УДК 519.6

О. В. Семенив, В. А. Яценко

Інститут космічних досліджень Національної академії наук України та Національного космічного агентства України, Київ

ИДЕНТИФИКАЦИЯ ДИНАМИЧЕСКИХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ ДИНАМИКИ $D_{\rm s}$ -ИНДЕКСА

Запропоновано новий метод ідентифікації моделей динаміки та прогнозування D_{sr} -індексу за експериментальними значеннями характеристик сонячного вітру. Він базується на реконструкції нелінійної дискретної динамічної системи, що описує поведінку геомагнітного індексу із високим ступенем кореляції з експериментальними даними. Для структурно-параметричної ідентифікації моделі прогнозування D_{sr} -індексу вперше запропоновано використати еволюційний алгоритм—метод генетичного програмування. Отримано прогнозні оцінки зміни D_{sr} -індексу на 1-9 годин наперед.

ВВЕДЕНИЕ

Состояние космической погоды в большой мере зависит от изменения солнечной активности [17]. Выбросы корональных масс порождают возмущения солнечного ветра и приводят к изменению нелинейных процессов в магнитосфере Земли [17]. Известно, что геомагнитные бури влияют на работу спутников, телекоммуникационных систем, динамику роста растений, самочувствие и здоровье людей [11, 17]. Долговременное прогнозирование состояния космической погоды особо важно при планировании космических миссий [11], так как робота электрических систем спутника может зависеть от изменения уровня солнечной активности.

Активное развитие космических исследований в направлении изучения солнечно-земных связей привело к появлению ряда новых методов описания динамики состояния магнитосферы в зависимости от влияния солнечного ветра. Ряд методов основывается на использовании физических свойств магнитосферы и гипотез [2, 6, 10] для построения эмпирической модели. Другие — на построение регрессионных моделей, нейронных сетей, RBF-сетей и вейвлет-анализа [4, 5, 7, 8, 14, 19—21].

Учитывая сложность внутренней природы явлений магнитосферы, отображающих цепоч-

ку взаимосвязанных физических процессов, не всегда возможно построить адекватную математическую модель, которая в полной мере описывает основные динамические свойства. Поэтому целесообразно использовать нелинейные динамические модели, структура которых изначально неизвестна, а их параметры зависят от экспериментальных данных наблюдений. При таком подходе характеристики солнечного ветра можно рассматривать как входы, а геомагнитные индексы — как выходы системы.

Индекс D_{st} является ключевым параметром, который характеризирует турбулентное состояние магнитосферы Земли во время магнитной бури [17]. Используя экспериментальные данные, можно построить математическую модель минимальной сложности типа «вход-выход» для прогноза значения D_{st} -индекса. Моделирование поведение D_{st} -индекса позволяет спрогнозировать будущее состояние магнитосферы. В работе предполагается, что нелинейная динамическая модель имеет один вход и один выход. При этом произведение скорости солнечного ветра и южной магнитной составляющей $B_z \cdot v$ выступают в качестве входа модели, а значения D_{st} -индекса — в качестве выхода.

ДИНАМИЧЕСКИЙ ПОДХОД К ПРОГНОЗИРОВАНИЮ $D_{\rm sr}$ -ИНДЕКСА

Для прогнозирования поведения D_{st} -индекса использовались статистические данные за 44

года, полученные из электронных web-ресурсов OMNI 2 [http://nssdc.gsfc.nasa.gov/omniweb] и Kyoto WDC (Киотского университета) [http:// swdcdb.kugi.kyoto-u.ac.jp]. В ходе построения модели прогнозирования использовалось предположение о том, что слаботурбулентное состояние магнитосферной плазмы в основном зависит от значения произведение $B_z \cdot v$ (южной составной магнитного поля Земли и скорости солнечного ветра) [1, 9, 13, 18]. В качестве переменной, характеризующей протекание процессов в магнитосфере и ионосфере, использовался D_{st} -индекс. Отбор наиболее информационных признаков, которые отвечают за существенные изменения в магнитосфере, проводился на этапе реконструкции модели прогнозирования.

