

УДК 550.38

А. С. Парновский¹, А. Ю. Полонская^{1,2}, В. Н. Шевченко¹, И. Т. Жук¹, Н. В. Маслова¹

¹ Институт космических исследований Национальной академии наук Украины
та Национального космического агентства Украины, Київ

² Национальный авиационный университет, Київ

РЕГРЕССИОННОЕ МОДЕЛИРОВАНИЕ ИНДЕКСОВ a_p И K_p : ПЕРВЫЕ РЕЗУЛЬТАТЫ

На основе методу регресійного моделювання була розроблена модель, що дозволяє прогнозувати індекси a_p та K_p на 3 години вперед. Для a_p -індексу коефіцієнт кореляції складає 0.86 (0.75 для тривіальної моделі), ефективність прогнозування 0.87, стандартне відхилення 9.15 нТл. Для K_p -індексу 96 % точок мають відхилення не більше за 1 та 83 % точок мають відхилення не більше за 1/3. Подібно до D_{st} -індексу, a_p -індекс має пам'ять з глибиною порядку 1000 годин. Він також містить помітні циклічності з періодами 12 год (добові), 27 діб (карінгтонівські) та 6 місяців (сезонні).

ВВЕДЕНИЕ

В более ранних работах [1, 2, 6–8] нами был описан метод регрессионного моделирования и применен к задаче прогнозирования D_{st} -индекса. Ниже этот метод будет применен к индексам a_p и K_p .

Результатом применения данного метода являются феноменологические модели, описывающие зависимость выхода нелинейной дискретной динамической системы от ее входов и предыдущих значений выхода. Эти модели могут использоваться, в частности для прогнозирования выхода системы на некоторое количество шагов вперед (с некоторой заблаговременностью). При этом критерии качества прогноза также являются критериями качества модели.

ОПИСАНИЕ ДАННЫХ И МЕТОДИКИ

В рамках данного метода прогнозное значение выхода $y^*(T + \Theta)$, где Θ — заблаговременность, ищется в виде

$$y^*(T + \Theta) = C_0 + \sum_{i=1}^m C_i x_i(T - l), l = \overline{0, L}, \quad (1)$$

где $x_i (i = \overline{1, \dots, m})$ — регрессоры, являющиеся членами полинома Колмогорова—Габорта от значений входов $u_k(t)$, известных на момент прогноза T :

$$x_i(t) = \prod_{k=1}^K u_k^{p_k}(t). \quad (2)$$

Здесь K — количество измеряемых входов, p_k — показатели степени ($\sum_k p_k \leq 4$), $C_i (i = 0, \dots, m)$ — коэффициенты регрессии, C_0 — коэффициент при постоянном регрессоре $x_0 \equiv 1$, m — количество переменных регрессоров, l — запаздывание, L — максимальное запаздывание.

Естественно, истинное значение выхода $y(T + \Theta)$ зависит также от неизвестных на момент прогнозирования значений входов $u_k(T + 1), \dots, u_k(T + \Theta)$. Это означает, что повышая заблаговременность Θ мы жертвуем возможностью описания процессов с меньшими временными масштабами. Учитывая этот факт, а также то, что нам доступны усредненные по времени значения входов, мы можем записать

$$y(T + \Theta) = y^*(T + \Theta) + \Delta y(T + \Theta), \quad (3)$$

где $\Delta y(T + \Theta)$ — ошибка прогноза, которую мы будем считать случайной.

Коэффициенты регрессии мы определяли по методу наименьших квадратов, используя алгоритм Гаусса—Жордана [9], так как нам требовалось знание ковариационной матрицы. Статистическая значимость регрессоров определялась по тесту Фишера [3, 4, 10]. Регрессор считался существенным, если его нельзя было отбросить на уровне значимости 10 %.

© А. С. ПАРНОВСКИЙ, А. Ю. ПОЛОНСКАЯ, В. Н. ШЕВЧЕНКО,
И. Т. ЖУК, Н. В. МАСЛОВА, 2011

Регрессоры добавлялись в следующем порядке: предыдущие значения выхода (авторегрессия), затем входные параметры, и наконец — комбинации входных параметров согласно (2). На каждом этапе несущественные регрессоры отбрасывались.

В данной работе в качестве выхода y используется геомагнитный индекс a_p . Этот индекс изменяется в диапазоне от 0 до 400 нТл и обновляется каждые 3 ч. Он измеряется по данным 13 среднеширотных станций, из которых девять расположены в северном полушарии, а две — в южном. С ним связан индекс K_p , который является основным в этой паре, однако неудобен для прогнозирования из-за небольшого диапазона принимаемых значений. По этой причине, говоря о прогнозе K_p -индекса, мы будем подразумевать прогноз a_p -индекса, пересчитанный по таблице соответствия этих двух индексов. В качестве данных для построения модели использовалась выборка из каталога OMNI2 [http://nssdc.gsfc.nasa.gov/omniweb/] с 01.01.1976 г. по 31.12.2000 г. Для проверки результатов прогнозирования использовалась выборка с 01.01.2001 г. по 31.12.2003 г.

