

УДК 550.38

А. С. Парновский

Институт космических исследований Национальной академии наук Украины и Национального космического агентства Украины, Киев

## Прогнозирование $D_{st}$ -индекса методом линейного регрессионного анализа

Надійшла до редакції 09.08.07

Метою роботи є побудова феноменологічної регресійної залежності  $D_{st}$ -індексу від параметрів сонячного вітру і використання цієї залежності для прогнозування. Магнітосфера розглядається як чорний ящик, тобто не використовуються жодні моделі або припущення. В роботі отримана залежність, що забезпечує прогнозування  $D_{st}$ -індексу на 9 годин вперед. При цьому коефіцієнт кореляції регресійного значення  $D_{st}$  з вимірним складас від 98.6 % при прогнозуванні на 1 годину до 79.3 % при прогнозуванні на 9 годин. У роботі також обговорюється, як вигляд статистично значущих регресорів може сприяти розумінню фізичного механізму впливу параметрів сонячного вітру на геомагнітну активність.

### ВВЕДЕНИЕ

Прогнозирование космической погоды является одной из основных задач современных космических исследований. Одной из ее важнейших подзадач является прогнозирование геомагнитной активности на поверхности Земли, так как именно она определяет влияние космической погоды на самочувствие людей, качество радиосвязи, работоспособность линий электропередачи и многие другие области современной жизни. На сегодняшний день есть несколько количественных показателей геомагнитной активности, самыми распространенными из которых являются шторм-тайм-вариация геомагнитного поля  $D_{st}$  и планетарный индекс геомагнитной активности  $K_p$ . Для прогнозирования геомагнитной активности индекс  $D_{st}$  более удобен, так как непосредственно отображает вариацию магнитного поля на поверхности Земли в нанотесла, в то время как индекс  $K_p$  является интегральным и более искусственным, поскольку является балльной характеристикой степени возмущенности магнитного поля (в этом смысле он схож с магнитудой землетрясений по шкале Рихтера). Кроме того, в используемых нами данных ряд

измерений  $D_{st}$  непрерывен с 1 января 1963 г. до 30 апреля 2006 г., в то время как ряд измерений  $K_p$  содержал разрывы. Подробнее с определением и методикой измерения  $D_{st}$ -индекса можно ознакомиться в работах [1, 3, 6]. Следует отметить, что предсказанием геомагнитной активности по параметрам солнечного ветра занимались и ранее [2, 5—7].

Целью работы является прогнозирование  $D_{st}$ -индекса без использования каких-либо физических моделей или предположений на основе спутниковых данных 1-ч разрешения о параметрах солнечного ветра, перечисленных ниже. Для этого мы воспользуемся стандартным хорошо зарекомендовавшим себя методом линейного регрессионного анализа. В рамках этого метода прогнозируемое значение  $D_{st}$  можно представить в виде

$$D_{st}(j) = \sum_i C_i x_i(j),$$

где  $j$  — текущий шаг (количество часов с 1 января 1963 г.),  $C_i$  — коэффициенты регрессии, а  $x_i(j)$  — регрессоры, являющиеся функциями входных величин или их комбинаций. Значения коэффициентов  $C_i$  определяются по методу наи-

меньших квадратов (МНК) с единичными статистическими весами всех точек, а значимость регрессоров — по методу Фишера.

Во многих работах по предсказанию геомагнитной активности используется лишь один входной параметр — электрическое поле утро-вечер ( $vB_z$ ). Забегая наперед, отметим, что соответствующая ему группа регрессоров действительно является наиболее статистически достоверной после предыдущих значений  $D_{st}$ , однако использование регрессии, содержащей только эти величины, приводит к существенно худшим результатам, чем при использовании нескольких входных параметров. Во многих работах также используются подходы, связанные с нейронными сетями. Хотя эти методы по сложности значительно превосходят метод линейного регрессионного анализа, но также обеспечивают лишь кратковременный прогноз (например, 4 ч в работе [6]). Кроме того, в ряде случаев эти методы испытывают значительные сложности при предсказании интенсивных бурь с  $K_p > 5$ , которые характеризуются резким глубоким падением  $D_{st}$ -индекса. Метод линейного регрессионного анализа при большей простоте лишен этого недостатка, и в нашем случае обеспечивает уверенное прогнозирование геомагнитной активности на глубину до 9 ч.