ИДЕНТИФИКАЦИЯ ДИСКРЕТНЫХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ

Для построения прогноза D_{st} -индекса использовалась модель типа «вход-выход» [1, 9, 13, 18]:

$$y(k) = F'[y(k-1),...,y(k-n_y),...,u(k-1),...,u(k-n_u),e(k-1),...,e(k-n_e)+e(k)],$$
 (1)

где $F'[\bullet]$ — полином порядка l от переменных u(k), y(k) и e(k); u(k) — входной сигнал; y(k) — выходной сигнал (D_{st} -индекс); e(k) — переменная, моделирующая шум в момент времени k, $t_k = kT$, T = const — период квантования, n_u , n_y , n_e — соответствующие значения входа и выхода. Нелинейная функция F' может быть полиномом, рациональной функцией, рядом из радиальных базовых функций или любой другой функцией [3, 4, 15, 19, 20].

Предполагая, что $e(k), e(k-1), ..., e(k-n_e) =$ = const ≈ 0 , представим уравнение (1) в виде

$$\hat{y}(k) = F'[y(k-1),...,y(k-n_y),...,u(k-1),...,u(k-n_u)].$$
 (2)

Пусть в модели есть как линейные, так и нелинейные члены. Обозначим набор линейных членов по выходу до k-го момента времени через вектор \mathbf{y}_{n_y} , $n_y = 1, \dots, N_y$, по входу — \mathbf{u}_{n_u} , $n_u = 0, 1, \dots, N_u$. При построение модели будем учитывать критерий физической реализуемости

$$P(z) = \frac{\sum_{j=0}^{j=J} c_j z_y^j}{\sum_{i=0}^{j=I} b_i z_u^i} \text{ при } J < I ,$$
 (3)

где P(z) — передаточная функция, c_j и b_i — неизвестные параметры модели. Она предполагает, что «глубина» отбора предвестников (регрессоров) по выходу должна быть меньше «глубины» отбора предвестников по входу ($N_y < N_u$). Будем учитывать только нелинейные члены до второго порядка включительно. Тогда их можно представить через компоненты векторов \mathbf{y}_{n_u} и \mathbf{u}_{n_u} как

$$\prod_{\substack{s=N_{y}\\s=N_{u}\\s=1}}^{r=N_{y}}y_{r,n_{y}}u_{s,n_{u}},\prod_{\substack{r=1\\r^{*}=1\\r^{*}=1}}^{r=N_{y}}y_{r,n_{y}}y_{r^{*},n_{y}},\prod_{\substack{s=N_{u}\\s^{*}=1\\s^{*}=1}}^{s=N_{u}}u_{s,n_{u}}u_{s^{*},n_{u}},$$

а модель представима в виде

$$\hat{y}(k) = \sum_{m_{1}=1}^{M_{1}} \theta_{m_{1}}^{1} y_{m_{1},n_{y}} + \sum_{m_{2}=1}^{M_{2}} \theta_{m_{2}}^{2} u_{m_{2},n_{u}} +$$

$$+ \sum_{m_{3}=1}^{M_{3}} \theta_{m_{3}}^{3} \left(\prod_{\substack{r=1\\s=1}}^{r=N_{y}} y_{r,n_{y}} u_{s,n_{u}} \right)_{m_{3}} + \sum_{m_{4}=1}^{M_{4}} \theta_{m_{4}}^{4} \left(\prod_{\substack{r=1\\r^{*}=1}}^{r=N_{y}} y_{r,n_{y}} y_{r^{*},n_{y}} \right)_{m_{4}} +$$

$$+ \sum_{m_{5}=1}^{M_{5}} \theta_{m_{5}}^{5} \left(\prod_{\substack{s=N_{u}\\s^{*}=N_{u}}}^{s=N_{u}} u_{s,n_{u}} u_{s^{*},n_{u}} \right)_{m_{5}} + \theta^{0},$$

$$(4)$$

где θ^0 , θ^1 , ..., θ^5 — наборы неизвестных параметров модели (для простоты в дальнейшем обозначим их через вектор θ), величины M_1 ,..., M_5 определяют количество представителей разных наборов возможных комбинации предвестников до k момента времени. Уравнение (4) представим в виде

$$\hat{\mathbf{y}}(k) = \sum_{i=1}^{M} \theta_i F_i^l(\mathbf{x}(k)), \qquad (5)$$

где $F_1^l,...,F_M^l$ — нелинейная функция (без параметров), $\theta_1,...,\theta_M$ — параметры модели, $\hat{y}(k)$ — значение выхода модели в k-й момент времени, $\mathbf{x}(k)$ — вектор всех возможных регрессоров до k-го момента времени, содержащий значения y,u.

