

ТИПОЛОГИЗАЦИЯ ОБЪЕКТОВ ЦИРКУМПОЛЯРНОГО ВИХРЯ СЕВЕРНОГО ПОЛУШАРИЯ

Шипко Ю.В., Шувакин Е.В.

ФГВКОУ ВПО Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж), Россия, (394064, Воронеж, ул. Старых Большевиков, 54А), e-mail: argentina_85@mail.ru

На базе моделирования циркумполярных образований Северного полушария на изобарическом уровне 500 гПа для января проведен многомерный статистический анализ, позволивший выявить три кластера объектов. Учитывались следующие характеристики базовой модели циркумполярного вихря: положение центра условных масс; площадь, масса и плотность вихря; ориентация оси эллипса рассеяния масс, главные средние квадратические отклонения и коэффициент сжатия эллипса. Использовались срочные данные реанализа параметров атмосферы NCEP/DOE AMIP-II за период 1979–2013 гг. Построена факторная модель циркумполярного вихря, дана содержательная интерпретация факторов. Объекты выделенных кластеров геометрически проявляются на факторных плоскостях. Дана оценка статистического различия распределений факторов в кластерах. В результате проведенного дискриминантного анализа построены функции, дискриминирующие объекты выделенных кластеров, определены факторы, оказывающие наибольшее влияние на дискриминацию между группами.

Ключевые слова: циркумполярный вихрь Северного полушария, кластеризация, факторная модель, дискриминантный анализ.

TYPOLGIZATION OF CIRCUMPOLAR VORTEX OBJECTS OF THE NORTHERN HEMISPHERE

Shipko Y.V., Shuvakin E.V.

Federal State Official Military Educational Institution of Higher Professional Education Military Educational-Research Centre of Air Force «Air Force Academy named after professor N.E. Zhukovsky and Y.A. Gagarin» (Voronezh) of the Ministry of Defense of the Russian Federation, (396064, Voronezh, street Starich Bolshevikov, 54A), e-mail: argentina_85@mail.ru

On the basis of modeling of circumpolar formations of the Northern hemisphere at the isobaric level 500 hPa for January the multidimensional statistical analysis which allowed to reveal three clusters of objects is carried out. The following characteristics of basic model of circumpolar vortex were considered: position of the center of masses; area, weight and density of vortex; orientation of dispersion ellipse axis, main standard deviations and coefficient of ellipse compression. Data of NCEP/DOE AMIP-II reanalysis of atmospheric parameters of 1979–2013 years period were used. The factorial model of circumpolar vortex is constructed, interpretation of factors is given. Objects of the allocated clusters are shown geometrically on the factorial planes. The estimation of statistical difference of factors distributions in clusters is given. As a result of the carried-out discriminant analysis the functions discriminating objects of the allocated clusters are constructed, the factors having the greatest impact on discrimination between groups are defined.

Keywords: circumpolar vortex of the Northern hemisphere, clusterization, factorial model, discriminant analysis.

В целях обеспечения безопасного и устойчивого развития страны в условиях изменяющегося климата принята Климатическая доктрина Российской Федерации, что указывает на актуальность развития научной основы политики в области климата.

В формировании климатической изменчивости важная роль принадлежит изменчивости циркуляции атмосферы. Глобальная атмосферная циркуляция является одним из механизмов, определяющих все многообразие региональных особенностей климата. Основным звеном атмосферной циркуляции Северного полушария является циркумполярный вихрь (ЦПВ) с центром в районе полюса и западно-восточным движением воздуха. Его характеристики

существенно влияют на формирование и распределение аномальных погодных условий. Поэтому надежное метеорологическое обеспечение оперативной и плановой деятельности потребителей метеорологической информации невозможно без всестороннего изучения внутренней структуры этой компоненты общей циркуляции атмосферы.

Цель работы – совершенствование моделей климатической системы и прогностических схем за счет выявления и учета структурных особенностей ЦПВ Северного полушария.

Для достижения поставленной цели решается задача так называемого разведочного статистического анализа [1], позволяющего выявить вероятностную и геометрическую природу типового метеорологического материала – многомерных данных барической топографии. При этом используются методы факторного, кластерного, дискриминантного анализа.

Исходная информация. В качестве исходного материала использовались данные реанализа параметров атмосферы NCEP/DOEAMIP-II [6] по Северному полушарию. Рассматривались значения геопотенциальной высоты H_{500} (на изобарическом уровне 500 гПа) в узлах регулярной широтно-долготной сетки с шагом $2,5^\circ$ по срокам 00, 06, 12, 18 час. (Всемирного скоординированного времени) за период 1979–2013 гг. По срочным данным геопотенциальной высоты проведен расчет среднесуточных значений. Задача решалась для центрального зимнего месяца (января), когда циклонический характер ЦПВ наиболее выражен.