#### ИСПОЛЬЗУЕМЫЕ ДАННЫЕ

В качестве входных данных использовалась база данных OMNI 2, созданная в Национальном центре научных космических данных NSSDC (США), и доступная по адресу <http://nssdc.gsfc.nasa.gov/omniweb/>. Она содержит усредненные на 1-ч интервале значения межпланетного магнитного поля, параметры плазмы солнечного ветра и потоки энергетичных частиц, измеренные различными КА в точке Лагранжа между Землей и Солнцем и пересчитанные в баллистическом приближении на магнитопаузу, а также индексы геомагнитной и солнечной активности, измеренные на поверхности Земли (всего 49 параметров начиная с 1963 г.).

На рис. 1 представлена временная развертка  $D_{st}$ -индекса за 43 года наблюдений (с 1963 по 2005 г.). Отчетливо видно, что нижняя огибающая графика, соответствующая наиболее мощ-

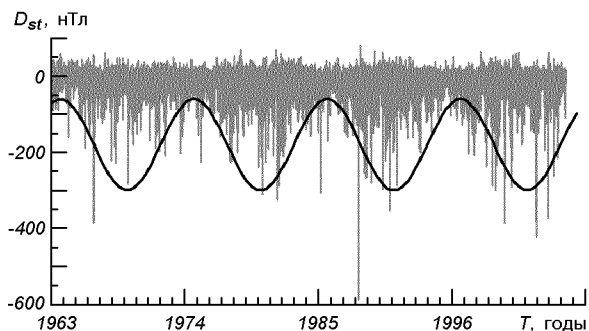


Рис. 1. Вариации индекса  $D_{st}$  за 1963—2006 гг. Жирная линия — синусоида с периодом 11 лет

ным геомагнитным бурям (напомним, что возмущения геомагнитного поля тем больше, чем меньше  $D_{st}$ ), воспроизводит 11-летний солнечный цикл. Если построить этот же график в более крупном масштабе, можно будет увидеть, что средний уровень  $D_{st}$  колеблется с периодом, близким к периоду обращения Солнца 27 сут. Все это указывает на правильность выбора  $D_{st}$ -индекса в качестве характеристики геомагнитной активности.

К сожалению, в базе данных OMNI 2 были участки пропусков данных. Поэтому нам пришлось отбросить эти участки, называемые далее выбросами. Конкретная процедура их отбрасывания приведена ниже.

#### МЕТОДИКА И ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ

В отличие от некоторых других работ по прогнозированию геомагнитной активности, мы строим не степенную регрессию от одного входного параметра, а регрессию от большого количества входных параметров и их комбинаций, в том числе нелинейных. Используя прямой метод вычисления коэффициентов регрессии без ортогонализации входных параметров, мы можем сохранить их физический смысл и, таким образом, получить оценки геоэффективности различных физических параметров и их комбинаций.

Отбор регрессоров производился с учетом корреляционных функций различных параметров и известных моделей взаимодействия солнечного ветра с магнитосферой (например, электриче-

ское поле утро-вечер). Первоначальное количество регрессоров бралось заведомо избыточным, чтобы отбор по методу Фишера выделил из них наиболее статистически существенные. Для прогнозирования на различное количество часов вперед набор регрессоров получался независимо. При этом использовались все данные за 43 года, за исключением тех, где отсутствовали данные либо по  $D_{st}$ , либо по одному из параметров, входящих в регрессию. Качество полученного прогноза оценивалось по среднему квадратичному отклонению предсказанного значения от измеренного и по коэффициенту корреляции между этими двумя величинами. При прогнозировании на 1 ч среднее квадратичное отклонение составило около 3.75 нТл (0.56 % от размаха колебаний  $D_{st}$ , равного 670 нТл), а при устремлении времени прогнозирования к бесконечности оно сравняется со стандартным отклонением выборки, равным 24 нТл (3.58 % от размаха). Предсказанное значение при этом будет соответствовать среднему значению по выборке ( $D_{st} = -17.89$  нТл). Среднее квадратичное отклонение, среднее между этими двумя предельными значениями, соответствует глубине прогнозирования 9 ч (рис. 2). Именно это значение характеризует характерную глубину прогнозирования по данному методу.