РЕЗУЛЬТАТЫ

Глубина памяти a_p -индекса, определенная тем же методом, что и для D_{st} -индекса в работах [2, 7, 8], оказалась порядка 1000 ч (1.5 каррингтоновских оборота). Это значение хорошо согласуется с результатом работы [5].

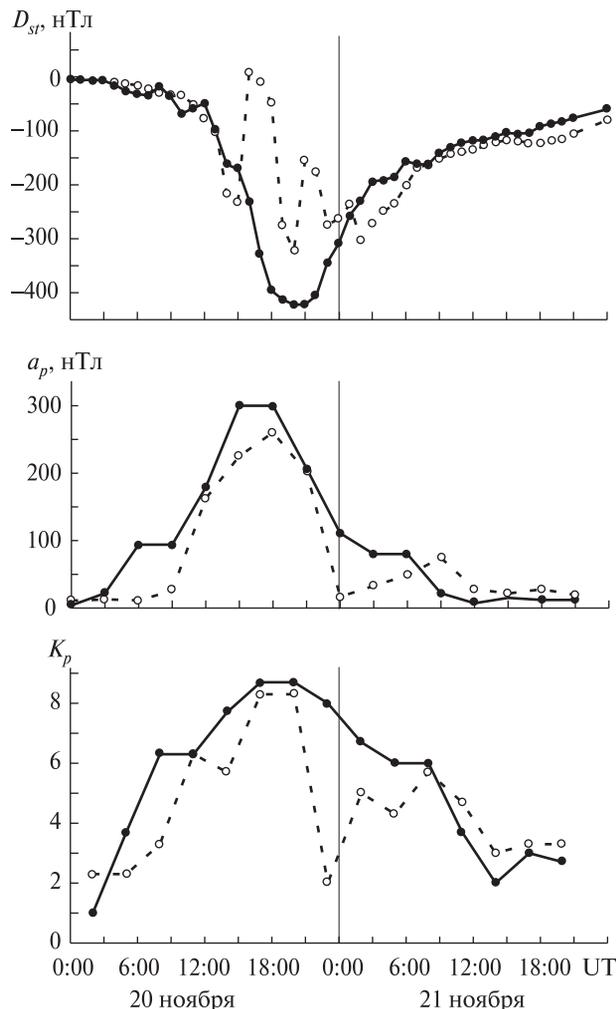
Основные характеристики разработанных моделей для заблаговременности 3 ч приведены

ны в таблице. В ней используются следующие обозначения: тривиальная модель — модель вида $y^*(T + \Theta) = y(T)$; авторегрессионная модель вида $y^*(T + \Theta) = C_0 + \sum_{l=0}^L C_l y(T-l)$; линейная модель — модель вида $y^*(T + \Theta) = C_0 + \sum_{l=0}^L C_l y(T-l) + \sum_{k=L+1}^{m+2} C_k u_k(T-l)$; нелинейная модель — частная регрессионная модель вида (1); СКО — среднее квадратичное отклонение; PE — эффективность прогноза $PE = 1 - \left(\frac{\sigma}{\sigma_s}\right)^2$, где σ_s — стандартное отклонение выборки; r — коэффициент линейной корреляции. Для K_p -индекса мы указываем долю точек с отклонениями не более 1/3 и 1. Видно, что лишь для нелинейной модели для D_{st} -индекса при переходе к тестовой выборке произошло существенное понижение статистических характеристик, что указывает на несколько избыточную сложность этой модели.

На рисунке приведено сравнение прогнозного и измеренного значений индексов D_{st} , a_p и K_p для бури 20—22 ноября 2003 г. Модель для D_{st} -индекса правильно предсказала начало и продолжительность бури. Колебания во время главной фазы связаны с малой статистикой сверхсильных бурь в учебной выборке. Модель для a_p - и K_p -индексов показала менее уверенный результат и требует доработки. Описываемый интервал не входит в учебную выборку.