Вектор $\mathbf{x}(k)$ представим в виде

$$\mathbf{x}(k) = (u(k - n_d - 1), ..., u(k - n_d - n_u),$$

$$y(k - n_d - 1), ..., y(k - n_d - n_v)),$$
(6)

где n_d — глубина отбора регрессоров. Уравнение (5) запишем в компактной матричной форме

$$\mathbf{Y} = \mathbf{X}\mathbf{\theta},\tag{7}$$

где $\mathbf{Y} = [y(k-n_d-n_y-1), y(k-n_d-n_y-2), ..., y(k)]^{\mathrm{T}}$ — матрица-столбец экспериментальных значений D_{sl} -индекса, $\mathbf{X}(k) = [x_1(k), x_2(k), ..., x_M(k)]$ — матрица всех возможных регрессоров на момент времени k, $\mathbf{\theta} = [\mathbf{\theta}_1, \mathbf{\theta}_2, ..., \mathbf{\theta}_M]$ — вектор неизвестных параметров модели.

ГЕНЕТИЧЕСКИЙ АЛГОРИТМ ИДЕНТИФИКАЦИИ МОДЕЛИ ПРОГНОЗИРОВАНИЯ

Из матричной формы модели (7) следует, что сложность ее идентификации возрастает с увеличением порядка нелинейности и количества вход-выходных данных и их комбинаций. Поэтому целесообразно использовать процедуру отбора наиболее информативных предвестников (регрессоров). Для этого воспользуемся процедурой минимизации ошибки между экспериментальными данными и выходом модели с одновременным поиском оптимальной структуры и параметров

$$\chi^{2} = \sum_{k=1}^{N_{b}} \left(y(k) - \sum_{i=1}^{M} \theta_{i} F_{i}^{l}(\mathbf{x}(k)) \right)^{2},$$
 (8)

где N_{lr} — размер обучающей выборки экспериментальных данных, χ^2 — ошибка между результатами моделирования и экспериментальными данными, а M — количество регрессоров. Представим набор параметров модели через вектор θ

$$\mathbf{\theta} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}\mathbf{Y} \ . \tag{9}$$

Предположим, что матрицу регрессоров \mathbf{X} можно представить в виде $\mathbf{X} = \mathbf{A}\mathbf{B}$, где \mathbf{B} — матрица $M \times M$, $B_{ij} = 0$ при i > j; \mathbf{A} — матрица $N_{lr} \times M$ с ортогональными столбцами; $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ — диагональная матрица $\mathbf{\Lambda}$. Введем вспомогательный вектор параметров $\mathbf{\omega}$:

$$\mathbf{\omega} = \mathbf{\Lambda}^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{Y} \,. \tag{10}$$

Дисперсия $\mathbf{Y}^{\mathsf{T}}\mathbf{Y}/N_{lr}$ описывается выражением [16]

$$\mathbf{Y}^{\mathsf{T}}\mathbf{Y} = \sum_{i=1}^{M} \omega_i^2 a_i^T a_i + \xi \,, \tag{11}$$

где ω_i — элемент искомого вектора ω , a_i — компоненты матрицы \mathbf{A} , ξ — ошибка. Используя функцию F_i^l , введем еще одну переменную — коэффициент компенсации ошибки ξ :

$$\xi_i^* = \frac{\omega_i^2 a_i^{\mathsf{T}} a_i}{\mathbf{Y}^{\mathsf{T}} \mathbf{Y}} \,. \tag{12}$$

Этот коэффициент позволяет упорядочить регрессоры в зависимости от степени их значимости в модели, что дает возможность провести отбор наиболее информативных из них.

