгоритмов. Данный метод дает возможность синтезировать математические модели любой сложности без предварительного выбора числа рядов селекции. Метод найдет применение в повышении точности прогнозов колебательных процессов, например уровня реки Днестр.

**Ключевые слова:** синтез математических моделей, генетический алгоритм, критерий селекции, популяция, хромосома.

Участники конференции, Национального первенства по научной аналитике

Целому ряду процессов экономического и экологического характера присуща гармоническая составляющая с некратными частотами, которая моделируется выражением:

$$G(t) = A_0 + \sum_{j=1}^{m}\left(A_j\,sin(t\omega_j) + B_j\,cos(t\omega_j)\right), \quad (1)$$

где $t$ - такты отсчета времени, $t = 1,2,3,...;N A_0, A_j, B_j$ - параметры гармонического ряда (1); $\omega_j = \omega_{j-1} + \Delta\omega_j$ - некратные частоты, $j = 1,2,3,...$

Для оценки параметров ряда (1) необходимо выполнение условия [1] $N \geq 3m+1$. Идентификация параметров модели (1) происходит в несколько этапов [2]. Первый этап - вычисление балансовых коэффициентов $\alpha_p$ из условия минимизации невязки

$$B = \sum_{i=m+1}^{N-m} b_i^2, \quad (2)$$

где

$$b_i = g(i+m) - \sum_{p=0}^{m-1}\alpha_p\left(g(i+p)+\right.$$
$$\left.+\,g(i-p)\right) + g(i-m)$$

$i = \overline{m+1, N-m}$; $g(t)$- отсчеты

реализации процесса в моменты времени, симметрично размещены относительно произвольной точки $i$.

Итак, будем решать задачу

$$\min_{\overline{\alpha}}: J(\overline{\alpha}) = \sum_{i=m+1}^{N-m}\left(z_{i,m} - \sum_{p=0}^{m-1}\alpha_p g_{i,p}\right)^2, (3)$$

где $\overline{\alpha} = (\alpha_0, \alpha_1, ..., \alpha_{m-1})^T$ - вектор весовых коэффициентов; $z_{i,m} = \overline{g}(i+m) + \overline{g}(i-m)$; $g_{i,p} = \overline{g}(i+p) + \overline{g}(i-p)$; $T$- символ транспонирования матриц.

Задачу (3) запишем в матрично-векторной форме

$$\min_{\overline{\alpha}}: J(\overline{\alpha}) = (\overline{z}_m - F_m\overline{\alpha})^T(\overline{z}_m - F_m\overline{\alpha}), (4)$$

Где; .

$$\overline{z}_m = \begin{bmatrix} \overline{g}(2m+1) + \overline{g}(1) \\ \overline{g}(2m+2) + \overline{g}(2) \\ ... \\ \overline{g}(N) + \overline{g}(N-2m) \end{bmatrix}$$

$$F_m = \begin{bmatrix} 2\overline{g}(m+1) & \overline{g}(m+2)+\overline{g}(m) & \overline{g}(m+3)+\overline{g}(m-1) & ... & \overline{g}(2m)+\overline{g}(2) \\ 2\overline{g}(m+2) & \overline{g}(m+3)+\overline{g}(m+1) & \overline{g}(m+4)+\overline{g}(m) & ... & \overline{g}(2m+1)+\overline{g}(3) \\ ... & ... & ... & ... & ... \\ 2\overline{g}(N-m) & \overline{g}(N-m+1)+\overline{g}(N-m-1) & \overline{g}(N-m+2)+\overline{g}(N-m-2) & ... & \overline{g}(N-1)+\overline{g}(N-2m+1) \end{bmatrix}$$

Минимизация выражения (4) приводит к нормальному уравнению Гаусса

$$F_m^T F_m \overline{\alpha} = F_m^T \overline{z}_m. \quad (5)$$

Из последнего уравнения можно найти

$$\overline{\alpha} = \left(F_m^T F_m\right)^{-1} F_m^T \overline{z}_m. \quad (6)$$

Использовать формулу (6) можно лишь тогда, когда размерность вектора $\overline{\alpha}$ невелика и матрица $F_m^T F_m$ является хорошо обусловленной. Если это условие не выполняется, то для нахождения $\overline{\alpha}$ следует решать уравнения (5) одним из численных методов [3].

