

Использование нейросетевой сегментации многомерных временных рядов для анализа геомагнитных возмущений

С. А. Доленко^а, И. Н. Мягкова^б, И. Г. Персианцев

Московский государственный университет имени М. В. Ломоносова,
Научно-исследовательский институт ядерной физики имени Д. В. Скобельцына (НИИЯФ
МГУ). Россия, 119991, Москва, Ленинские горы, д. 1, стр. 2.

E-mail: ^а dolenko@srd.sinp.msu.ru, ^б irina@srd.sinp.msu.ru

Статья поступила: 13.11.2015; подписана в печать 20.04.2016.

Настоящая работа посвящена предварительному анализу при помощи искусственных нейронных сетей результатов сегментации многомерных временных рядов следующих данных: параметров солнечного ветра, межпланетного поля и Dst-индекса, в период с ноября 1997 г. по март 2014 г. Показано, что данный метод исследования позволит продвинуться в понимании динамики состояний магнитосферы Земли.

Ключевые слова: геомагнитные возмущения, Dst-вариация, солнечный ветер, межпланетное магнитное поле, искусственные нейронные сети.

УДК: 550.385.4. PACS: 92.60.hw, 94.30.Xy, 96.60.Vg.

Введение

Прогнозирование состояния магнитосферы Земли — это сложная комплексная задача, поскольку магнитосфера представляет собой многокомпонентную динамическую систему, открытую для внешних воздействий. Динамика состояния системы «межпланетное космическое пространство — солнечный ветер — магнитосфера Земли» может быть описана многомерным временным рядом (ВР), который включает параметры межпланетного магнитного поля (ММП) — модуль вектора **В** ММП и три его компоненты; солнечного ветра (СВ) — скорость, плотность, температура плазмы СВ; геомагнитные индексы — Dst и Kp-индексы. В настоящей работе рассматриваются возможности исследования динамики магнитосферы Земли при помощи сегментации многомерных временных рядов данных о параметрах СВ и ММП, а также геомагнитного индекса Dst.

Уровень возмущенности магнитосферы Земли традиционно оценивают с помощью геомагнитных индексов. Амплитуду планетарных возмущений во время геомагнитных бурь, захватывающих всю магнитосферу, принято характеризовать с помощью амплитуды Dst-вариации, которая представляет собой максимальное отклонение вариации магнитного поля Земли от спокойного уровня, усредненное по значениям, измеренным на контрольной цепочке магнитных станций, расположенных в низких широтах, т.е. вблизи экватора [1]. Информацию об экспериментально измеренных значениях Dst-вариации предоставляет отделение Мирового центра данных (World Data Center), находящееся в Киото (Япония) [2]. Источниками геомагнитных возмущений могут быть как корональные выбросы массы (КВМ), достигающие орбиты Земли, так и высокоскоростные потоки СВ. Необходимым, а возможно и достаточным условием возникновения геомагнитного возмущения является присутствие южной (отрица-

тельной) z -компоненты ММП, т.е. отрицательно B_z , при наличии которого возможно передача энергии от солнечного ветра в магнитосферу.

Возмущения магнитного поля Земли (геомагнитные возмущения) или, как их чаще называют, магнитные бури — это одно из основных проявлений космической погоды. При этом под космической погодой в настоящее время принято понимать процессы и явления, происходящие на Солнце, в солнечном ветре, магнитосфере и ионосфере Земли, которые могут влиять на работу бортовых и наземных технологических систем, а также на самочувствие и здоровье людей. Экспериментальные исследования показывают, что геомагнитные возмущения оказывают непосредственное влияние на состояние околоземного космического пространства (ОКП) [3]. Известно также, что магнитные бури могут вызывать нарушения в работе телеграфных линий, трубопроводов, линий электропередач и энергосетей, приводить к нарушениям радиосвязи; во многих работах исследуются вопросы влияния магнитных бурь на самочувствие людей. Из всего вышесказанного следует, что детальное изучение процессов, происходящих в магнитосфере Земли, в том числе с целью улучшения методов прогнозирования ее состояния, является весьма важной задачей.

Цель настоящей работы — предварительный анализ результатов сегментации при помощи искусственных нейронных сетей (ИНС) многомерных временных рядов данных о параметрах солнечного ветра, межпланетного поля и геомагнитных индексов (амплитуды Dst-вариации) с декабря 1997 г. по март 2014 г.

1. Характеристика данных, составляющих многомерный временной ряд, описывающий состояние магнитосферы Земли

Так как основной причиной возникновения возмущений магнитосферы Земли являются вариации

параметров СВ и ММП [4], то в качестве входных параметров при сегментации ВР (как и при прогнозировании значений Dst-вариации) представляется разумным использовать параметры плазмы СВ, параметры ММП, а также саму Dst-вариацию.

При этом для того чтобы измеряемые вариации параметров СВ и ММП были заведомо временными, а не пространственными, необходимо, чтобы космический аппарат (КА), на борту которого выполняются измерения параметров СВ и ММП, постоянно находился в одной и той же точке, неподвижной как относительно Земли, так и относительно Солнца. Это условие выполняется при нахождении КА в точке Лагранжа L_1 между Солнцем и Землей, на расстоянии около 1,5 млн км от последней. На сегодняшний день в L_1 находятся несколько КА, которые предназначены для измерения и передачи на Землю в реальном времени параметров СВ и ММП. В настоящей работе используются данные КА ACE (Advanced composition explorer) [5].

В качестве примера влияния изменения параметров СВ и ММП на состояние магнитосферы Земли, а именно на Dst-индекс (верхняя панель рис. 1), показаны изменения скорости СВ и параметров ММП — полного вектора ММП $\mathbf{B}$ и его z -компоненты B_z (нижняя панель), происходившие в ноябре–декабре 2013 г.

На рис. 1 видно, что возрастания скорости СВ, происходящие в результате как прихода к Земле корональных выбросов массы (КВМ), так и высокоскоростных потоков СВ из корональных дыр, приводят к возникновению геомагнитных возмущений. Нетрудно заметить, что при сходных параметрах внешних воздействий на магнитосферу (скорости СВ, величине вектора и z -компоненты ММП) возникают геомагнитные возмущения разной амплитуды, длительности и с разным временным профилем; наоборот, похожий профиль Dst-вариации вызыва-

ется совершенно разными внешними воздействиями, т. е. одна и та же амплитуда Dst-вариации возникает в разных условиях.

