

УДК 551.513.2:551.509.314

**НЕКОТОРЫЕ ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ  
ПАРАМЕТРИЗОВАННЫХ ФОРМ КРУПНОМАСШТАБНОЙ  
АТМОСФЕРНОЙ ЦИРКУЛЯЦИИ В СИСТЕМЕ УПРАВЛЕНИЯ  
КАЧЕСТВОМ АТМОСФЕРНОГО ВОЗДУХА КАЗАХСТАНА**

Канд. геогр. наук      В.Г. Сальников

*Изучены различные подходы к параметризации форм крупномасштабной атмосферной циркуляции и возможность применения в системе управления качеством атмосферного воздуха Казахстана. Показана возможность применения алгоритма Головкина. Выделенные типы процессов являются определяющими в формировании различных уровней загрязнения в республике. Показана перспективность такого подхода в процессе разработки комплекса воздухоохраных мероприятий.*

Одним из самых перспективных подходов к задаче выделения множества погодных режимов является классификация долгоживущих структур циркуляции с точки зрения стационарных решений в упрощённых моделях, описывающих отклик планетарных волн на наблюдаемое термическое и орографическое воздействие. Современные гидродинамические модели общей циркуляции атмосферы такого рода, вероятно, в значительной степени идеализированы и вряд ли могут стать надёжной основой детального анализа и окончательного вывода о существовании множества погодных режимов в атмосфере.

В отсутствие развитого теоретического руководства вполне оправдан более субъективный подход для выделения типов циркуляции, связанных с преимущественными погодными режимами. Необходимо отметить, что качество идентификации режимов атмосферной циркуляции в значительной, если не в решающей, степени определяют качество гидродинамических прогнозов и предел предсказуемости мгновенного состояния атмосферы. Под режимом, как правило, понимают крупномасштабную барическую структуру, которая характеризуется определённой формой циркуляции (гребень, ложбина, циклон, антициклон) и продолжительностью существования, или устойчивостью. Решение проблемы долгосрочного предсказания динамики атмосферных процессов и, следовательно,

особенностей диффузии примесей в воздушной среде во многом зависит от возможности выделять и проследивать эволюцию этих крупномасштабных возмущений атмосферной циркуляции (режимов).

Для установления принадлежности некоторого объекта (предмета, процесса, явления, ситуации) к одному из классов или образов в последние годы достаточно активно используются методы распознавания образов [2...5, 7]. Объектами классификации могут служить синоптические ситуации, поля температуры и влажности воздуха, осадков, атмосферного давления, геопотенциала, атмосферные явления, потенциал загрязнения атмосферы и т.п.

Необходимо отметить, что, согласно современным представлениям существует два основных направления исследований, идентифицирующих режимы атмосферной циркуляции с помощью распознавания образов. Первое – это статистические процедуры, основанные на оценке плотностей распределения вероятностей. Результаты таких исследований представлены в работах [7, 9]. Так, в работе [9] выдвигается гипотеза, в соответствии с которой локальные максимумы плотности распределения вероятностей, существующие в фазовом пространстве, связаны с определёнными режимами атмосферной циркуляции. Для каждой из точек наблюдений  $Z_i$  рассчитывается евклидово расстояние и формируется ансамбль точек, для которых это расстояние меньше некоторого порогового значения. Плотность определяется как число точек  $Z_j$  (включая  $Z_i$ ), принадлежащих окрестности радиуса  $r$ . Выделяются шесть режимов (классов) атмосферной циркуляции, идентифицируемых в порядке убывания локальной плотности распределения вероятностей.

Подобные результаты получены в работе [12], где предварительно проведена фильтрация исходных данных с целью выделения стационарных и нестационарных составляющих. В качестве расстояния между объектами используется среднее квадратическое отклонение. Делается вывод о глобальном характере режимов атмосферной циркуляции.

Недостатками этого подхода являются громоздкость необходимых вычислительных процедур и субъективизм, связанный с выбором порогового расстояния  $r$ . Кроме того, остаётся открытым вопрос, действительно ли локальные максимумы плотности распределения вероятности связаны с определёнными режимами атмосферной циркуляции.