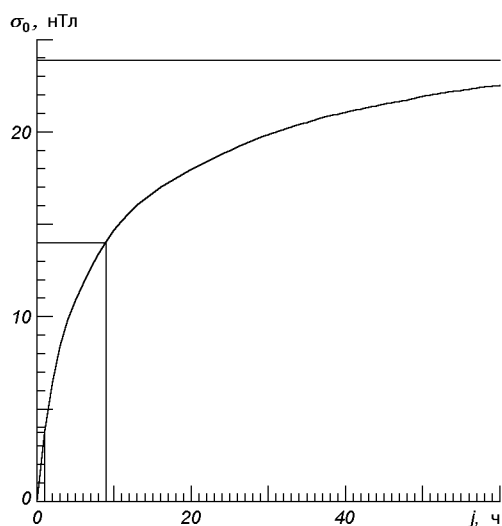


Рис. 2. Зависимость стандартного отклонения  $\sigma_0$  остатков регрессии от глубины прогнозирования

При получении набора статистически достоверных регрессоров мы начали с исходного большого числа регрессоров. После обработки данных по МНК для каждого из регрессоров определялся параметр  $F$ , характеризующий значимость по тесту Фишера. Эти параметры сравнивались со значениями 2.7055, 3.84, 5.02, 6.635, 7.879, 10.83 и 12.1, соответствующими уровням статистической значимости 10, 5, 2.5, 1, 0.5, 0.1 и 0.05 %. Максимальное значение коэффициента Фишера составило  $1.5 \cdot 10^5$  для регрессора, соответствующего значению  $D_{st}$  1 ч назад при прогнозировании на 1 ч вперед. После отбрасывания несущественных регрессоров процедура повторялась до тех пор, пока не получался нужный набор. Окончательное количество регрессоров зависело от принятого минимального уровня достоверности. Повышая уровень достоверности, мы получаем меньшее число регрессоров, каждый из которых заведомо статистически существен. Параметры регрессионной зависимости (1) для прогнозирования на 1 ч вперед приведены в таблице. Отметим, что в отличие от работы [6] мы не вводим подгоночных параметров, и все используемые регрессоры имеют очевидный физический смысл.

Обратим внимание, что эту процедуру необходимо применять к большой выборке. В качестве эксперимента мы провели эту процедуру для выборок, соответствующих отдельным годам. В них появлялись новые статистически существенные для конкретного года регрессоры, не существенные ни на большом интервале, ни для других годов. Они, по-видимому, связаны с конкретной фазой 11-летнего цикла и другими особенностями года, для которого они были построены. Также были случаи, когда существенность регрессора в большой выборке, понижалась при переходе к отдельным годам. Мы обнаружили, что для построения регрессии вполне достаточно использовать 22-летнюю выборку: коэффициенты, полученные для интервала 1963—1984 гг., использованы для прогнозирования на интервале 1985—2006 гг. По сравнению с полной выборкой коэффициент корреляции при прогнозировании на 1 ч понизился с 98.7 до 98.6 %, а при прогнозировании на 9 ч — с 79.8 до 79.3 %. Для сравнения в работе [6] при использовании данных с лучшим временным разрешением при прогнозировании на время распространения воз-