Как видно из рис. 1, за ноябрь–декабрь 2013 г. наблюдалось пять случаев возрастания скорости СВ выше 500 км/с (все они обозначены на рис. 1 римскими цифрами). Источником первых трех — I, II, III — были корональные выбросы массы (КВМ), дошедшие до орбиты Земли, а двух оставшихся — IV и V — высокоскоростные потоки СВ из корональных дыр (КД). Но, как видно из рис. 1, только два из указанных возрастаний (первое и четвертое) вызвали геомагнитные бури с амплитудой больше – 50 нТл, еще после двух возрастаний скорости СВ (второго и третьего) геомагнитные возмущения (амплитуда Dst-вариации) были малы, а пятое возрастание вызвало слабую магнитную бурю с максимальной амплитудой – 40 нТл, причем следующая похожая буря (25 декабря) произошла при скорости СВ меньше 400 км/с.

Мы предполагаем, что столь различная реакция магнитосферы Земли на внешние воздействия является следствием различных ее состояний, и надеемся, что с помощью кластеризации параметров СВ и ММП мы сможем объективно выделить принципиально различные состояния земной магнитосферы. Отметим, что в литературе уже имеются сообщения об успешном применении ИНС Кохонена для кластеризации массива данных отобранных по определенным критериям геомагнитных возмущений, с целью выделения нескольких типов геомагнитных возмущений, различных с точки зрения вызывающих их физических процессов [6]. В настоящей работе ставится несколько иная задача: с помощью кластеризации часовых данных о состоянии магнитосферы Земли осуществить сегментацию всего ВР, описывающего ее состояние. Это, с одной стороны, возможно, позволит продвинуться

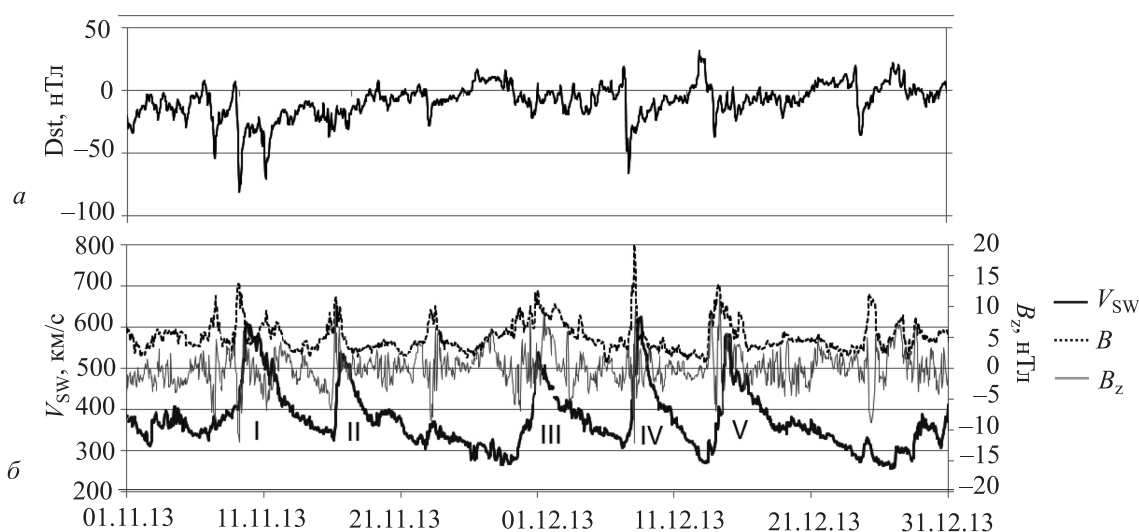


Рис. 1. Временные зависимости: а — Dst-вариации, б — скорости солнечного ветра V_{sw} (сплошная черная линия), амплитуды модуля вектора $\mathbf{B}$ (пунктир), его z -компоненты B_z (серая линия). Период — ноябрь–декабрь 2013 г.

в понимании механизмов возникновения и развития геомагнитных возмущений, а с другой, поможет лучше осуществлять прогнозирование при обучении ИНС не на всем массиве данных, а по отдельности для каждого кластера.

С учетом всего изложенного в качестве входных данных в настоящей работе использовались часовые значения следующих величин: параметры СВ (скорость, плотность, температура) с прибора SWEPAM KA ACE [7], три компоненты ММП в системе GSM и его модуль с прибора MAG KA ACE [7], а также значение Dst-вариации по данным Всемирного центра данных в Киото [8]. Применялись не очищенные и обработанные данные Level 2, а близкие к реальному времени данные (Browse/Quicklook), по двум причинам. Во-первых, модуль классификации должен в перспективе стать составной частью системы реального времени по прогнозированию значений Dst-индекса, работающей на сайте Центра оперативного космического мониторинга НИИЯФ МГУ (<http://swx.sinp.msu.ru/models/dst.php>), с использованием отдельных прогнозирующих моделей — своей для каждого кластера данных, и должен уметь работать с данными того качества, которое возможно в режиме реального времени. Во-вторых, данные Level 2 имеют значительно большее число пропусков.

Пропуски в данных не устранялись, использовались только примеры без пропусков. Полный размер массива данных для кластеризации (данные с конца октября 1997 г. по конец марта 2014 г.) за вычетом пропущенных данных составил 144 155 примеров. Каждая переменная по отдельности нормировалась приведением к нулевому среднему и единичному стандартному отклонениям.

Заметим, что состояние магнитосферы как динамической системы полностью не описывается комбинацией только текущих значений указанных параметров, но зависит также от предыстории их изменения. Для учета такой предыстории осуществлялось погружение временных рядов указанных параметров на глубину в 24 ч.

2. Методика сегментации многомерных временных рядов данных о параметрах солнечного ветра, межпланетного поля и Dst-вариации при помощи ИНС, примеры самоорганизующихся карт Кохонена

Как уже было сказано, основной нашей задачей является кластеризация (с помощью нейронных сетей Кохонена и альтернативных методов) данных многомерных ВР, описывающих изменение геомагнитных индексов, параметров ММП и СВ, измеренных за пределами магнитосферы (в так называемой точке Лагранжа L_1 между Солнцем и Землей). Цель кластеризации — выявление различных областей взаимосвязанных состояний цепочки «гелиосфера — солнечный ветер — магнитосфера Земли». Отметим, что поскольку кластеризации подлежат данные, не

являющиеся независимыми, а представляющие собой различные точки многомерного ВР, решаемая задача, по сути, представляет собой задачу сегментации ВР, что необходимо учитывать при осуществлении кластеризации.