На втором этапе составляется уравнение

$$\alpha_0 + \sum_{p=1}^{m-1}\alpha_p\,cos(p\omega) = cos(m\omega)$$

для произвольной частоты $\omega$, которое с помощью рекуррентного соотношения [1] $cos(p\omega) = 2cos((p-1)\omega)cos\omega - cos((p-2)\omega)$, $p = \overline{2,m}$ приводится к алгебраическому уравнению $m$-той степени относительно $z = cos\omega$

$$P_m z^m + P_{m-1}z^{m-1} + ... + P_1 z + P_0 = 0. (7)$$

Суть третьего этапа заключается в оптимальном синтезе гармонического ряда (1).

Предложен подход построения математических моделей, основанный на идеях генетических алгоритмов. Вся реализация исходной величины процесса или явления $N$ разбивается на три части в такой пропорции [2]: $N_R = 0.7N$, $N_Q = 0.2N$ и $N_R = 0.1N$. Для множества данных $N_R + N_Q$ определяются весовые коэффициенты $\alpha_p$ из уравнения (4) по методу исключения Гаусса с выбором главного элемента [3]. Решение уравнения (7) дает возможность найти частоты $\omega_j$, $j=\overline{1,m}$. Тогда, на множестве точек $N_R+N_Q$ необходимо найти параметры модели (1) $A_0$, $A_j$, $B_j$ [7]. Создадим хромосому длиной *m,* в которой на *i*-том месте будет стоять ноль или единица в зависимости от того частота $\omega_j$ изъята из выбранного полного ряда $m$ или оставлена. Набор хромосом образует популяцию. В задачи синтеза моделей колебательных процессов функцией приспособленности, которая позволяет выбрать наиболее приспособленные особы из популяции, выступает комбинированный критерий селекции [5]

$$\rho=\sqrt{n_d^2+B^2} \tag{8}$$

где $n_d^2$ - критерий смещения, который вычисляется по формуле:

$$n_d^2=\frac{\sum_{i=1}^{N}\left(g_i(R)-g_i(S)\right)^2}{\sum_{i=1}^{N}\overline{g}_i^2}, \tag{9}$$

где $B$ - функция невязки (2); $g_i(R)$, $g_i(S)$ - величины, вычислены на множестве точек $N$ по формуле (1), а коэффициенты модели (1) найдены на множествах $N_R+N_Q$ и $N_S$.

Генетический алгоритм состоит из следующих шагов [4].

Ш1. Случайным образом формируется популяция с $I$ особей, каждая из которых является хромосомой длиной $m$.

Ш2. Для каждой хромосомы вычисляется критерий селекции (8).

В соответствии с моделью (1) формируется матрица

$$F=\begin{bmatrix} 1 & \sin\omega_1 & \cos\omega_1 & \sin\omega_2 & \cos\omega_2 & \ldots & \sin\omega_m & \cos\omega_m \\ 1 & \sin(2\omega_1) & \cos(2\omega_1) & \sin(2\omega_2) & \cos(2\omega_2) & \ldots & \sin(2\omega_m) & \cos(2\omega_m) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & \sin(N\omega_1) & \cos(N\omega_1) & \sin(N\omega_2) & \cos(N\omega_2) & \ldots & \sin(N\omega_m) & \cos(N\omega_m) \end{bmatrix}$$

В хромосоме удваиваем единицы и нули потому, что каждой частоте $\omega_j$ соответствует пара коэффициентов $A_j$, $B_j$. В модели (1) всегда присутствует коэффициент $A_0$, то в хромосому на первую позицию добавляем единичный ген. С матрицы $F$ формируют новую матрицу $F_{new}$ путем изъятия тех столбцов с матрицы $F$, что отвечают нулям хромосомы $Б_{d0}$. Матрицу $F_R$ образуют первые $N_R+N_Q$ столбцы матрицы $F_{new}$, а вторую – последние $N_S$. На множествах точек $N_R+N_Q$ и $N_S$ вычисляются ненулевые коэффициенты $A_0$, $A_j$, $B_j$ модели (1) путем решения нормального уравнения Гаусса

$$F_R^T F_R \overline{A}_R = F_R^T \overline{g}_R, \tag{10}$$

$$F_S^T F_S \overline{A}_S = F_S^T \overline{g}_S, \tag{11}$$

$$\overline{A}_R=\left(A_0^{(R)},A_1^{(R)},B_1^{(R)},A_2^{(R)},B_2^{(R)},\ldots,A_{m_1}^{(R)},B_{m_1}^{(R)}\right)^T$$

$$\overline{A}_S=\left(A_0^{(S)},A_1^{(S)},B_1^{(S)},A_2^{(S)},B_2^{(S)},\ldots,A_{m_1}^{(S)},B_{m_1}^{(S)}\right)^T$$