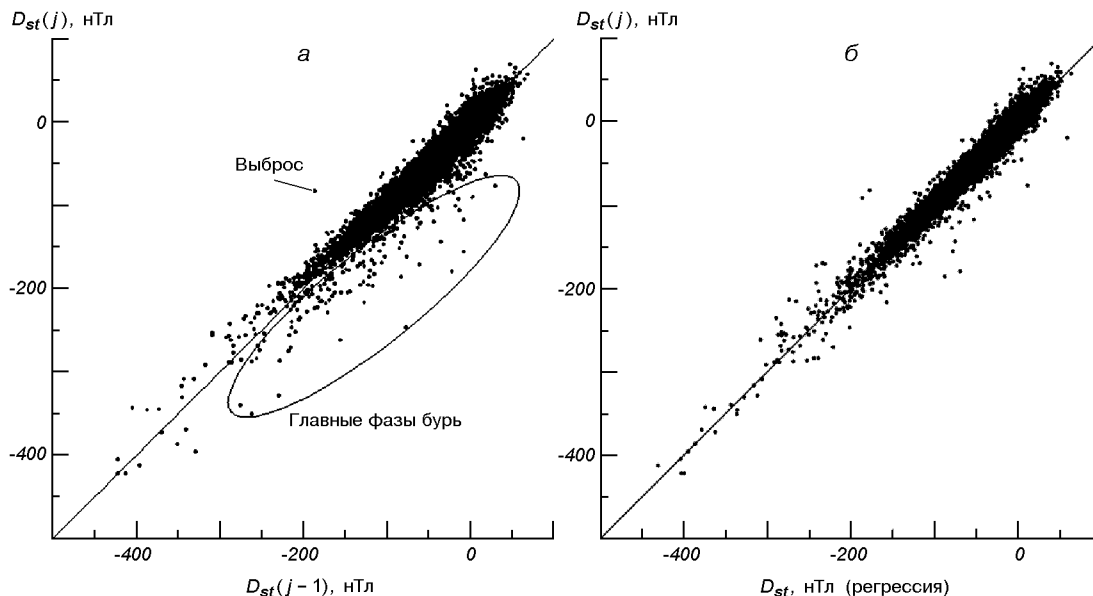


Рис. 3. Зависимость измеренных значений  $D_{St}(j)$ : *a* — от предсказанных значений  $D_{St}(j-1)$  в простейшей модели с глубиной предсказания 1 ч, *б* — от предсказанных значений  $D_{St}$  в полной модели с глубиной предсказания 1 ч

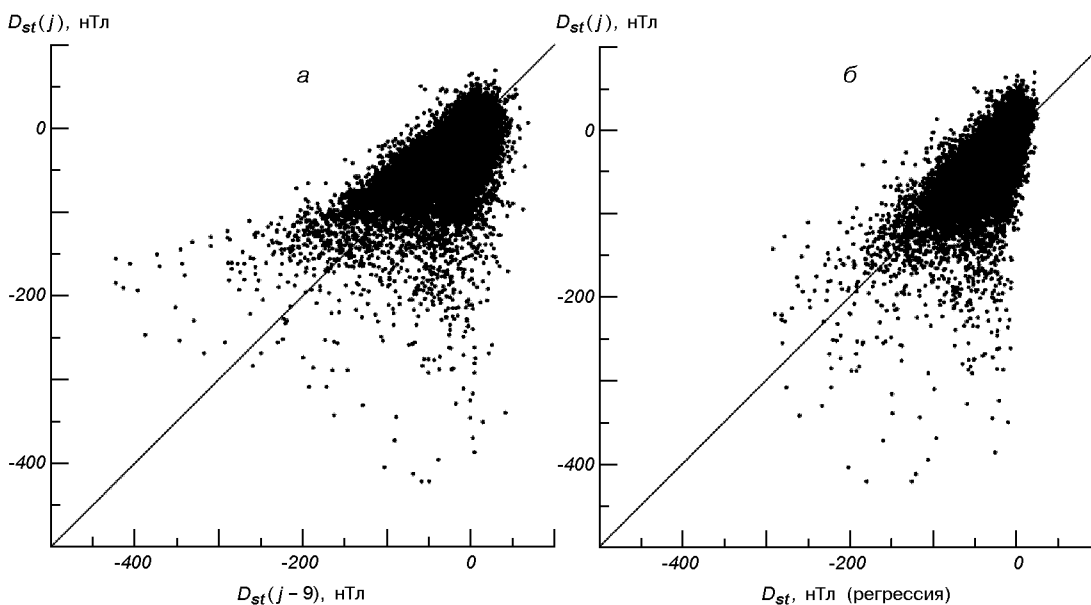


Рис. 4. То же для моделей с глубиной предсказания 9 ч

мущений от орбиты спутника до Земли (десятки минут) за 1995—1999 гг. получена зависимость, обеспечивающая линейную корреляцию 93.9 % и среднее квадратичное отклонение 6.4 нТл, а за первую половину 2000 г. — линейную корреляцию

95 % и среднее квадратичное отклонение 7.9 нТл.