Известно, что любой алгоритм кластеризации использует понятие расстояния между точками кластеризуемого множества. Из-за этого результат кластеризации может существенно зависеть как от набора входных параметров, так и от их взаимной нормировки. Поэтому для решения основной задачи необходимо определить оптимальный набор входных параметров (и способов их нормировки), в пространстве которых получаемая кластеризация является наиболее контрастной и объяснимой с точки зрения ранее известных физических процессов.

В настоящей работе кластеризация данных применяется для адаптивного выделения областей в фазовом пространстве динамической системы «гелиосфера — солнечный ветер — магнитосфера Земли», соответствующих разным классам (типам состояний) этой системы. Фактически речь идет о сегментации многомерного ВР состояний. В качестве подхода к первичной сегментации ВР можно использовать кластеризацию входящих в него состояний в небольшое количество классов (кластеров), с последующим возможным учетом ограничений, накладываемых с целью устранения возможных «переколебаний» по ходу ВР при отнесении каждого состояния к тому или иному кластеру. При этом первичная кластеризация (до наложения ограничений) вообще не требует никаких предположений о времени существования и характере смены типов состояний.

Авторы настоящей работы имеют опыт работы в области применения ИНС к космофизическим данным [9]. Кроме того, была создана собственная система прогноза амплитуды Dst-вариации на один час вперед с помощью модели, основанной на нейросетевых технологиях [10].

В качестве основного алгоритма кластеризации в настоящей работе использовалась НС Кохонена [11, 12]. Результатом обучения сети Кохонена является распределение немногочисленных нейронов сети в исходном пространстве признаков (ИПП) в соответствии с плотностью распределения многочисленных примеров обучающей выборки в том же пространстве. Это позволяет, задавая количество нейронов N сети Кохонена, осуществлять кластеризацию примеров, выделяя не более N кластеров. Таким образом, при малом N НС Кохонена применяется для решения задачи кластеризации. Обычно такой режим используется при $N \leq 20$.

При сравнительно больших значениях N и двумерной топологии сети можно рассматривать пространство сети как двумерную карту (так называемую самоорганизующуюся карту Кохонена — СОК)

и анализировать положение отдельных примеров на этой карте.

Самоорганизующаяся карта Кохонена осуществляет нелинейное отображение ИПП, имеющего высокую размерность, на двумерное пространство карты Кохонена, с сохранением соотношений близости между разными примерами. Это позволяет эффективно использовать такие карты для визуализации данных, первоначально заданных в многомерных пространствах. В частности, на двумерной карте могут быть достаточно легко проведены границы между кластерами, соответствующими различным областям ИПП. Анализ дополнительно облегчается, если взять в качестве третьей координаты на карте значение того или иного анализируемого признака ИПП.

В настоящей работе была использована реализация СОК и НС Кохонена в программном пакете Deductor 5.3 [13]. Кластеризация с помощью СОК в пакете Deductor осуществляется в два этапа. На первом этапе строится двумерная СОК заданного размера, т.е. осуществляется кластеризация двумерной сетью Кохонена в заданное большое число первичных кластеров — ячеек карты $K = X \times Y$ (где X и Y — размеры СОК), а полученная двумерная карта может быть использована для визуализации данных (рис. 2). На втором этапе объектами кластеризации являются уже не исходные примеры, а центроиды первичных кластеров, полученных на первом этапе; при этом кластеризация в иско-

мое небольшое число кластеров N осуществляется алгоритмом k -средних. Классификация исходных примеров в полученные таким образом N классов осуществляется по минимуму расстояния от данного примера до центроидов выделенных N кластеров. В настоящей работе СОК состояла из 16×12 или из 8×6 шестиугольных ячеек с евклидовой функцией расстояния. Начальная скорость обучения составляла 0.5, конечная скорость обучения — 0.005. Длительность обучения — 500 эпох. Количество искоемых кластеров варьировалось от 3 до 5.

В таблице представлены средние для каждого кластера значения основных физических признаков — индекса Dst, скорости СВ, z -компоненты ММП в системе GSM и модуля ММП для шести вариантов кластеризации — в 3, 4 и 5 кластеров с помощью СОК и алгоритма k -средних.

На рис. 2 показаны СОК с ячейками 16×12 , полученные для трех кластеров: для текущих значений индекса Dst (слева сверху), амплитуды вектора ММП (справа сверху) и скорости СВ (слева снизу). Окраска каждой ячейки карты отвечает среднему значению соответствующего параметра по всем примерам, попавшим в данную ячейку. Четвертая панель показывает принадлежность ячеек к кластерам. Подчеркнем, что на всех четырех картах изображена одна и та же сеть Кохонена, а одинаковые координаты на разных картах имеет одна и та же ячейка, включающая одни и те же примеры. Более высокие значения амплитуд параметров СВ и ММП пока-