Используемая модель ЦПВ. При построении математической модели ЦПВ выбран подход [2–4], определяющий аналогию между проекцией поверхности H_{500} на плоскость XOY и двумерным нормальным распределением на этой плоскости случайных величин X, Y с весами p_{ij} для точек (x_i, y_j) . При этом веса проекций элементов объема «чаши», образованной топографией изобарической поверхности горизонтальной плоскостью, проходящей через крайнюю изогипсу H_0 , определяются по формуле [4]:

$$p_{ij} = \frac{(H_0 - H_{ij}) \cos \varphi_i}{\sum_i \sum_j (H_0 - H_{ij}) \cos \varphi_i}, \quad (1)$$

где H_{ij} – значение высоты изобарической поверхности в узле (x_i, y_j) ; φ_i – широта точки (x_i, y_j) ; учитываются только точки (узлы), для которых выполняется условие $H_{ij} < H_0$.

В декартовой системе координат XOY центр совмещен с географическим полюсом, ось X направлена на юг вдоль гринвичского меридиана, ось Y – на восток (90° западной – 90° восточной долготы). В этой системе центр циркуляции (x_c, y_c) определяется как центр масс системы материальных точек.

Характеристиками модели циркумполярного вихря являются: площадь S , ограниченная с юга изогипсой H_0 ; условная «масса» M рассматриваемого объема; условная плотность ЦПВ $z = M/S$; геометрические параметры концентрации двумерного распределения около

центра масс: угол θ между главной осью эллипса рассеяния и осью X ; σ_u , σ_v – полуоси или главные средние квадратические отклонения эквивалентного единичного эллипса рассеяния; коэффициент сжатия эллипса $k = \sigma_v / \sigma_u$ [2–4].

Используемая модель ЦПВ [2–4] программно реализована для поверхности H_{500} . В качестве краевой изогипсы H_0 выбрана характерная для января изогипса, совпадающая с осью планетарной высотной фронтальной зоны, $H_0 = 540$ дам [2].

Кластерный анализ. С целью определения однородных по структуре групп множества циркумполярных образований $\{O_i\}$ ($i = 1, 2, \dots, n$; где n – объем выборки суточных значений характеристик ЦПВ, $n=1085$) с описательными признаками x_{ij} ($j = 1, 2, \dots, p$; к расчету приняты параметры модели вихря: $x_c, u_c, S, M, z, \theta, \sigma_u, \sigma_v, k$), проведена процедура иерархической кластеризации объектов многомерного пространства. При этом не требуется априорной информации о распределении генеральной совокупности [1]. Мерой близости, «похожести» различных объектов O_l, O_m выбрано хеммингово расстояние («городских кварталов») [1]:

$$d_H(O_l, O_m) = \sum_{j=1}^p |x_{lj} - x_{mj}|. \quad (2)$$

Предварительно проведена стандартизация параметров по формуле $x_j^* = (x_{ij} - \bar{x}_j) / \sigma_{x_j}$, чтобы компоненты вектора признаков были однородны по своему физическому смыслу. Для меры (2) влияние отдельных больших разностей (выбросов) уменьшается, так как они не возводятся в квадрат как в евклидовой метрике.

Дерево иерархической кластеризации имеет три отчетливо выделяющиеся ветви. Таким образом, при рассмотрении среднесуточных характеристик циклонических образований можно говорить, в первом приближении, о трех группах (типах) ЦПВ. При этом объекты циркуляции 1-го типа составляют 30,1 % случаев, 2-го – 29,6 %, 3-го – 40,3 %.

В таблице 1 сведены статистические оценки основных характеристик ЦПВ отмеченных кластеров. Из данных таблицы 1 видно, что объекты циркуляции 3-го кластера, в сравнении с объектами кластеров 1, 2, имеют в среднем большую «массу» и «плотность», более сжатый эллипс рассеяния масс. Выявлена особенность распределений угла ориентации эллипса рассеяния по кластерам: для кластера 1 (в основном) $\theta > 0$; для кластера 2 – $\theta < 0$ (т. е. объекты циркуляции кластера 1 имеют направление главной оси эллипса рассеяния восточной ориентации, а кластера 2 – западной); для кластера 3 угол θ имеет бимодальное распределение.