В основе второго направления лежат эвристические алгоритмы, основанные на интуитивных соображениях выделения групп точек в ис-

ходном пространстве. В работе [10] классификация исходной совокупности объектов производится в соответствии с выбранными эталонами. Входные параметры (пороговое расстояние и число классов) задаются в зависимости от поставленной задачи.

Основные трудности и элементы субъективизма, связанные с использованием эвристических алгоритмов, возникают при выборе расстояния между объектами, в том числе и порогового, а также оператора существенных свойств признаков, определяющих результаты классификации. Для улучшения качества классификации и получения более достоверных оценок расстояния между объектами кроме традиционных, среднего квадратического отклонения и коэффициента корреляции, могут использоваться и дополнительные параметры, к примеру, различные функции зонального и меридионального градиентов, вычисленных для каждой точки [8].

В ряде исследований [8, 11] при решении задачи автоматической классификации предварительно используется метод главных компонент с целью ортогонализации исходных данных и выделения крупномасштабных составляющих. В работе [11] за меру сходства объектов принимается нормализованная сумма квадратов разностей от 2-го до 100-го временного коэффициента разложения двухдневных средних полей  $H_{500}$  во внетропических широтах северного полушария. Сделан вывод, что аналогичные ситуации для всего северного полушария – явление весьма редкое, они составляют 0,2 % всех рассмотренных случаев. Это обуславливает необходимость регионального подхода к анализу режимов атмосферной циркуляции.

В целом, анализ проведённых исследований показывает, что уровень результатов по идентификации режимов атмосферной циркуляции с помощью метода распознавания образов во многом определяется оптимальностью набора решающих правил, характерных признаков объекта и критериев качества классификации.

В данной работе представлены результаты классификации осреднённых полей геопотенциала, учитывающих основные особенности развития синоптических процессов на поверхности  $H_{500}$ , которые во многом определяют особенности погодных условий у поверхности Земли.

Одним из путей преодоления трудностей и элементов субъективизма, возникающих при использовании эвристических алгоритмов, выборе расстояния между объектами, в том числе и порогового, представляется классификация на основе алгоритма Б.А. Головкина [6], реализованная

для всей исходной совокупности объектов без использования порогового расстояния и процедуры ранжирования.

Теоретической основой автоматической классификации служит алгоритм объективного распознавания образов (алгоритм Головкина) [6]. В соответствии с этим методом в матрице мер близостей  $\{R'\}$  размером  $k' \times N$ , где  $N$  – число объектов, в каждом  $s$ -м столбце определяется минимальный элемент  $R'_{st}$  ( $s = 1, \dots, k; t = 1, \dots, N$ ). Объект  $Y_i$ , номер которого совпадает с номером столбца, считается принадлежащим классу (эталону), номер которого совпадает с  $s$ -м номером строки. Такая процедура выполняется при  $k = N, N - 1, \dots, 1$ .

До проведения классификации выбирается набор признаков, которые описывают наиболее важные, с точки зрения проводимого исследования, черты классифицируемых объектов. В зависимости от конкретных особенностей классифицируемого множества, а также цели, для которой проводится классификация устанавливается мера близости объектов. Последняя представляет собой числовую функцию, заданную на парах наборов признаков. Говоря о близости или удаленности объектов друг от друга, имеют в виду значение меры близости для данной пары объектов: чем оно меньше, тем «ближе» расположены последние в пространстве признаков. Значения меры близости для каждой пары объектов поступают на вход алгоритма классификации. В процессе реализации алгоритма из всех возможных разбиений классифицируемого множества выбирается оптимальное, которое представляет собой искомую классификацию. Оптимальность здесь понижается в заранее установленном смысле, что позволяет говорить об объективности получаемого результата.

Пусть классифицируемое множество состоит из  $N$  объектов, причем они пронумерованы. Поскольку число классов не задано заранее, в процессе работы алгоритма последовательно строятся наилучшие деления множества на  $K, N - 2, \dots, 2$  класса. Для этого при каждом фиксированном числе  $K$  ищутся такие  $K$ -объектов, для которых значения показателя достигает минимума на множестве всех различных  $K$  – элементных объектов.

$$F(K) = \sum_{i \in P} \sum_{j \in Q_i} S_{ij}. \quad (1)$$

Здесь  $P$  – множество номеров  $K$ -объектов,  $Q_i$  – множество номеров объектов, для которых наиболее близким с номером из  $P$  является  $i$ -ый объект;  $S_{ij}$  – значение меры близости для пары объектов с номерами  $i$  и  $j$ .