- векторы параметров модели;

$$\overline{g}_R=\left(\overline{g}^{(1)},\overline{g}^{(2)},\ldots,\overline{g}^{(N_R+N_Q)}\right)^T$$

, $\overline{g}_S=\left(\overline{g}^{(1)},\overline{g}^{(2)},\ldots,\overline{g}^{(N_S)}\right)^T$ - векторы экспериментальных данных на множестве точек $N_R+N_Q$ и $N_S$.

По известной совокупностью коэффициентов $\overline{A}_R$ и $\overline{A}_S$ модели (1) на множестве точек $N$ вычисляют $g(R)=F_{new}\overline{A}_R$, $g(S)=F_{new}\overline{A}_S$. Значение критерия селекции находят для каждой хромосомы по формуле (9) и в результате получают множество значений $\rho_i$.

*Ш3.* Определяют

$$\rho_{min}=\min_{i\in M}\rho_i. \tag{12}$$

Если минимальное значение (12) критерия селекции (9) не превосходит некоторого положительного значения $\varepsilon$, то происходит остановка алгоритма. Остановка алгоритма также может произойти в случае, когда его выполнение не приводит к улучшению функции приспособления или в том случае, когда алгоритмом уже выполнено заданное число итераций. После выполнения одной из трех условий из популяции выбирается хромосома $Б^*$, для которой выполняется условие (12). После операции удвоения и присоединения единичного гена получаем - $Б^*_{d0}$. Эта хромосома задает структуру модели оптимальной сложности и формирует матрицу $F^*$. Перерасчет параметров модели (1) осуществляется на множестве всех точек исходного массива данных.

*Ш4.* В данном алгоритме использован метод турнирной селекции. Все хромосомы разбиваются на подгруппы, чаще всего по 2 - 3 особи в каждой, с последующим выбором из каждой образованной подгруппы хромосомы с лучшей приспособленностью.

*Ш5.* Оператор мутации с вероятностью $P_m$ изменяет значение гена в хромосоме на противоположное, то есть с 1 на 0 или с 0 на 1. Оператор скрещивания состоит из двух этапов. На первом этапе выбирается лучшая хромосома из подгруппы особей по критерию селекции. В результате получаем новую популяцию хромосом, в которой применяют оператор скрещивания. После выполнения оператора скрещивания происходит переход к Ш2.

Анализ изменения уровня воды в р. Днестр показывает, что со временем имеет место тренд $h(t)$, который носит линейный характер, и существует гармоническая составляющая $G(t)$ обусловлена сезонным изменением метеорологических условий [6], т.е.

$$\widetilde{H}_t=H_t+G(t)+h(t). \tag{13}$$

Из зависимости (13) был выделен линейный тренд

$$h(t)=\theta_0+\theta_1 t, \tag{14}$$

где $\theta_0$, $\theta_1$ - параметры линейного тренда, найденные по методу наименьших квадратов.

На рис. 1 показан результат работы программы, написанной с использованием разработанного метода в среде MatLab, где знаком «о» отмечены экспериментальные данные, а «+» - результат расчета по формуле (1). После выделения из экспериментальных данных гармонического тренда получили остаток - $H_t=\widetilde{H}_t-(G(t)+h(t))$.

Величина $H_t$ является функцией параметров, определяющих погодные условия, т.е.