Таблица 1

Статистические оценки параметров ЦПВ по кластерам (500 гПа, январь)

Параметр ЦПВ	Кластер 1		Кластер 2		Кластер 3	
	средняя	ср.кв. откл.	средняя	ср.кв. откл.	средняя	ср.кв. откл.
x_c , км	-420,02	394,43	-552,63	330,64	-503,71	284,68
y_c , км	-91,15	328,29	-59,09	326,82	27,47	264,89
$S \cdot 10^{-6}$, км ²	63,39	3,83	65,34	3,32	65,57	2,92
$M \cdot 10^{-9}$, км ² дам	1,37	0,16	1,38	0,17	1,56	0,15
z , дам	21,72	2,54	21,19	2,71	23,86	2,42
θ , град.	32,34	14,34	-31,98	17,75	2,80	31,80
σ_u , км	2453,85	147,77	2411,18	127,15	2217,08	119,02
σ_v , км	1617,53	138,59	1688,51	143,82	1788,39	129,93
k	0,66	0,07	0,70	0,07	0,81	0,07

Факторная модель ЦПВ. С целью сжатия информации и определения структуры данных проведен факторный анализ, позволяющий обнаружить скрытые факторы, объясняющие связи между наблюдаемыми признаками объекта. В факторном анализе применяется модель описания переменных $X_j (j = 1, 2, \dots, p)$ линейной комбинацией r ненаблюдаемых факторов f_1, f_2, \dots, f_r , при этом $r < p$:

$$X_j^* = a_{j1}f_1 + a_{j2}f_2 + \dots + a_{jr}f_r + e_j, \quad (3)$$

где X_j^* – центрированные и нормализованные исходные данные; $f_i, i = 1, 2, \dots, r$, называются общими факторами, так как они являются общими для представления всех признаков; e_j – остаток или специфичный фактор для переменной X_j ; a_{ji} – неизвестные коэффициенты, имеют специальное название – факторные нагрузки (задача факторного анализа состоит в том, чтобы оценить их некоторым оптимальным образом).

В результате факторного анализа определены общие факторы модели ЦПВ на уровне 500 гПа, объясняющие 89,5 % общей дисперсии признаков. Содержательная интерпретация факторов дана в соответствии с факторными нагрузками (после «варимакс-вращения») на параметры ЦПВ, превышающими по модулю 0,7 [1].

Первый фактор наиболее тесно связан с предикторами σ_u, σ_v, k , что дает основание интерпретировать f_1 как фактор формы рассеяния условных масс ЦПВ (объясняет 28,9 % общей дисперсии признаков).

Фактор f_2 , судя по нагрузкам на параметры M, z , можно назвать фактором массы ЦПВ (объясняет 24,0 % общей дисперсии).

Соответственно другим факторам даны следующие интерпретации:

- f_3 – фактор положения центра вихря (15,7 % общей дисперсии);
- f_4 – фактор ориентации эллипса рассеяния (11,3 % общей дисперсии);
- f_5 – фактор площади вихря (9,6 % общей дисперсии).

На рисунках 1–3 объекты выделенных кластеров представлены в разных проекциях: на плоскости факторов f_1, f_2 (рисунки 1, 2) и f_1, f_4 (рисунок 3).