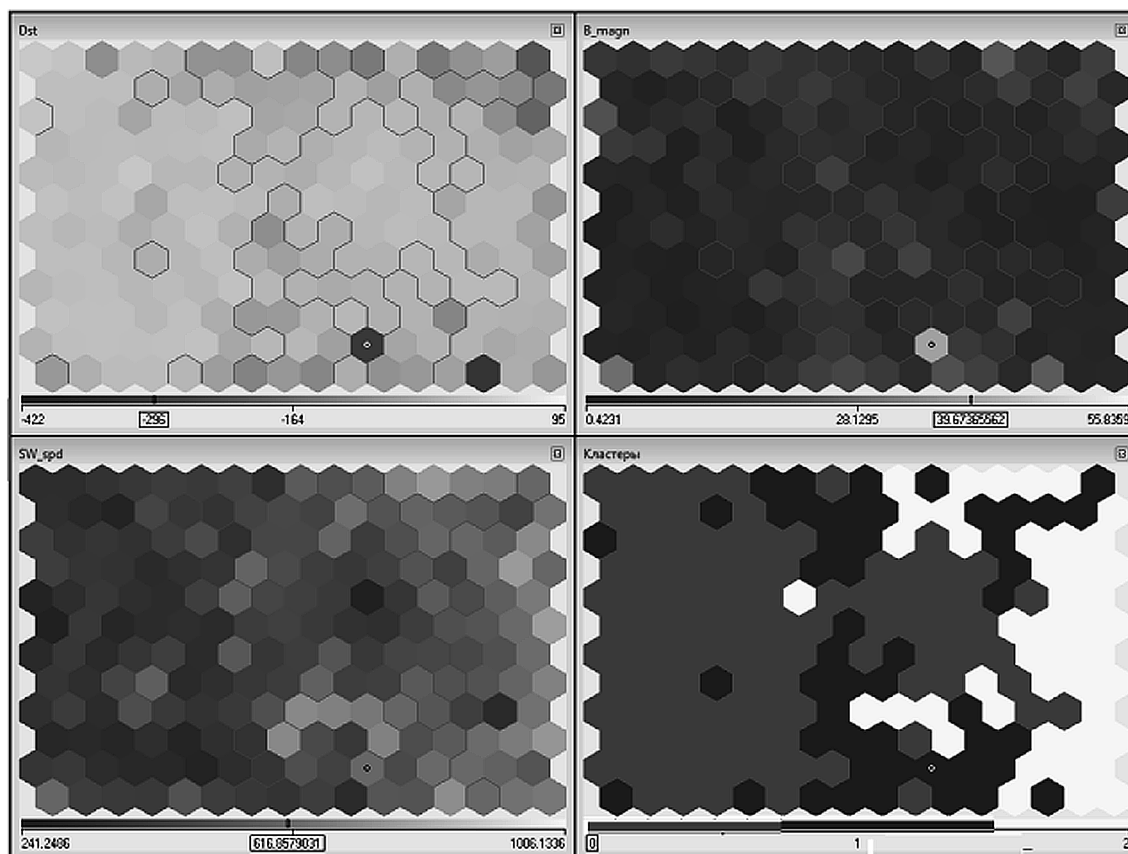


Рис. 2. Самоорганизующиеся карты для исследования Dst для трех кластеров

**Средние по кластерам значения основных физических признаков
для разбиений на 3, 4 и 5 кластеров с помощью СОК
и с помощью алгоритма k -средних**

Алгоритм, число кластеров	Кластер	Количество примеров	Dst, нТ	Скорость СВ, км/с	B_z , нТ	$ \mathbf{B} $, нТ
k -средних, 3	0	59221	−3.0	350.8	0.04	5.5
	1	47271	−16.1	449.1	−0.01	6.3
	2	20165	−25.8	601.2	−0.07	5.6
k -средних, 4	0	56503	−2.6	348.6	0.04	5.4
	1	37123	−11.3	435.9	0.02	5.0
	2	19008	−24.8	604.1	−0.03	5.5
	3	14023	−30.1	482.4	−0.12	9.8
k -средних, 5	0	40212	−0.7	335.0	0.03	5.0
	1	28412	−12.1	413.2	−0.09	6.5
	2	24999	−13.6	422.9	0.18	6.9
	3	20085	−18.9	504.7	−0.08	5.2
	4	12949	−28.2	634.2	−0.06	5.7
СОК, 3	0	102293	−7.2	389.6	0.03	5.6
	1	17850	−24.9	607.4	0.00	5.4
	2	6514	−42.6	528.0	−0.44	9.9
СОК, 4	0	76564	−4.2	365.1	0.07	5.4
	1	27704	−20.5	468.9	−0.13	7.2
	2	21614	−22.4	590.0	0.02	5.3
	3	775	−103.4	557.6	−2.28	12.9
СОК, 5	0	79226	−4.3	367.8	0.08	5.5
	1	22229	−16.8	491.1	0.11	6.4
	2	17216	−23.8	607.8	0.01	5.4
	3	7250	−35.3	442.3	−0.89	8.1
	4	736	−101.4	553.5	−2.38	12.5

заны более светлым цветом (см. цветовую легенду под каждой из панелей). Например, кружком на всех картах отмечена ячейка, в которую попали примеры с большим по модулю отрицательным средним значением Dst, большой средней амплитудой вектора ММП и средней скоростью СВ около 700 км/с; эта ячейка попала в кластер 1, помеченный на правой нижней карте темно-серым цветом.

Рис. 2 показывает, что кластеру «0» (серый цвет на правой нижней панели) соответствуют минимальные значения скорости СВ и небольшие амплитуды Dst (таблица), что мы интерпретируем как отсутствие бурь или спокойный период. Кластер 2 (белый цвет на правой нижней панели) характеризуется максимальными амплитудами Dst (более темный цвет, поскольку во время бурь значение Dst отрицательно), высокими значениями скорости СВ (светлые участки) и максимальными значениями модуля ММП. Это кластер интерпретируется нами как главная фаза бури. Оставшийся кластер 1 с высокой скоростью СВ и промежуточными значениями Dst мы относим к фазе восстановления магнитной бури. Представляется, что это и есть наиболее простой вариант кластеризации состояний магнитосферы Земли.

Сделанные предположения были проверены при помощи сравнения размеров получившихся кластеров с изменением числа бурь с течением времени (в зависимости от года). Результаты сравнения приведены на рис. 3.

Из рис. 3 видно, что кластер 0, который мы интерпретировали как отсутствие бурь, действительно является таковым, поскольку изменение его размера хорошо антикоррелирует с числом бурь. Наиболее показателен тут минимальный размер кластера 0 в 2003 г., когда в период экстремальных солнечных событий наблюдался максимум числа бурь, и его максимальный размер в 2009 г., когда имел место аномальный минимум солнечной и соответственно геомагнитной активности. Размер кластера 1 — «фаза восстановления», ожидаемо антикоррелирует с размером кластера 0 и соответственно коррелирует с числом бурь. Кластер 2 слишком мал, чтобы что-то уверенно утверждать о нем, хотя минимум в 2009 г., когда скорость СВ и число и амплитуда геомагнитных возмущений были аномально низкими, на рис. 3 просматривается.