Каждый из  $K$ -объектов, найденных согласно минимуму показателя  $F(K)$ , является наиболее типичным представителем некоторого класса и называется его эталоном. Сам класс формируется возле эталона путем объединения с последним тех объектов, для которых данный эталон является самым близким из всех  $K$ -эталонов. Поиск наилучших разбиений осуществляется по схеме пошагового исключения наименее предпочтительного объекта из текущего набора объектов, с помощью которого получаем приближённое решение.

Таким образом, имеем  $N - 2$  наилучших разбиений. Каждое из них характеризуется значением показателя качества разбиения  $C_{max}$ , позволяющего сравнить разбиение классифицируемого множества на различное число классов. Показатель  $C_{max}$  учитывает кучность группирования объектов в классах, а также изолированность классов друг от друга. Сопоставление значений  $C_{max}$  для различных  $K$  позволяет выбрать оптимальное число классов. Соответствующее этому числу наилучшее разбиение и является решением задачи.

В качестве исходной информации использовались ежедневные ряды поля  $H_{500}$  в 96 точках первого естественного синоптического района за период с 1989 по 1991 годы. Первая задача состояла в оптимальном описании пространственного распределения поля аномалий  $H_{500}$  с целью сокращения размерности вектора  $X$  и получения более компактных и легче делимых областей. Для этого используются лингвистические и структурные методы. Предлагаемый язык имеет простейшую грамматику: описание изображения в ней есть перечисление того, какие отличительные фрагменты в каком «месте» изображения расположены. Характерными фрагментами распределения аномалии поля  $H_{500}$  будем считать:

- степень аномальности этих полей, определяемую с помощью показателя  $K$ , введённого Н.А. Багровым (2) [12]

$$K_j = \frac{1}{m} \sum_{i=1}^m \left( \frac{\Delta H_{ij}}{\sigma_i} \right)^2, \quad (2)$$

где  $\Delta H_{ij}$  – аномалия  $H_{500}$  в точке  $i$ ;  $\sigma_i$  – её среднее квадратическое отклонение;  $m$  – общее число точек.

- знак поля, характеризуемый отношением (3):

$$P_j = \frac{m_j}{m}, \quad (3)$$

где  $m_j$  – число точек с положительными значениями аномалии  $H_{500}$  для данного поля.

- внутреннюю структуру изображения, представленную средними квадратическими значениями меридионального и зонального барических градиентов, вычисленных для каждой точки и осреднённых в пределах заданной территории:

$$PR_j = \left[ \frac{1}{N_2(N_1-1)} \sum_{j=1}^{N_2} \sum_{i=1}^{N_1-1} \left( \frac{\Delta H_{ij} - \Delta H_{ij+1}}{l_{ii+1}} \right)^2 \right]^{1/2} + \left[ \frac{1}{N_1(N_2-1)} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2-1} \left( \frac{\Delta H_{ij} - \Delta H_{ij+1}}{l'_{jj+1}} \right)^2 \right]^{1/2}, \quad (4)$$

где  $l_{ii+1}$  и  $l'_{jj+1}$  – расстояния между точками;  $N_1$  и  $N_2$  – число точек, взятых соответственно вдоль круга широты и долготы.

Вся территория разбивалась на четыре «квадрата», для каждого из которых рассчитывались параметры  $K_j$ ,  $P_j$  и  $PR_j$  ( $j = 1, \dots, N$ ) и соответствующие средние по выборке  $\bar{K}$ ,  $\bar{P}$  и  $\bar{PR}$ . В зависимости от наличия или отсутствия в квадрате значений параметров  $K_j > \bar{K}$ ,  $P_j > \bar{P}$  и  $PR_j > \bar{PR}$  изображению присваивалось значение 1 или  $-1$ , т.е. было выполнено следующее преобразование вектора  $X$ :

$$Y = (Y_1, \dots, Y_{12}) = \text{sign}(X_1, \dots, X_{120}). \quad (5)$$

В результате размерность исходного пространства признаков уменьшилась в 10 раз.