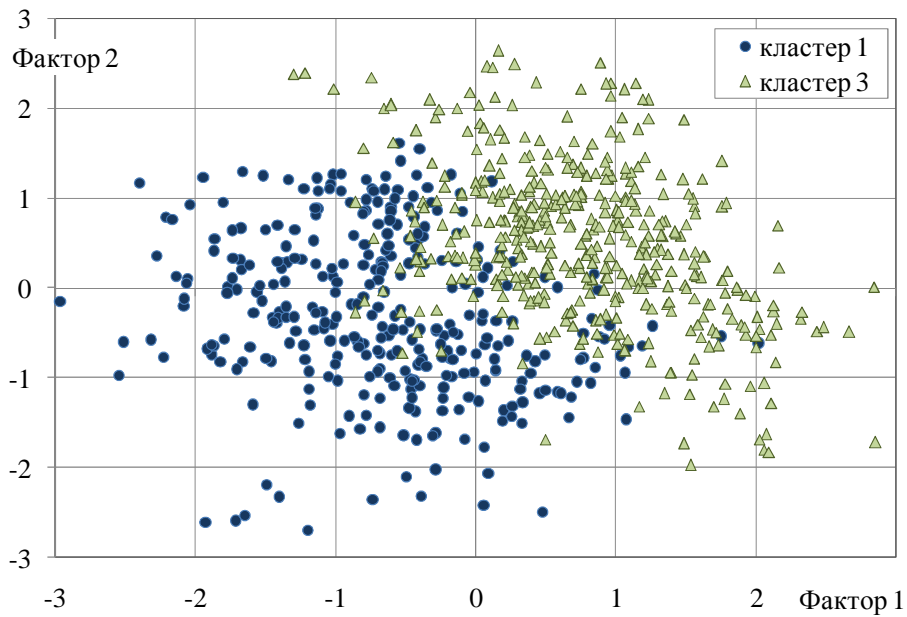


Рис.1. Распределение объектов ЦПВ на плоскости факторов f_1, f_2 с маркировкой групп, полученных кластеризацией (без объектов кластера 2)

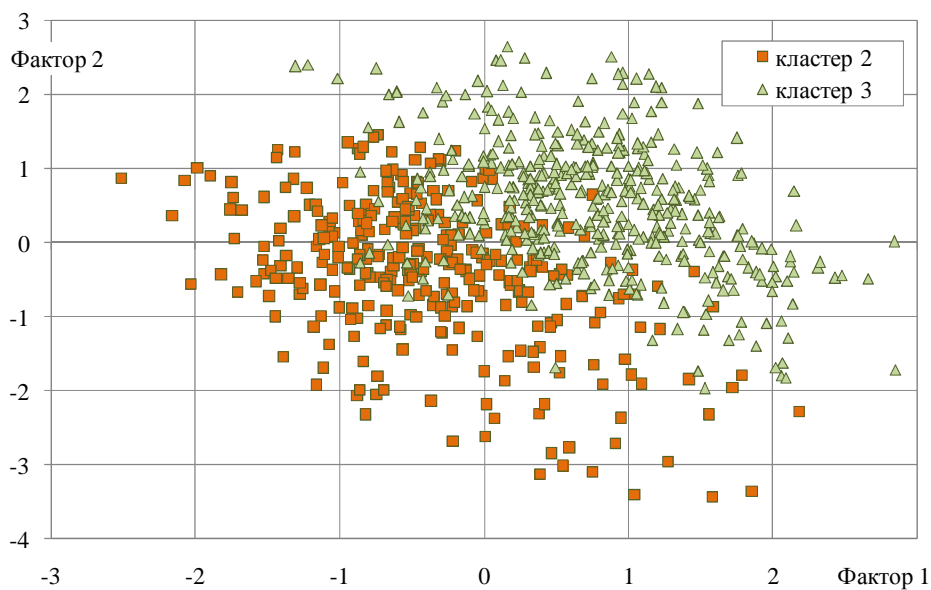


Рис.2. Распределение объектов ЦПВ на плоскости факторов f_1, f_2 с маркировкой групп, полученных кластеризацией (без объектов кластера 1)

Как показано на рисунках, «сгустки» объектов ЦПВ полученных кластеров выделяются визуально (чтобы не затенять картину отличия скоплений объектов определенных групп, на каждом рисунке не показаны объекты одного из кластеров).