Идентификацию структуры модели проведена на основе метода генетического программирования (ГП) [1, 12, 16, 18]. В отличие от стандартных оптимизационных методов, в которых результаты расчетов представлены в виде набора чисел, генетический алгоритм позволяет представить результат в виде структурно-упорядоченных символов. Для получения приемлемой модели генетическим алгоритмом следует осуществить некоторую последовательность процедур. В первую очередь из экспериментальных данных проводится отбор обучающей и тестовой выборок и вводится глубина отбора входных и выходных значений (значения $v \cdot B_z$ и D_{ct} -индекса), формируя таким способом массив возможных регрессоров к моменту времени k. После чего случайным способом из этого массива отбирается некоторое количество начальных элементов (регрессоров) — начальная популяция особи. Процесс поиска регрессоров, наиболее влияющих на прогноз, похож на процесс выживания некого организма (особи) в природе. Для этого проводится оценка уровня приспосабливания (выживаемости) особи (величина, которая характеризует возможность использования того или иного регрессора на следующем этапе). Для проведения отбора в генетических алгоритмах используются ряд операторов: скрещивания (crossover), мутации (mutation) и воспроизведения (reproduction).

Для отображения результатов работы алгоритма ГП (регрессоров и их комбинации) использовалась бинарное дерево, где конечное

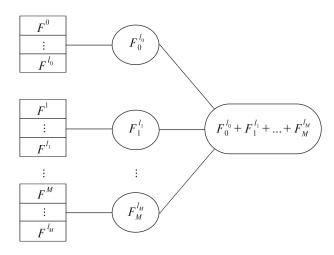


Рис. 1. Схема поиска оптимальной структуры модели

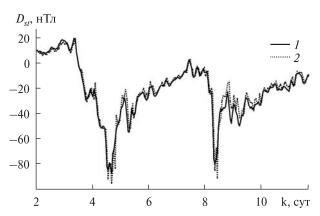


Рис. 2. Прогнозирование значения D_{sr} -индекса на один час вперед (апрель 2006 г.): I— экспериментальные данные, 2— прогноз

Значения коэффициента корреляции на интервале времени 1000 ч

Год	На 1 ч вперед	На 3 ч вперед	На 6 ч вперед	На 9 ч вперед
1999	0.954	0.751	0.593	0.488
2000	0.963	0.815	0.664	0.562
2001	0.965	0.839	0.698	0.601
2002	0.961	0.786	0.592	0.492
2003	0.955	0.804	0.661	0.557
2004	0.953	0.768	0.568	0.411
2005	0.963	0.855	0.742	0.665
2006	0.975	0.804	0.666	0.576

множество регрессоров T вмещает значения $\{x_1(k),...,x_n(k)\}$. Функциональный набор задавался арифметическими операторами $F^l\{+,*\}$ [12, 16].

При идентификации дерева модели из регрессоров и их комбинации создается ряд ячеек, которые формируют ветви основного дерева модели. Назовем места стыка ячеек узлами, которые получают значения функционального набора F^l {+,*}. Для понижения размерности структуры модели используем в алгоритме идентификации коэффициент компенсации ошибки (12), а оператор назовем оператором «обрезки». Он также будет проводить процедуру отбрасывания идентичных ячеек, т. е. если $T_i = T_{i+n}$, то T_{i+n} отбрасывается.

Перед началом работы алгоритма установим ряд ограничений: диапазон отбора данных (интервал времени, на котором проводится реконструкция модели), количество циклов отбора регрессоров, размер исходной популяции, количество возможных комбинации регрессоров, количество линейных параметров и глубина отбора регрессоров. Пусть i — количество циклов отбора регрессоров. Предположим i = 0, проведем отбор ячеек и обозначим результат F_i^I . Продолжая итерации i = i + 1, конечным результатом получаем искомую структуру $F_0^{I_0} + F_1^{I_1} + F_2^{I_2} + ... + F_M^{I_M}$, которую можно представить в виде (рис. 1):

$$\hat{y}(k) = \theta_0 F_o^{l_0} + \theta_1 F_1^{l_1} + \theta_2 F_2^{l_2} + \dots + \theta_M F_M^{l_M}. \quad (13)$$