Отметим, что достоверность постоянного регрессора увеличивается с увеличением глубины прогнозирования, а достоверность регрессора,

Коэффициенты регрессии при прогнозировании на 1 час вперед. Минимальный уровень статистической достоверности регрессоров — 90 % ( $F > 2.7055$ ), среднее квадратичное отклонение — 3.76 нТл, объем выборки — 139692 точки, коэффициент корреляции — 98.7 %

$i$	$C_i$	$\Delta C_i$	$F$	$x_i$
1	0.115555E+01	0.287866E-02	0.161137E+06	$D_{st}(j-1)$
2	-0.299802E+00	0.436327E-02	0.472112E+04	$D_{st}(j-2)$
3	0.690698E-01	0.437462E-02	0.249285E+03	$D_{st}(j-3)$
4	-0.737033E-02	0.318689E-02	0.534734E+01	$D_{st}(j-4)$
5	0.158539E-01	0.291081E-02	0.296649E+02	$D_{st}(j-7)$
6	0.182679E-01	0.236903E-02	0.594585E+02	$D_{st}(j-8)$
7	0.113505E-02	0.300712E-04	0.142473E+04	$B_z(j-1) \cdot v(j-1)$
8	-0.871260E-03	0.120588E-04	0.522020E+04	$B_z(j-2) \cdot v(j-2)$
9	-0.249543E-01	0.398636E-03	0.391867E+04	$B_z^2(j-1)$
10	-0.175310E+00	0.136431E-01	0.165114E+03	$ B  (j-1)$
11	0.589849E-01	0.160058E-01	0.135818E+02	$ B  (j-3)$
12	-0.142064E+00	0.208747E-01	0.463169E+02	$ B  (j-6)$
13	0.115483E+00	0.208237E-01	0.307497E+02	$ B  (j-7)$
14	-0.651049E-01	0.206890E-01	0.990531E+01	$ B  (j-10)$
15	0.449028E-01	0.192162E-01	0.545848E+01	$ B  (j-11)$
16	0.317899E-01	0.843823E-02	0.141955E+02	$ B  (j-20)$
17	-0.199444E-02	0.554909E-03	0.129226E+02	$ B ^2(j-1)$
18	0.111616E-01	0.521612E-03	0.457888E+03	$ B ^2(j-2)$
19	-0.379923E-02	0.671470E-03	0.320156E+02	$ B ^2(j-3)$
20	-0.440797E-02	0.415798E-03	0.112387E+03	$ B ^2(j-4)$
21	0.841704E-02	0.705181E-03	0.142467E+03	$ B ^2(j-6)$
22	-0.782535E-02	0.777916E-03	0.101186E+03	$ B ^2(j-7)$
23	0.287269E-02	0.520874E-03	0.304134E+02	$ B ^2(j-8)$
24	-0.147823E-02	0.524860E-03	0.793590E+01	$ B ^2(j-9)$
25	0.614276E-02	0.767092E-03	0.641318E+02	$ B ^2(j-10)$
26	-0.240567E-02	0.655135E-03	0.134852E+02	$ B ^2(j-11)$
27	0.251051E-02	0.386572E-03	0.421782E+02	$ B ^2(j-15)$
28	-0.358427E-02	0.520243E-03	0.474631E+02	$ B ^2(j-16)$
29	0.314027E-02	0.515821E-03	0.370644E+02	$ B ^2(j-17)$
30	-0.138873E-02	0.413629E-03	0.112727E+02	$ B ^2(j-18)$
31	-0.141097E-02	0.353481E-03	0.159300E+02	$ B ^2(j-20)$
32	0.425745E+00	0.143661E-01	0.878246E+03	$B_z(j-1)$
33	-0.189150E+00	0.189042E-01	0.100114E+03	$B_z(j-3)$
34	0.433738E-01	0.216148E-01	0.402906E+01	$B_z(j-4)$
35	-0.963558E-01	0.215678E-01	0.199564E+02	$B_z(j-5)$
36	0.775695E-01	0.176548E-01	0.193025E+02	$B_z(j-6)$
37	0.116744E+00	0.451274E-02	0.669242E+03	$n_i(j-1)$
38	-0.961877E-02	0.315033E-02	0.932420E+01	$v(j-2)$
39	0.156108E-01	0.358553E-02	0.189546E+02	$v(j-3)$
40	-0.776699E-02	0.193840E-02	0.160545E+02	$v(j-6)$
41	-0.172396E-04	0.683267E-06	0.636608E+03	$v(j-1)^2$
42	0.230969E-04	0.298000E-05	0.600727E+02	$v(j-2)^2$
43	-0.142711E-04	0.333875E-05	0.182716E+02	$v(j-3)^2$
44	0.893491E-05	0.182477E-05	0.239719E+02	$v(j-6)^2$