Статистические характеристики разработанных моделей при прогнозировании на 3 ч вперед

Модель	Индекс									
	D_{st}				a_p				K_p	
	СКО, нТл	PE, %	r , %	r_0 , %	СКО, нТл	PE, %	r , %	r_0 , %	$\pm 1/3$, %	± 1 , %
<i>Учебная выборка (01.01.1976 г. — 31.12.2000 г.)</i>										
Тривиальная	10.38	82.2	91.1	91.1	13.65	53.5	76.8	76.8	46.2	81.8
Авторегрессия	9.92	83.7	91.5	91.1	12.74	59.6	77.2	76.8	80.1	93.9
Линейная	7.89	85.6	92.5	89.2	8.06	68.5	82.9	71.6	80.0	94.3
Нелинейная	7.38	87.8	93.7	89.3	7.69	71.3	84.6	71.6	81.6	95.6
<i>Экзаменационная выборка (01.01.2001 г. — 31.12.2003 г.)</i>										
Тривиальная	12.00	82.3	91.2	91.2	15.72	53.8	76.9	76.9	46.9	82.6
Авторегрессия	11.59	83.9	91.6	91.2	15.07	58.8	76.9	76.8	80.3	94.0
Линейная	9.26	87.7	93.7	90.6	10.35	67.2	82.4	76.6	81.1	94.8
Нелинейная	12.11	79.8	89.6	90.9	10.80	64.3	80.9	76.6	83.2	96.1



Значения D_{st} -, a_p - и K_p -индексов для бури 20–22 ноября 2003 г.: сплошная линия — наблюдения, штриховая — прогноз на 3 ч вперед

ВЫВОДЫ

Таким образом, метод регрессионного моделирования, разработанный для прогнозирования D_{st} -индекса, оказался вполне применим и к индексам a_p и K_p . Полученные с его помощью результаты уже представляют практическую ценность благодаря высокой точности (не менее 80–90 %) и высокому быстродействию (не более 1–2 мин).

В данной работе использовалась база данных OMNI2 (<http://nssdc.gsfc.nasa.gov/omniweb/>). Данная работа была частично поддержана НАН Украины и

НКА Украины в рамках программы GEO-UA (Распоряжения Президиума НАН Украины № 322 от 30.05.2008 г. и № 377 от 24.06.2010 г.), а также молодежным грантом (Распоряжение Президиума НАН Украины № 244 от 13.04.2009 г. и Постановление Президиума НАН Украины № 189 от 24.06.2009 г.).

1. Парновский А. С. Прогнозирование D_{st} -индекса методом линейного регрессионного анализа // Космічна наука і технологія. — 2008. — **14**, № 3. — С. 48–54.
2. Парновский А. С. Метод регрессионного моделирования и его применение к задаче прогнозирования космической погоды // Проблемы управления и информатики. — 2009. — № 3. — С. 128–135.
3. Fisher R. A. Statistical methods for research workers: 12th ed. — London, Oliver and Boyd, 1954. — 356 p.
4. Hudson D. J. Statistics. Lectures on Elementary Statistics and Probability. — Geneva, CERN, 1964. — 320 p.
5. Johnson J. R., Wing S. A cumulant-based analysis of non-linear magnetospheric dynamics. // Report PPPL-3919-rev. — 2004. — (http://www.pppl.gov/pub_report/2004/PPPL-3919rev.pdf).
6. Parnowski A. S. Statistical approach to Dst prediction // J. Phys. Studies. — 2008. — **12**, N 4. — P. 4003.
7. Parnowski A. S. Statistically predicting Dst without satellite data // Earth, Planets and Space. — 2009. — **61**, N 5. — P. 621–624.
8. Parnowski A. S. Regression modeling method of space weather prediction // Astrophys. and Space Sci. — 2009. — **323**, N 2. — P. 169–180.—doi:10.1007/s10509-009-0060-4. [arXiv:0906.3271].
9. Press W. H., Teukolsky S.A., Vetterling W.T., Flannery B.P. Numerical Recipes in FORTRAN 77 and 90. The Art of Scientific Computing. — 2nd ed. — Cambridge — New York — Melbourne: Cambridge Univ. Press, 1992.
10. Seber G. A. F. Linear Regression Analysis. — New York: Wiley, 1977.

Надійшла до редакції 17.12.10

A.S. Parnowski, A.Yu. Polonskaya, V.N. Shevchenko, I.T. Zhuk, N.V. Maslova

REGRESSION MODELLING FOR a_p AND K_p INDICES: FIRST RESULTS

On the basis of the regression modelling method, the model is developed which allows us to forecast a_p and K_p indices three hours ahead. For a_p index, the linear correlation coefficient is 0.86 (0.75 for the trivial model), the prediction efficiency is 0.87, and the standard deviation is 9.15 nT. For K_p index, 96 % of all the points have the absolute error not more than 1 unit of K_p , and 83 % of all the points have the absolute error not more than 1/3 units of K_p . Similar to the D_{st} index, the a_p index has a memory of its previous values for approximately 1000 hours. It also has evident temporal variations with typical periods of 12 hours (diurnal), 27 days (Carrington) and 6 months (seasonal).