Для проверки качества модели использовалось выражение [16]

$$\Psi = \frac{R}{1 + \exp(\alpha(Q - \beta))}, \qquad (14)$$

где Ψ — величина, характеризирующая адекватность полученной модели, R — коэффициент корреляции между набором экспериментальных данных и выходом модели, Q — сложность модели (количество ветвей дерева), α и β — параметры ограничения. Дополнительным статистическим параметром оценивания пригодности модели является средняя квадратичная ошибка

$$E_{lr} = \frac{1}{N_{lr}} \sum_{k=1}^{N_{lr}} \left(y(k) - \sum_{i=1}^{M} \theta_i F_i^l(\mathbf{x}(k)) \right)^2.$$
 (15)

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ

Для проведения численного моделирования использовался временной ряд длиной 500 часов. В результате реконструкции найдена модель вида

$$\hat{y}(k+1) = -0.434 \cdot u(k-3) \cdot u(k-4) + +0.05 \cdot y(k) \cdot u(k-3) + 0.672 \cdot u(k-4) \cdot u(k-2) + +0.713 \cdot u(k-1) - 0.472 \cdot y(k-1) + 0.163 \cdot y(k-3) + +0.488 \cdot u(k-2) + 0.211 \cdot u(k-4) - 0.485 \cdot u(k-3) - -1.482 \cdot u(k) - 0.035 \cdot y(k-1) + 1.27 \cdot y(k) - -0.02 \cdot y(k-4) - 0.077,$$
 (16)

с помощью которой было проведено прогнозирования значения D_{st} -индекса на один час вперед (рис. 2).

Для поиска модели прогнозирования на несколько часов вперед, была проведена модификация уравнения (1)

$$y(k) = F'[y(k-1),...,y(k-n_y),...,u(k-\delta),...,u(k-n_u),e(k-1),...,e(k-n_e)+e(k)],$$
 (17)

где $\delta = 10$.

В результате использования уравнения (17), (16) и алгоритма реконструкции получена модель прогнозирования

$$\hat{y}(k+2) = 0.11 \cdot u(k-25) + 1.34 \cdot \hat{y}(k+1) - 0.60 \cdot y(k) - 0.006 \cdot u(k-22) \cdot y(k-14) \cdot u(k-19) \cdot u(k-28) + 0.062 \cdot u(k-15) - 0.006 \cdot y(k-23) \cdot u(k-9) \cdot u(k-25) \times u(k-13) + 0.144 \cdot u(k-28) - 0.0016 \cdot y(k-23) \times y(k-12) - 0.085 \cdot u(k-30) + 0.003 \cdot \hat{y}(k+1) \times y(k-17) - 0.003 \cdot y(k-2) \cdot y(k-14) - 0.003 \cdot y(k-11) + 0.203 \cdot y(k-1) - 0.11 \cdot u(k-21) \cdot u(k-20) + 0.046.$$
(18)

В таблице приведены результаты расчетов коэффициентов корреляции между экспериментальными значениями D_{st} -индекса и прогнозированными значениями.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Предложен новый метод структурно-параметрической идентификации модели прогнозирования D_{st} -индекса на основе эволюционного алгоритма. Данный метод показал ряд преимуществ по сравнению с раннее предложенными

методами прогнозирования. Использование динамического подхода и генетического программирования позволяет проводить автоматический отбор наиболее значимых предвестников (регрессоров) по экспериментальным данным и реконструировать математическую модель в символьном виде. Такой способ отображения результата моделирования позволяет из возможных альтернатив моделей прогнозирования D_{st} индекса отобрать наиболее оптимальную по нескольким важным критериям (сложность структуры, точность прогноза, адекватность модели и т. д.). Разработан алгоритм поиска нелинейной дискретной динамической модели для прогнозирования поведения D_{st} индекса.

Авторы выражают благодарность О. К. Черемных и О. С. Парновскому за полезные замечания и обсуждения результатов роботы.