Таким образом, мы предполагаем, что использование НС и самоорганизующихся карт Кохонена позволит осуществить кластеризацию и визуализацию

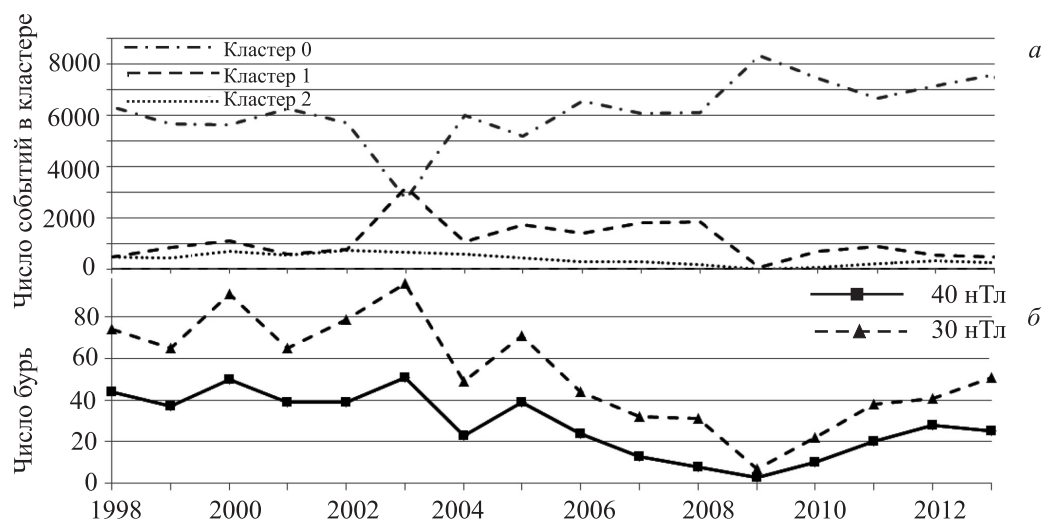


Рис. 3. Сравнение изменения размеров кластеров (а) с изменением числа бурь разной амплитуды (б) с течением времени (в зависимости от года)

многомерных данных, характеризующих состояние рассматриваемой динамической системы. Помимо этого нами применялись более традиционные алгоритмы кластеризации, такие как алгоритм k -средних. Проведено сравнение результатов работы.

3. Примеры результатов сегментации для различных вариантов кластеризации и разных периодов солнечной активности

Ниже представлены примеры результатов сегментации для разных эпох солнечной, а следовательно, и геомагнитной активности (ГА), демонстрирующие особенности работы разных алгоритмов кластеризации. Ставится задача выяснить, одинаково ли работают предложенные методы сегментации в эти периоды.

На рис. 4, а–в приведены результаты сегментации ВР НС Кохонена (пунктир) и методом k -средних (сплошная линия) для периода времени, близкого к максимуму солнечной активности (СА) с 11 марта по 20 мая 2012 г. — варианты кластеризации с тремя, четырьмя и пятью кластерами, полученными при помощи сетей Кохонена (сплошная серая линия) и k -средних (пунктир). На рис. 4, г приведены временные зависимости скорости СВ, модуля вектора ММП и вертикальной компоненты ММП.

Как видно из рис. 4, при помощи сегментации за данный период времени обоими методами были уверенно выделены моменты влияния на магнитосферу Земли высокоскоростных потоков СВ.

Для случая трех кластеров (верхняя панель, серая линия), по нашему мнению, более адекватно работает метод сетей Кохонена; полученные кластеры можно достаточно уверенно (по крайней мере, для данного периода) отождествить следующим образом, как уже было указано выше: 0 — нет бури, 1 — фаза восстановления, 2 — короткая главная фаза. Для метода k -средних главная фаза, как видно из рисунка, попадает в тот же кластер, что и

окончание фазы восстановления, что с физической точки зрения неприемлемо.

Для случая четырех кластеров, наоборот, адекватнее работает метод k -средних — полученные кластеры мы отождествили как: 0 — нет бури, 1 — окончание фазы восстановления, 2 — начало фазы восстановления и 3 — главная фаза. Для метода сетей Кохонена, как видно из рис. 4, б, для данного периода времени удалось выделить только три кластера: 0 — нет бури, 1 — главная фаза и окончание фазы восстановления (как это было для метода k -средних в случае трех кластеров) и 2 — начало фазы восстановления. Из рис. 4 также видно, что начальную фазу бури и ее внезапное начало (sudden storm commencement — SSC) на данном этапе развития алгоритмов сегментации, к сожалению, выделить не удалось.

Разбиение на 5 кластеров (рис. 4, в), на наш взгляд, не имеет значимых преимуществ по сравнению с четырьмя кластерами, но выглядит менее устойчивым к случайным флуктуациям. Также следует отметить, что для метода СОК в указанном временном периоде удалось выделить только 4 кластера.

Из рис. 4, очевидно, следует, что в используемом варианте кластеризации существенное влияние на результат оказывает величина скорости СВ, из-за чего буря, произошедшая в начале апреля 2012 г. и имевшая амплитуду — 53 нТ, но происходившая при невысокой скорости СВ, была отнесена обоими методами к кластеру 0, который мы идентифицировали как «нет бури». Это очевидный недостаток используемых методов. В будущем будут предприняты попытки по преодолению этой проблемы.

Рассмотренный период времени — с 11 марта по 20 марта 2012 г. — соответствует максимуму текущего цикла солнечной активности (СА), когда источником большей части геомагнитных возмущений с большой амплитудой были КВМ.