Отождествляя значения признаков  $Y$  со значениями координат пространства Хемминга, в качестве меры сходства между объектами будем использовать расстояние  $Y$ .

$$R(Y_i, Y_j) = \frac{1}{4} \sum_{i=1}^{12} (Y_{in} - Y_{jn})^2, \quad (6)$$

где  $i, j = 1, \dots, N$ . Параметр  $R$  удовлетворяет аксиомам метрики (тождественности, симметричности и треугольника).

Вторая задача заключается в распознании полученных образов распределения  $H_{500}$ , которая реализуется с помощью рассмотренного выше алгоритма Головкина.

На третьем этапе решается вопрос оценки качества полученных  $N$ -классификаций. Функционал, который необходимо экстремизировать, конструируется следующим образом [4]. Для каждого объекта  $Y_i \in A_i$  определяются два числа: мера «притяжения» объекта  $Y_i$  к объектам той же группы  $A_S$ :

$$M_i = \frac{1}{N_S} \sum_{Y_t \in A_S} (R_{st} - R^*), \quad (7)$$

где  $R^*$  – заданный порог ( $R = 7$ );  $N_S$  – число объектов в группе  $A_S$  и мера «притяжения» объекта ко всем остальным объектам:

$$\bar{M}_i = \frac{1}{N - N_S} \sum_{Y_i \in A_S} (R_{st} - R^*). \quad (8)$$

С использованием величин  $M_i$  и  $\bar{M}_i$  вводится понятие «стабильности» объекта:

$$Q_i = \frac{M_i - R^*}{1 - R^*} + \frac{R^* - \bar{M}_i}{R^*} \quad (9)$$

и стабильность всего разбиения

$$Q = \frac{1}{N} \sum_{i=1}^N Q_i. \quad (10)$$

Величина  $Q$  рассчитывается для значений  $k' = N, N - 1, \dots, 1$ . Максимальное значение функции  $Q(k)$  отмечается при  $k' = 4$ . Поэтому в качестве окончательного принимаем разбиение данной совокупности объектов на 4 класса.

Для практического распознавания синоптических воздействий была отработана программа для персонального компьютера (ПК), в которой используются ежесуточные данные геопотенциальной высоты поверхности 500 гПа в узлах сеточной области с шагом  $5^\circ \times 10^\circ$  и координатами  $35 \dots 70^\circ$  с.ш.,  $-20^\circ$  з.д.  $\dots 70^\circ$  в.д.

В результате проведённых исследований на выборке обучения (набор наиболее характерных примеров разновидностей развития синоптических ситуаций) были отобраны случаи с аналогичным типом развития синоптической ситуации в казахстанском регионе. Проведено временное осреднение этих полей для 96 узлов ( $8 \times 12$ ) используемой сеточной области. На основе чего выделено 4 основных типа полей  $H_{500}$ , различающихся особенностями конфигурации и географического положения высотной фронтальной зоны, а, следовательно, и особенностями синоптических процессов у поверхности Земли. Для каждого из выделенных типов построены средние композиционные поля  $H_{500}$  и аномалий поля  $H_{500}$ , выполнен анализ соответствующих синоптических процессов. При всём разнообразии проанализированных процессов выделяются ситуации, связанные с режимами зональных (тип 1) и меридиональных потоков (типы 2 и 3), а также их блокированием (тип 4). На рисунке 1 и 2 в качестве примера приводятся композиционные средние поля  $H_{500}$  и аномалий  $H_{500}$  для типа 4.