Окончание таблицы

$i$	$C_i$	$\Delta C_i$	$F$	$x_i$
45	-0.855826E+00	0.167418E-01	0.261317E+04	$p(j-1)$
46	0.268689E+00	0.183159E-01	0.215198E+03	$p(j-2)$
47	-0.783364E-01	0.182902E-01	0.183477E+02	$p(j-3)$
48	-0.721052E-01	0.175859E-01	0.168108E+02	$p(j-4)$
49	0.557193E-01	0.164529E-01	0.114670E+02	$p(j-5)$
50	0.117062E+00	0.167627E-01	0.487684E+02	$p(j-6)$
51	-0.749044E-01	0.141459E-01	0.280377E+02	$p(j-7)$
52	-0.579183E-01	0.140430E-01	0.170078E+02	$p(j-9)$
53	0.426967E-01	0.126796E-01	0.113344E+02	$p(j-10)$
54	0.457738E-01	0.128966E-01	0.126012E+02	$p(j-15)$
55	-0.572935E-01	0.169735E-01	0.113934E+02	$p(j-16)$
56	0.353073E-01	0.133981E-01	0.694885E+01	$p(j-17)$
57	0.160221E-01	0.807798E-02	0.393141E+01	$p(j-20)$
58	-0.182053E+00	0.390814E-01	0.216982E+02	$E(j-3)$
59	0.216142E+00	0.438804E-01	0.242601E+02	$E(j-4)$
60	-0.347304E+00	0.438600E-01	0.627032E+02	$E(j-5)$
61	0.138882E+00	0.361020E-01	0.148013E+02	$E(j-6)$
62	0.232382E+00	0.902961E-02	0.662319E+03	$E(j-8)$
63	0.134628E+01	0.207456E+00	0.421157E+02	1

описывающего последнее измеренное значение  $D_{st}$ , снижается. При прогнозировании на 1 ч соответствующие коэффициенты Фишера составляют 42 и  $1.6 \cdot 10^5$  соответственно, а при прогнозировании на 9 ч — 1002 и 1482. Увеличение достоверности постоянного регрессора связано с тем, что при увеличении глубины прогнозирования регрессионное значение  $D_{st}$  стремится к среднему значению по выборке. Что же касается последнего измеренного значения  $D_{st}$ , то уменьшение его достоверности связано с его «устареванием» по мере увеличения глубины прогнозирования. Тем не менее, регрессия, состоящая из одного лишь этого регрессора, является простейшей моделью описания поведения  $D_{st}$ -индекса. Она соответствует предположению, что  $D_{st}$ -индекс не изменился по сравнению с последним измеренным значением. Уровень корреляции регрессионного значения  $D_{st}$  в этой модели является минимальным среди всех возможных моделей, поэтому о качестве модели можно судить по разности корреляций рассматриваемой и простейшей моделей. Для иллюстрации приведем рис. 3 и 4, на которых для

глубины прогнозирования в 1 и 9 ч соответственно по оси ординат отложены измеренные значения  $D_{st}$ , а по оси абсцисс — регрессионные. При этом рис. 3, а и 4, а соответствуют простейшей модели, а рис. 3, б и 4, б — полной. Видно, что полные модели обеспечивают заметно меньшие отклонения от измеренных значений. Кроме того, рис. 3, а очень удобен для анализа исходных данных. Точки, лежащие вблизи линии одинаковых значений, соответствуют спокойным состояниям магнитосферы. Точки, лежащие ниже этой линии, соответствуют главным фазам бурь. Точки, лежащие слегка выше диагонали, соответствуют восстановительным фазам бурь. Единственная точка, лежащая значительно выше, является выбросом, произошедшим 21 октября 1989 г. в 17:00.

**ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ И ВЫВОДЫ**

Обсудим физический смысл наиболее существенных регрессоров. Максимальную достоверность имеет группа регрессоров  $D_{st}(j - i)$ , где  $j$  —

момент прогнозирования,  $i$  — количество шагов предыстории (см. таблицу). Они играют двойную роль: с одной стороны, они задают предысторию поведения  $D_{st}$ , а с другой — неявно содержат в себе невязки модели в предыдущие моменты времени, вводя тем самым в модель некоторую коррекцию. Следующая по достоверности группа регрессоров —  $v \times B_z(j - i)$ . Это электрическое поле утро-вечер [1], наводимое солнечным ветром в магнитосфере. До настоящего времени этот параметр считался наиболее геоэффективным. Далее идут регрессоры  $B_z^2(j - 1)$  и  $n_i(j - 1)$ , а также группы регрессоров  $p(j - i)$ ,  $|B|^2(j - i)$  и  $v^2(j - i)$ . Эти регрессоры описывают полное давление солнечного ветра. Далее следуют южный компонент магнитного поля солнечного ветра  $B_z(j - i)$ , считавшийся наиболее геоэффективным параметром в теории Данжи [1, 4], электрическое поле солнечного ветра  $E(j - i)$ , скорость солнечного ветра  $v(j - i)$  и модуль напряженности магнитного поля солнечного ветра  $|B|(j - i)$ . Кроме перечисленных, достоверен и постоянный регрессор.

Таким образом, в данной работе была получена регрессионная зависимость индекса  $D_{st}$  от параметров солнечного ветра вида (1), обеспечивающая его прогнозирование на несколько часов вперед. Характерная глубина прогнозирования — 9 ч. Коэффициент корреляции между регрессионным и измеренным значениями  $D_{st}$  при прогнозировании на 1 ч составил 98.6 %, а на 9 ч — 79.3 %. Регрессоры, входящие в указанную зависимость, были выбраны таким образом, что их статистическая достоверность, определенная по методу Фишера, составляла не менее 90 %. Физический смысл указанных регрессоров достаточно понятен: они описывают предысторию  $D_{st}$ , электрическое поле утро-вечер, полное давление солнечного ветра, а также скорость, электрическое поле, модуль и южный компонент магнитного поля. Отметим, что достоверность полного давления оказалась выше достоверностей скорости потока и южного компонента магнитного поля. Таким образом, по важности южный компонент магнитного поля уступает уже не только электрическому полю утро-вечер, но и полному давлению солнечного ветра, что идет в разрез с привычным представ-

лением о механизме развития геомагнитной бури [1, 4], где она играет определяющую роль.

Разработанный автором по результатам данной работы программный код, позволяющий предсказывать  $D_{st}$ -индекс на 9 ч вперед, загружен на сайт <http://www.spaceweather.org.ua/>, и в ближайшее время будет готов к использованию.

Автор благодарит О. К. Черемных за постановку задачи, В. М. Кунцевича, Ю. А. Селиванова, В. А. Яценко и других сотрудников ИКИ НАНУ и НКАУ за обсуждение результатов и ценные замечания.

1. Акасофу С.-И., Чепмен С. Солнечно-земная физика. — М.: Мир, 1975.—900 с.
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 // *Geophys. Res. Lett.*—1990.—17.—P. 41—44.
3. Campbell W. H. Geomagnetic storms, the Dst ring-current myth and lognormal distributions // *J. Atmos. and Terr. Phys.*—1996.—58, N 10.—P. 1171—1187.
4. Dungey J. W. Interactions of solar plasma with the geomagnetic field // *Planet. Space Sci.*—1963.—10.—P. 233—237.
5. Feldstein Y. I. Modelling of magnetic field of magnetospheric ring current as function of interplanetary medium parameters // *Space Sci. Rev.*—1992.—59.—P. 83—105.
6. Temerin M., Xinlin Li. A new model for the prediction of  $D_{st}$  on the basis of the solar wind // *J. Geophys. Res.*—2002.—107, N A12.—P. SMP 31-1.
7. Wu J.-G., Lundstedt H. Geomagnetic storm predictions from solar wind data with the use of dynamic neural networks // *J. Geophys. Res.*—1997.—102, N A7.—P. 14255—14268.

#### $D_{st}$ PREDICTION USING THE LINEAR REGRESSION ANALYSIS

*A. S. Parnowski*

The aim of our investigation is to derive the phenomenological regression of  $D_{st}$  in relation to solar wind parameters and to use the regression for  $D_{st}$  prediction. The magnetosphere is considered as a black box, i.e., no models or assumptions are used. We derived the regression providing predicting  $D_{st}$  for nine hours ahead. The correlation between predicted and measured  $D_{st}$  values varies from 98.6 % for one-hour prediction to 79.3 % for nine-hour prediction. We also discuss how the form of statistically significant regressors can help understanding the physical mechanism of solar wind influence on geomagnetic activity.