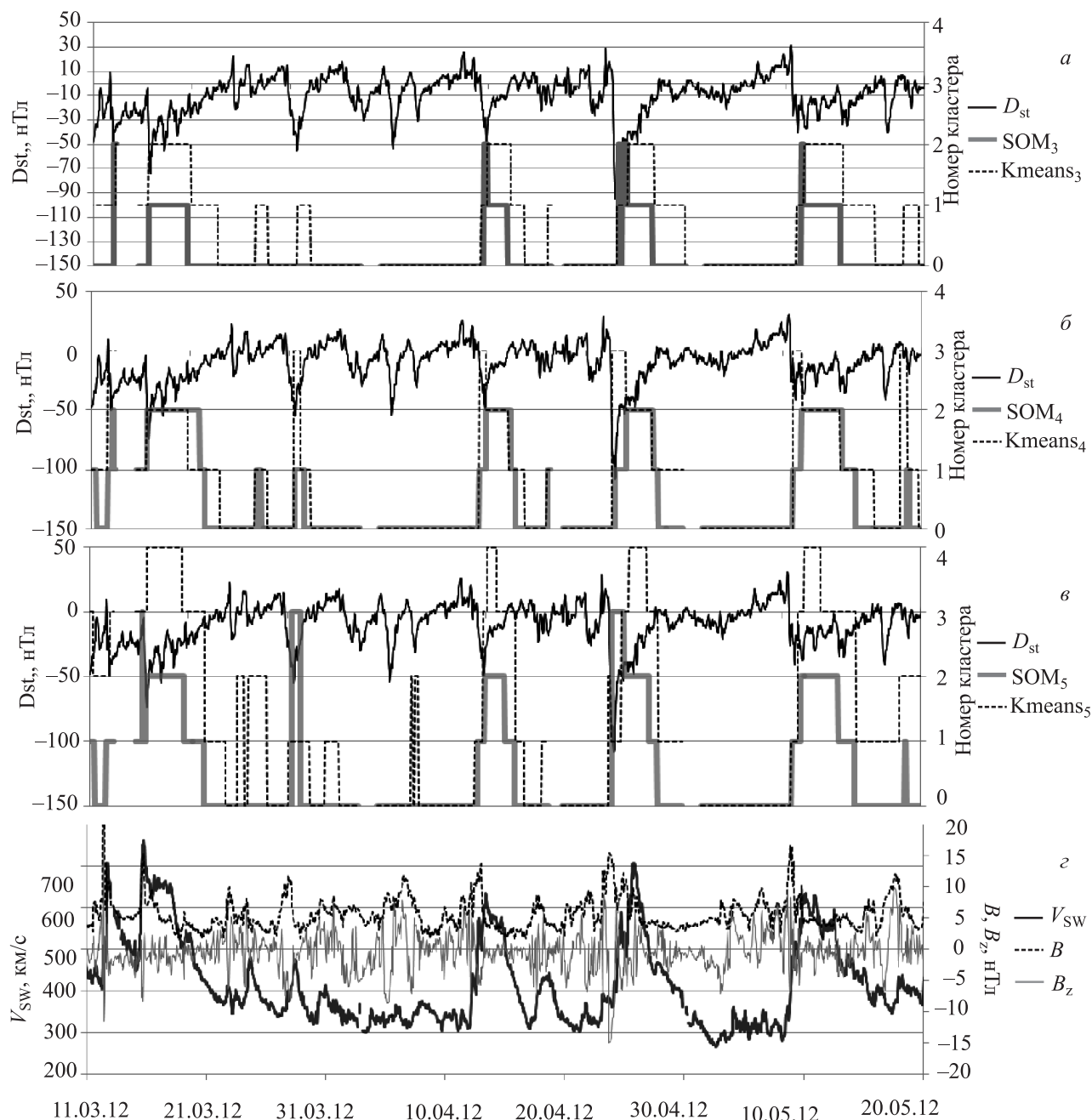


Рис. 4. Пример результатов кластеризации на 3 (а), 4 (б), 5 (в) кластерах и параметры ММП и СВ (г) за период, соответствующий максимуму СА с 11 марта по 20 мая 2012 г. V_{sw} — скорость СВ; SOM_3 , SOM_4 , SOM_5 — результат сегментации на три, четыре и пять классов с помощью нейронных сетей Кохонена; $K_{mean\ 3}$, $K_{mean\ 4}$, $K_{mean\ 5}$ — результат сегментации на три, четыре и пять классов с помощью алгоритма k -средних

Чтобы убедиться, что предлагаемые методы кластеризации работают и в другие периоды СА, т.е. для других состояний магнитосферы и среднего уровня геомагнитной возмущенности, мы применили кластеризацию на три и четыре кластера методами СОК и k -средних для других периодов времени. Были выбраны периоды с 11 февраля по 20 мая 2007 г. (спад 23-го цикла СА, рис. 5) и с 5 августа по 31 ноября 2008 г. (начало минимума СА, рис. 6).

Представленные на рис. 5 и 6 возрастания скорости потока СВ, вызвавшие геомагнитные возмущения, в большинстве своем являются результатом

прихода к Земле высокоскоростных потоков СВ. Как видно из этих рисунков, качество и особенности прогнозирования методами СОК и k -средних сохраняются, несмотря на уменьшение амплитуды геомагнитных возмущений при уменьшении уровня СА.

4. Обсуждение

Как можно заключить из сравнения рис. 4, 5 и 6, основные черты сегментации, отмеченные нами для марта-мая 2013 г. (максимума СА), остаются теми же самыми как для начала спада СА (февраль-май 2007 г.), так и для начала минимума СА (август-но-