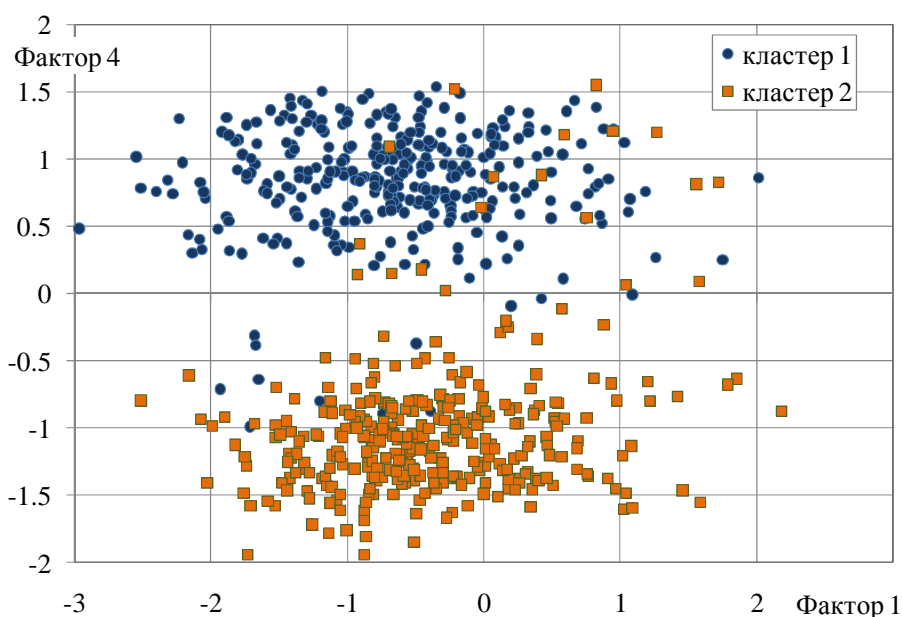


Рис.3. Распределение объектов ЦПВ на плоскости факторов f_1, f_4

с маркировкой групп, полученных кластеризацией (без объектов кластера 3)

Оценка однородности кластеров. В целях определения однородности групп, проведена проверка гипотезы о принадлежности факторов рассматриваемых кластеров одним и тем же распределениям. Использовались непараметрические методы (или свободные от распределения): на основе критериев Колмогорова – Смирнова и рангового Манна – Уитни[6], – не требующие предпосылки существования закона распределения.

В таблице 2 представлены результаты оценки по совокупности двух указанных критериев статистического различия распределений соответствующих факторов в рассматриваемых группах на уровне значимости $\alpha = 0,05$ (знаком «+» отмечены случаи принятия статистической гипотезы, что две совокупности имеют одно и то же распределение; знаком «-» – принятия альтернативной гипотезы).

Таблица 2

Оценки статистического различия распределений факторов в кластерах ($\alpha= 0,05$)

Кластеры	f_1	f_2	f_3	f_4	f_5
1, 2	-	+	+	-	-
1, 3	-	-	-	-	-
2, 3	-	-	+	-	+

Как следует из данных таблицы 2, существенно различаются распределения всех факторов в кластерах 1, 3; в кластерах 1, 2 различаются только распределения факторов f_1, f_4, f_5 ; в кластерах 2 и 3 – f_1, f_2, f_4 .

На рисунке 4 даны в сравнении статистические характеристики факторов по группам (мода, 50-ти процентный интервал и размах варьирования), графическое представление которых визуально подтверждает данное утверждение.

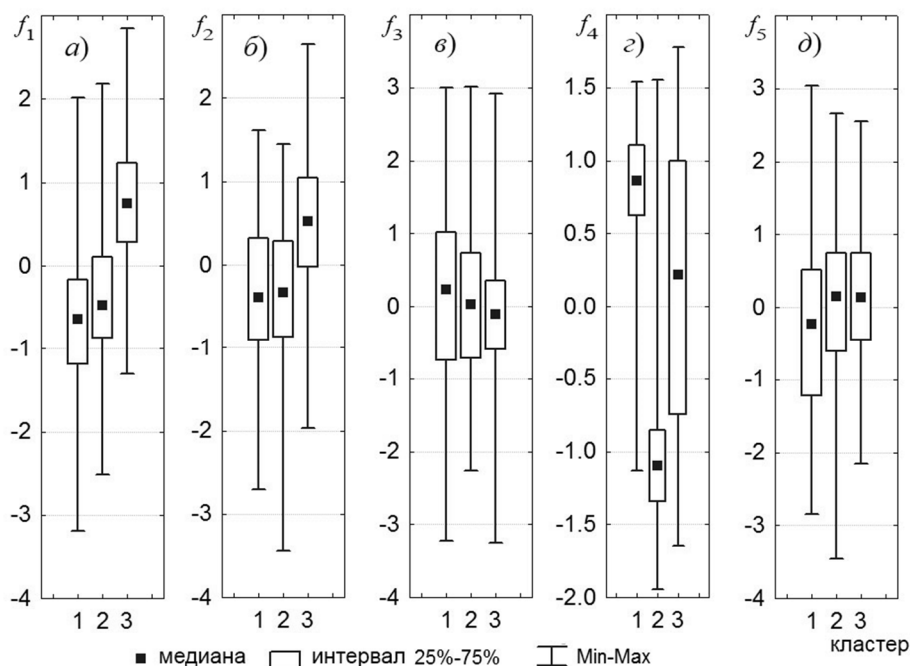


Рис. 4. Сравнение для выделенных кластеров статистических характеристик факторов: а) f_1 ; б) f_2 ; в) f_3 ; г) f_4 ; д) f_5

Дискриминантный анализ. Для выяснения, какие факторы вносят свой вклад в дискриминацию между полученными совокупностями объектов ЦПВ, в задаче использовался дискриминантный анализ. В результате дискриминантного анализа, во-первых, построены классификационные функции, предназначенные для определения того, к какой группе наиболее вероятно может быть отнесен отдельный объект. Во-вторых, определены дискриминантные функции, разделяющие возникающие совокупности.