$$H_t=\mu(T_t, f_t, f_{t-1}, f_{t-2}, f_{t-3}, \nu_t, p_t), \tag{15}$$

где $T_t$ - среднесуточная температу-

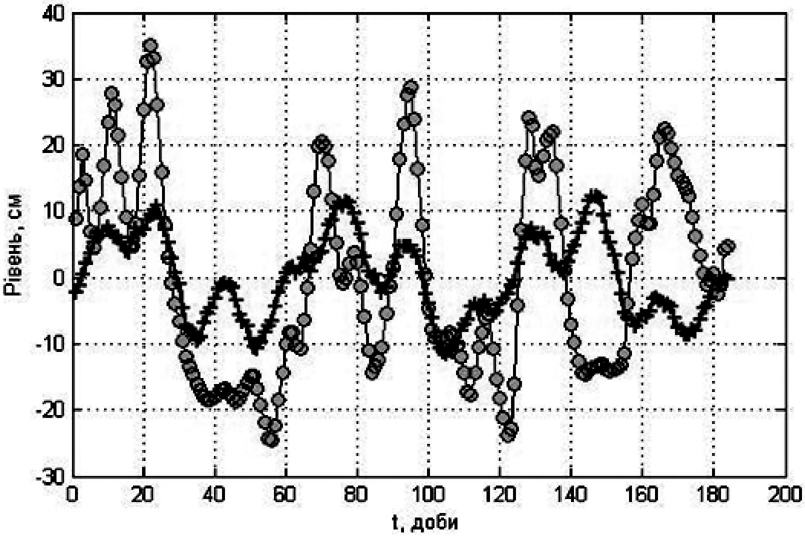


**Рис. 1. Гармонический тренд колеблющегося процесса (р. Днестр)**

ра воздуха, °*C*; $f_t$ - количество осадков, *мм/сутки*; $\nu_t$ - среднесуточная скорость ветра, *м/с*; $p_t$ - среднесуточное барометрическое давление, *мм. рт. ст.*, $t$ - текущее дискретное время, $k = 1,2,3$ - смещение во времени.

Соотношение (15) будем искать в виде полинома

$$y_t = \sum_{i=0}^{M-1} a_i \prod_{j=1}^{k} \left(x_j^{(t)}\right)^{s_j} ,$$

$$t = \overline{1,N}. \qquad (16)$$

где $M = \frac{(m+n)!}{m!\,n!}$ - количество членов полинома (16) [7]; $a_i$ - коэффициенты полинома; $\sum_{j=1}^{n} s_j \le m$ - степени аргументов, $x_j^{(i)}$ - входные величины в каждом наблюдении $t$ ( $x_1 = T_t$, $x_2 = f_t$, $x_3 = f_{t-1}$, $x_4 = f_{t-2}$, $x_5 = f_{t-3}$, $x_6 = \nu_t$, $x_7 = p_t$).

Систему уравнений (16) удобно представить в матрично-векторной форме

$$\overline{y} = F\overline{a} , \qquad (19)$$

где $\overline{y} = (y_1, y_2, ..., y_N)^T$ - вычисленное значение выхода модели (16) в каждой точке наблюдений; $F$ - матрица размером $N \times M$, элементы которой произведения аргументов при параметрах $a_i$; $\overline{a} = (a_0, a_1, ..., a_{M-1})^T$ - вектор параметров модели (16).

Зная $Y_t$ и $y_t$, $i = \overline{1,N}$, можно вычислить критерий аппроксимации

$$J(\overline{a}) = \sum_{t=1}^{N} (Y_t - y_t)^2 .$$ Минимизация критерия аппроксимации приводит к соотношению $F^T F\overline{a} = F^T \overline{Y}$, которое называют нормальным уравнением метода наименьших квадратов. Из этого соотношения можно найти

$$\overline{a} = \left(F^T F\right)^{-1} F^T \overline{Y}. \qquad (22)$$

Использовать формулу (22) можно лишь тогда, когда размерность вектора параметров $\overline{a}$ невелика и матрица $F^T F$ является хорошо обусловленной [8]. Если это условие не выполняется, то для решения уравнения (22) следует использовать один из численных методов, например, метод Гаусса с выбором главного элемента [3].

Как правило, структура модели (16) неизвестна, что приводит к необходимости произвольного выбора как числа функций, так и вида самих функций в модели (16). Поэтому был предложен индуктивный метод самоорганизации моделей [1], идейную сторону которого определяет теорема Геделя.

Для выбора структуры модели используют критерии регулярности

$$\Delta^2(Q) = \frac{\sum_{i=1}^{N_Q} (Y_i - y_i)^2}{\sum_{i=1}^{N_Q} Y_i^2} \qquad (23)$$

и минимума смещения

$$\Delta^2(R,Q) = \frac{\sum_{i=1}^{N} (y_i(R) - y_i(Q))^2}{\sum_{i=1}^{N} Y_i^2} . \qquad (24)$$

Если выбран критерий регулярности (23), то данные эксперимента распределяют [2]: $N_R = 0.7N$ и $N_Q = 0.3N$, а при выборе критерия (24) - $N_R = 0.5N$ і $N_Q = 0.5N$.

Реализация индуктивного метода самоорганизации моделей осуществляется поэтапно: первый этап - генерация моделей-претендентов (в определенном порядке повышения сложности), второй этап - отбор лучшей модели по критерию селекции (23) или (24).