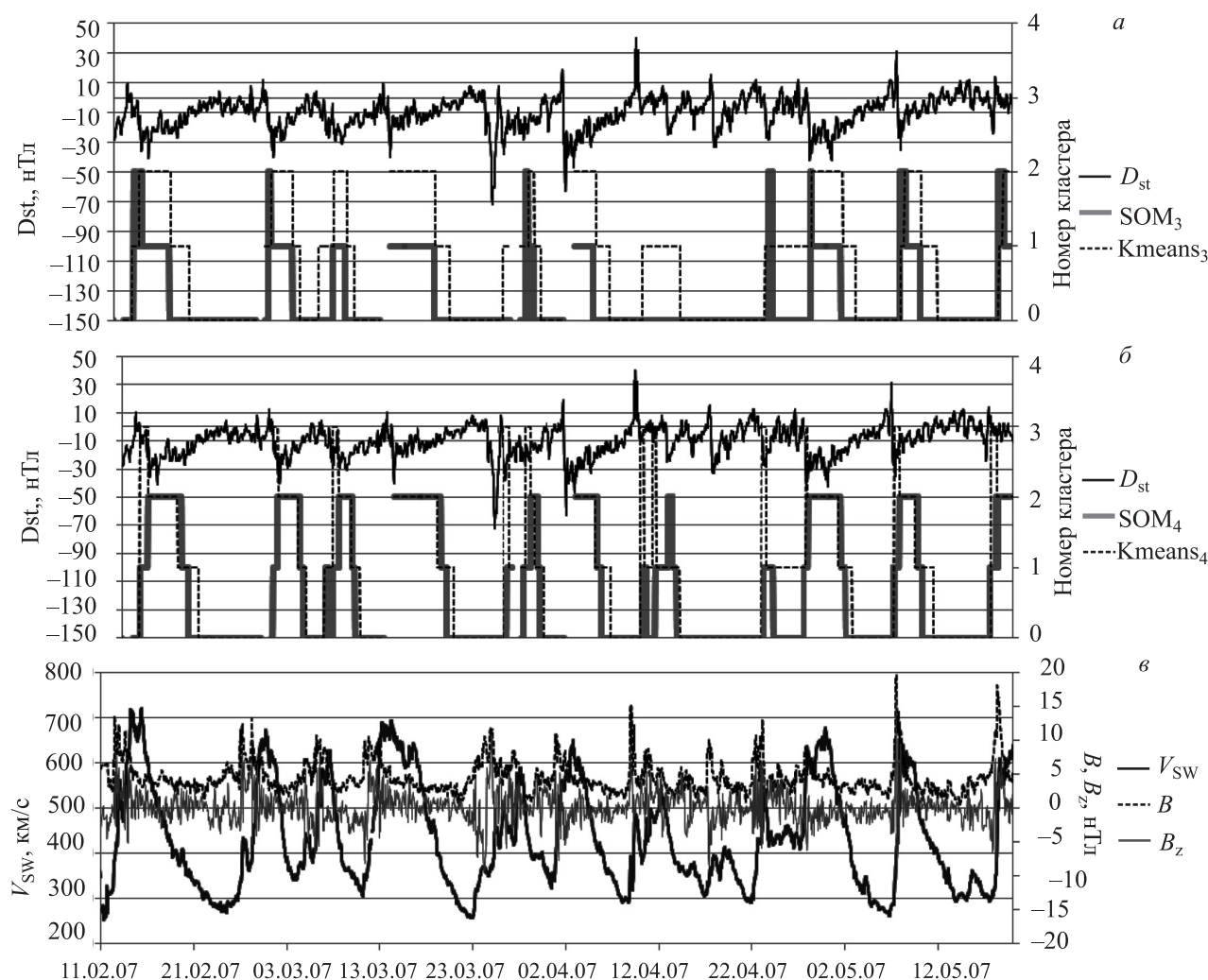


Рис. 5. Пример результатов кластеризации на 3 (а) и 4 (б) кластера и параметры ММП и СВ (в) за период, соответствующий началу спада СА 23-го цикла с 11 февраля по 20 мая 2007 г. Обозначения соответствуют обозначениям на рис. 4

ябрь 2008 г.). Ведущим параметром сегментации во всех трех случаях является скорость СВ. В результате геомагнитные возмущения, происходящие при низкой скорости СВ при длительных отрицательных значениях B_z (z -компоненты ММП), оказываются в кластере, соответствующем невозмущенному состоянию магнитосферы.

Значения Dst-индекса и модуля ММП также являются существенными классификационными признаками (см. табл.). Впрочем, как и ожидалось, не всегда различным кластерам соответствуют сильно различающиеся средние значения признаков, так как не существует четких и однозначных взаимосвязей между значениями признаков (например, нельзя утверждать, что большим по величине отрицательным значениям индекса Dst в любой момент времени соответствуют высокая скорость СВ и большое значение модуля ММП), в то время как алгоритмы кластеризации учитывают значения всех признаков сразу.

Следует отметить, что для метода k -средних отнесение к «буревым» кластерам периодов с незначительной геомагнитной возмущенностью происходит

чаще, чем для метода ИНС Кохонена, т.е. первый более чувствителен. Это может быть плюсом для слабовозмущенных периодов и минусом при частых и сильных возмущениях, поскольку тот же метод зачастую относит к «буревым» кластерам моменты, когда геомагнитные возмущения отсутствуют.

Для всех трех периодов времени, приведенных на рис. 4, 5 и 6, при разделении на три кластера в случае применения метода k -средних главная фаза магнитной бури во многих случаях попадает в тот же кластер, что и окончание фазы восстановления. При использовании СОК такой эффект отсутствует. Данную проблему планируется детально исследовать при дальнейшей работе.

Следует также подчеркнуть, что во всех случаях в разные кластеры попадают не различные типы геомагнитных возмущений, а их различные фазы (или отсутствие возмущения). Это означает, что различия между физическими процессами, происходящими на разных фазах геомагнитных возмущений, оказываются сильнее, чем различия между физическими процессами, происходящими во время геомагнитных возмущений разных типов [6].