Каждой группе (совокупности объектов) соответствует своя классификационная функция d_{ci} ($i = 1, 2, 3$), определяемая вектором коэффициентов и константой:

$$d_{c1} = -2,734 - 1,663f_1 - 0,808f_2 + 0,396f_3 + 1,779f_4 - 0,692f_5, \quad (4)$$

$$d_{c2} = -2,726 - 0,999f_1 - 0,986f_2 + 0,056f_3 - 2,212f_4 + 0,178f_5, \quad (5)$$

$$d_{c3} = -2,079 + 1,979f_1 + 1,329f_2 - 0,337f_3 + 0,294f_4 + 0,387f_5, \quad (6)$$

Данные функции классификации (14)–(16) можно использовать для прямого вычисления показателя классификации для некоторых новых значений. Наблюдение (объект циркуляции) считается принадлежащим той совокупности, для которой получен наивысший показатель классификации.

В результате проведенного анализа определены дискриминантные функции d_1 и d_2 . Функция d_1 в основном дискриминирует между группой 3 и объединением групп 1 и 2. Функция d_2 дискриминирует между группой 1 и 2. По коэффициентам полученных дискриминантных функций, представленных в стандартизованном виде:

$$d_1 = -1,004f_1 - 0,781f_2 + 0,226f_3 - 0,070f_4 - 0,267f_5, \quad (7)$$

$$d_2 = -0,095f_1 + 0,125f_2 + 0,101f_3 + 1,001f_4 - 0,281f_5, \quad (8)$$

можно судить о степени влияния каждого предиктора на дискриминацию групп. Так, согласно (7), (8), наибольшее влияние на дискриминацию между группой 3 и совокупностью групп 1, 2 оказывают факторы f_1 и f_2 (формы и массы ЦПВ); на дискриминацию между группами 1 и 2 наибольшее влияние оказывает фактор f_4 (ориентации эллипса рассеяния условных масс ЦПВ), что подтверждает и распределение объектов на рисунках 1–3.

Другой целью применения дискриминантного анализа является проведение классификации. Если модель установлена и получены дискриминирующие функции, возникает вопрос о том, как хорошо они могут предсказывать, к какой совокупности принадлежит конкретный объект циркуляции. Такой анализ не проводился, поскольку это цель следующего этапа – модельного дискриминантного анализа. Цель разведочного анализа – предварительная оценка внутренней структуры исходной совокупности.

Достоверность полученных результатов подтверждается большим объемом данных, привлеченных для исследования, и применением классических методов статистической обработки информации.

Полученные результаты позволяют говорить о дифференцированном использовании данных о состоянии ЦПВ в разрабатываемых погодно-климатических моделях, тем самым повысить эффективность метеорологического обеспечения.

Список литературы

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. Справочное издание, [под ред. С.А. Айвазяна]. – М.: Финансы и статистика, 1989. – 607 с.
2. Багров Н. А., Орлова И. И. К вопросу определения центра циркуляции атмосферы // Труды Гидрометцентра СССР. – Вып. 211. – Л.: Гидрометеиздат, 1978. – С. 3–14.
3. Багров Н.А. О центре циркуляции // Метеорология и гидрология. – 1975. – № 2. – С. 3–11.
4. Глызь Г. А. О некоторых характеристиках циркумполярного вихря // Труды ВНИИГМИ-МЦД. – Вып. 58. – М.: Гидрометеиздат, 1979. – С. 98–104.
5. Справочник по прикладной статистике. В 2 т. Т. 2: [под ред. Э. Ллойда, У. Ледермана; пер. с англ. под ред. С.А. Айвазяна, Ю.Н. Тюрина]. – М.: Финансы и статистика, 1990. – 526 с.
6. Kanamitsu M. et al. NCEP/DOEAMIP-II Reanalysis // Bull. American Meteor. Soc., 2002. – Vol. 83. – P. 1631–1643.

Рецензенты:

Куролап С.А., д.г.н., профессор, заведующий кафедрой геоэкологии и мониторинга окружающей среды, ФГБОУ ВПО «Воронежский государственный университет», г. Воронеж;

Кузнецов И.Е., д.т.н., начальник 13 кафедры гидрометеорологического обеспечения, ФГБОУ ВПО «Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина» Министерства обороны Российской Федерации, г. Воронеж.