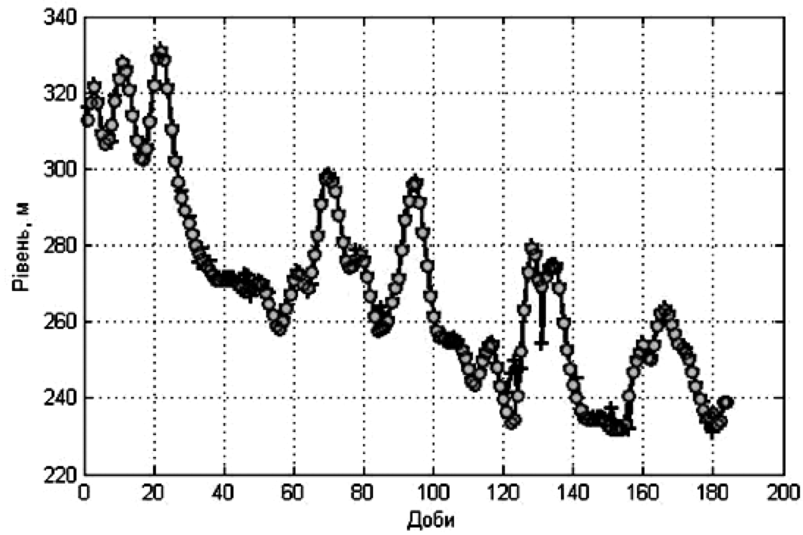


**Рис. 2. Зависимость уровня воды в р. Днестр от погодных условий**

Создадим хромосому длиной $M$, в которой на $i$-том месте будет стоять единица или нуль в зависимости от того, параметр $a_i$, $i=\overline{0,M-}$ модели (16) отличный от нуля или нулевой. С начальной популяции хромосом путем эволюционного отбора нужно выбрать такую, хромосому, которая обеспечивает минимальное значение критерия селекции (23) или (24). Алгоритм решения поставленной задачи аналогичен ранее разработанному с той лишь разницей, что не осуществляется операция удвоение генов в хромосомах.

Найденные зависимости $h(t)$, $G(t)$ и $y_t$ дают возможность найти

$$\hat{H}_t = G(t) + h(t) + y_t \, . \qquad (25)$$

На рис. 2 «+» обозначены вычисление значения по формуле (25), а значком «o» - экспериментальные значения уровня воды в р.Днестр. Адекватность модели проверялась с помощью коэффициента корреляции $K_Y = 0{,}98746$, что свидетельствует о высокой степень корреляции между величинами $Y_i$ и $y^{(i)}$.

Таким образом, применение идей генетических алгоритмов к построению сложных математических моделей дает возможность выбрать оптимальную по структуре адекватную модель и значительно уменьшить объем вычислений.

**References:**


1. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем: монография / А.Г. Ивахненко. – К.: Наукова думка, 1981. – 296 с.
2. Ивахненко А.Г. Справочник по типовым программам моделирования. / А.Г. Ивахненко, Ю. Коппа, В.С. Степашко и др.; под ред. А.Г. Ивахненко – К.: Техніка, 1980. – 180 с.
3. Вержбицкий В.М. Основы численных методов: учебник для вузов / В.М. Вержбицкий. – М.: Высшая школа, 2002. – 840 с.
4. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский; пер. с польск. И.Д. Рудинского. – М.: Горячая линия-Телеком, 2004. – 452 с.
5. Ивахненко А.Г. Помехоустойчивость моделирования: монография / А.Г. Ивахненко, В С. Степашко - Киев: Наук. думка, 1985. - 216 с.
6. Горбійчук М.І. Математичні моделі прогнозування стоку р. Дністер для запобігання техногенних аварій магістральних газопроводів / М.І. Горбійчук, О.В. Пендерецький // Науковий вісник Івано-Франківського національного технічного університету нафти і газу. – 2008. - № 2 (18). – С. 30 – 33.
7. Горбійчук М.І. Індуктивний метод побудови математичних моделей газоперекачувальних агрегатів природного газу / М.І. Горбійчук, М.І. Когутяк, Я.І. Заячук // Нафтова і газова промисловість. – 2008. - № 5. – С. 32 – 35.
8. Ермаков С.М. Математическая теория оптимального эксперимента / С.М. Ермаков, А.А. Жиглявский: учеб. пособие. – М.: Наука, 1987. – 320 с.