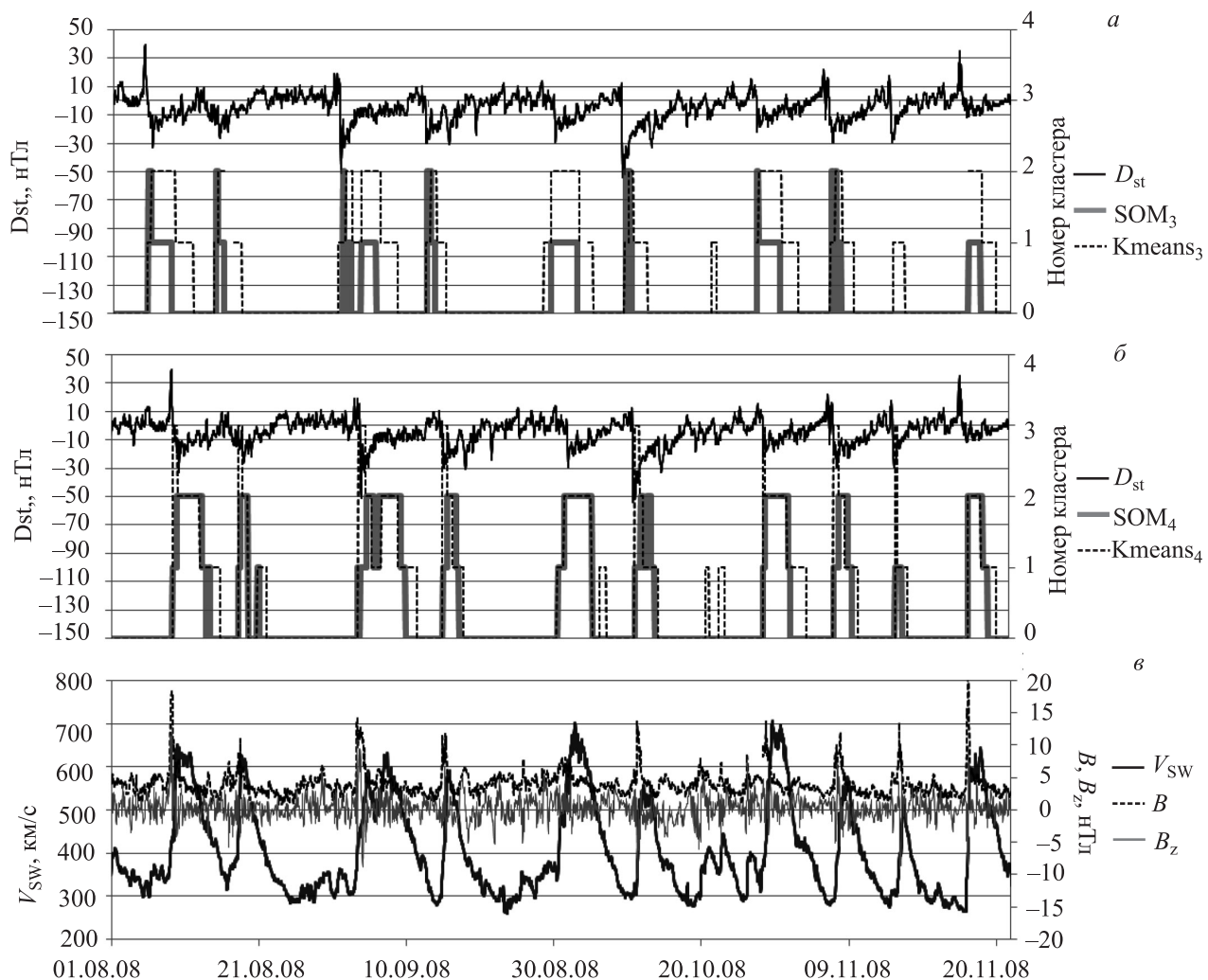


Рис. 6. То же за период, соответствующий окончанию спада и началу минимума СА 23-го цикла с 1 августа по 21 ноября 2008 г. Обозначения соответствуют обозначениям на рис. 4

Заключение

В работе представлены первые результаты сегментации многомерных временных рядов параметров солнечного ветра (скорость, плотность и температура СВ), межпланетного магнитного поля (амплитуда и три компоненты вектора) и геомагнитного индекса Dst путем кластеризации данных при помощи сетей Кохонена и k -средних и проведено их сравнение.

Обнаружено, что ведущим параметром сегментации является скорость СВ, что создает проблемы при кластеризации геомагнитных возмущений, происходящих при низкой скорости СВ за счет длительной отрицательной z -компоненты ММП.

Показано, что полученная при помощи данных методов метода сегментация ВР может иметь адекватную физическую интерпретацию в случае как СОК, так и k -средних, что говорит о применимости данного метода для исследования геомагнитных возмущений.

Наилучшие с точки зрения физического смысла варианты сегментации были получены при помощи СОК с топологией 16×12 ячеек, с тремя кластерами и методом k -средних для 4 кластеров.

Сравнение результатов сегментации, полученных в разные периоды СА, показало, что предлагаемая методика сегментации может быть применима для разных фаз цикла СА, по крайней мере, для 23-го и 24-го циклов СА.

Работа выполнена при финансовой поддержке РФФИ (грант 14-01-00293-а).

Список литературы

1. Лазутин Л.Л. Мировые и полярные магнитные бури. МГУ, 2012.
2. Интернет-портал Мирового центра данных (Geomagnetic Equatorial Dst Index Home Page. <http://wdc.kugi.kyoto-u.ac.jp/dstdir/index.html>).
3. Мягкова И.Н., Шугай Ю.С., Веселовский И.С., Яковчук О.С. // Астрон. вестн. 2013. № 7. С.141. (Myagkova I.N., Shugay Yu.S., Veselovsky I.S., Yakovchouk O.S. // Solar System Research. 2013. **47**, N 2. P. 141.) 10.1134/S0038094613020068
4. Burton R.K., McPherron R.L., Russell C.T. // J. Geophys. Res. 1975. **80**. P. 4204.
5. Real-Time Solar Wind Data. <http://www.swpc.noaa.gov/products/ace-real-time-solar-wind>

6. Бархатов Н.А., Левитин А.Е., Ревунова Е.А. // Геомагнетизм и аэрономия. 2014. **54**, № 2. С. 185.
7. Caltech, Advanced Composition Explorer Science Center. <http://www.srl.caltech.edu/ACE/ASC/>
8. Всемирный центр данных в Киото (Geomagnetic Data Service). <http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html>.
9. Dolenko S.A., Orlov Yu.V., Persiantsev I.G., Shugai Ju.S. // Lecture Notes in Computer Science. 2005. **3697**. P. 527.
10. Прогноз значения амплитуды Dst-вариации в режиме реального времени // Сайт Центра анализа космической погоды НИИЯФ МГУ. <http://swx.sinp.msu.ru/models/dst.php>
11. Kohonen T. Self-Organizing Maps. 3d ed. Springer, 2001.
12. Vrieze O.J. Kohonen Network // Artificial Neural Networks. An introduction to ANN theory and practice. Lecture Notes in Computer Science / Ed. by P.J. Braspenning, F. Thuijsman, A.J.M.M. Weijters. Springer, 1995. **931**. P. 83.
13. Аналитическая платформа Deductor 5.3. <http://www.basegroup.ru>

The use of artificial neural network segmentation of multivariate time series for the analysis of geomagnetic disturbances

S. A. Dolenko^a, I. N. Myagkova^b, I. G. Persiantsev

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow 119991, Russia.
E-mail: ^a dolenko@srd.sinp.msu.ru, ^b irina@srd.sinp.msu.ru.

This study is devoted to a preliminary analysis of the segmentation of multivariate time series of the parameters of the solar wind, interplanetary magnetic field, and the Dst index using artificial neural networks, for the period from November 1997 to March 2014. It is shown that this technique allows progress in understanding the dynamics of the Earth's magnetosphere.

Keywords: geomagnetic disturbances, solar wind, segmentation of multivariate time series, clustering, artificial neural network.

PACS: 92.60.hw, 94.30.Xy, 96.60.Vg.

Received 13 November 2015.

English version: *Moscow University Physics Bulletin*. 2016. **71**, No. 4. Pp. 454–463.

Сведения об авторах

1. Доленко Сергей Анатольевич — канд. физ.-мат. наук, зав. лабораторией; тел.: (495) 939-46-19, e-mail: dolenko@srd.sinp.msu.ru.
2. Мягкова Ирина Николаевна — канд. физ.-мат. наук, ст. науч. сотрудник; e-mail: irina@srd.sinp.msu.ru.
3. Персианцев Игорь Георгиевич — доктор физ.-мат. наук, профессор, вед. науч. сотрудник.