- 1. Семенив О. В., Сидоренко В. И., Черемных О. К. и др. Оптимизационный подход к прогнозированию космической погоды // Проблеми управленния и информатики. 2008. № 4. С. 115—130.
- 2. Baker D. N., Klimas A. J., McPherron R. L., Buchner J. The evolution from weak to strong geomagnetic activity: an interpretation in terms of deterministic chaos // J. Geophys. 1990. 17 (1). P. 41—44.
- Balikhin M., Bates I., Walker S. N. Identification of linear and nonlinear processes in space plasma turbulence // Adv. Space. – 2001. – 28. – P. 787–800.
- Billings S. A., Wei H. L. The wavelet-NARMAX representation: a hybrid model structure combining the polynomial models and multiresolution wavelet decompositions // Int. J. Systems Science. 2005. 36 (3). P. 137–152.
- Boaghe O. M., Balikhin M. A., Billings S. A., Alleyne H. Identification of nonlinear processes in the magnetosphere dynamics and forecasting of D_{st} index // J. Geophys. 2001. 106. P. 30047–30066.
- Burton R. K., McPherron R. L., Russell C. T. An empirical relationship between interplanetary conditions and D_{st} // J. Geophys. – 1975. – 80. – P. 4204–4214.
- Chen S., Billings S. A., Cowan C. F. N., Grant P. M. Practical identification of NARMAX models using radial basis functions // Int. J. Control. 1990. 52. P. 1327–1350.
- Chen S., Billings S. A. Neural networks for nonlinear dynamic system modelling and identification // Int. J. Control. – 1992. – 56. – P. 319–346.
- 9. *Cheremnykh O., Yatsenko V., Semeniv O., Shatokhina Iu.* Nonlinear dynamical model for space weather prediction // Укр. фіз. журн. 2008. **53**, № 5. С. 502—505.

- Goertz C. K., Shan L. H., Smith R. A. Prediction of geomagnetic activity // J. Geophys. 1993. 98. P. 7673–7684.
- Goodman J. M. Space Weather & Telecommunication. New York: Kluwer, 2005. – 382 p.
- 12. *Koza J. R.* Genetic Programming: On the programming of Computers by Means of Natural Evolution. MIT Press: Cambridge. 1992. 819 p.
- 13. Kuntzevich V. M., Cheremnykh O. K., Semeniv O. V., Yatsenko V. A. Space weather and its effects: modeling, prediction and risk analysis // Abstracts. 9th Ukrainian conference on space research, Yevpatoria, Crimea, Ukraine. August 31 September 5, 2009. P. 26.
- Leontaritis I. J., Billings S. A. Input-output parametric models for nonlinear systems // Int. J. Control. – 1985. – 41. – P. 303–344.
- Leontaritis I. J., Billings S. A. Model selection and validation methods for nonlinear systems // Int. J. Control. 1987. 45. P. 311–341.
- Madar J. Genetic programming for the identification of nonlinear input-output models // Industrial and Engineering Chemistry Res. – 2005. – 44 (9). – P. 3178–3186.
- Schindler K. Physics of Space Plasma Activity. New York: Cambridge Univ. Press, 2007. – 508 p.
- Semeniv O. V. Space weather prediction using evolution algorithms // Сборник тезисов «Восьмая украинская конференция по космическим исследованиям», Евпатория, Украина, 1–7 сентября 2008 г. — С. 60.

- Sharifi J., Araabi B. N., Lucas C. Multi-step prediction of D_{st} index using singular spectrum analysis and locally linear neurofuzzy modelling // Earth Planets Space. 2006. 58 (3). P. 331–341.
- 20. Wei H. L., Billings S. A. A unified wavelet-based modelling framework for nonlinear system identification: the WANARX model structure // Int. J. Control. 2004. 77 (4). P. 351–366.
- 21. Wei H. L., Zhu D. Q., Billings S. A., Balikhin M. A. Forecasting the geomagnetic activity of the D_{st} index using multiscale radial basis function networks // Adv. Space Res. -2007. -40. -P. 1863-1870.

Надійшла до редакції 01.10.09

O. V. Semeniv, V. O. Yatsenko

IDENTIFICATION OF DYNAMICAL MODELS FOR DST-INDEX FORECASTING

The new method of dynamic model identification for D_{sr} -index prediction by using experimental data is proposed. The method is based on the reconstruction of the nonlinear discrete dynamic system that gives the prediction of the geomagnetic index value with a high level of correlation to the real data. The genetic programming was used for the simulation of structure and parameters identification for D_{sr} -index prediction. Predictive values of the D_{sr} -index dynamics for 1-9 hours ahead are derived.