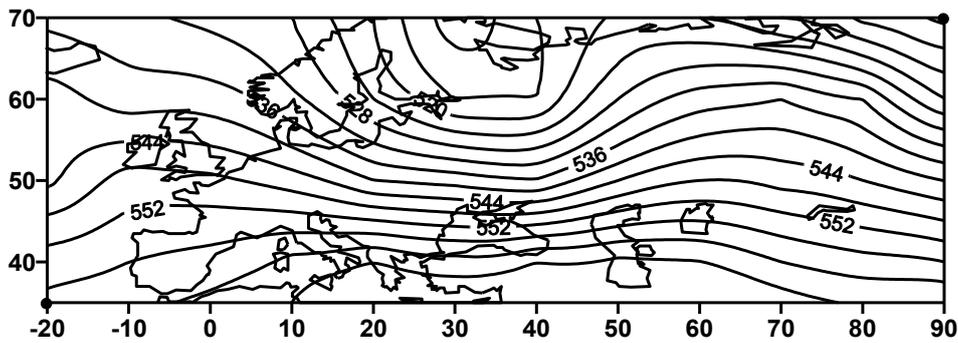


Рисунок 1 – Композиционное поле  $H_{500}$  для типичного блокирующего процесса.

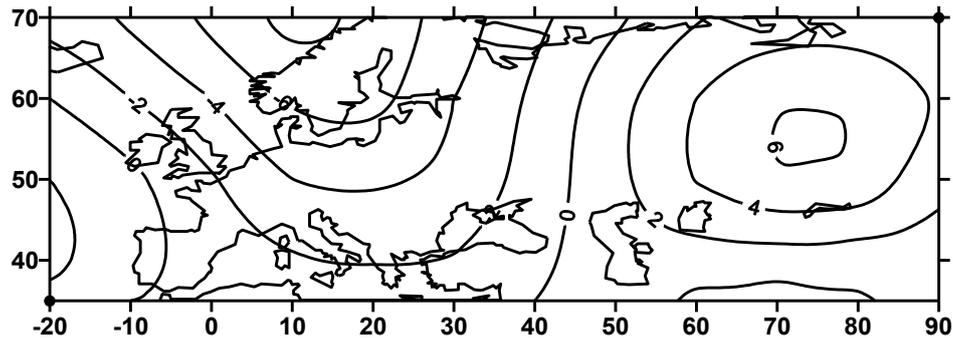


Рисунок 2 – Композиционное поле аномалий  $H_{500}$  для типичного блокирующего процесса.

Обширный блокирующий гребень располагается над территорией Урала и Западной Сибири. Область с положительными аномалиями  $H_{500}$  также указывает на более интенсивное, по сравнению с нормой, развитие высотного гребня над рассматриваемыми районами. К западу от гребня располагается ложбина, ось которой проходит по ЕТР с севера на юг. Такое поле приводит к тому, что над Казахстаном преобладают мало подвижные антициклоны, обуславливающие характерный режим погоды (частые инверсии, слабый ветер, отсутствие осадков, а при соответствующих условиях и туманы). С точки зрения величины потенциала загрязнения атмосферы этот тип является наиболее опасным, так как подобные погодные условия способствуют наиболее интенсивному накоплению примесей в воздушной среде.

На рисунке 3 и 4 приводятся композиционные средние поля типичного зонального процесса.

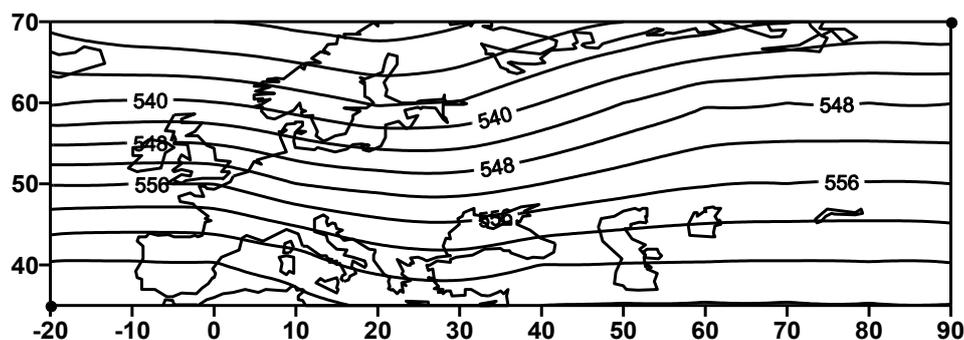


Рисунок 3 – Композиционное поле  $H_{500}$  для типичного зонального процесса.

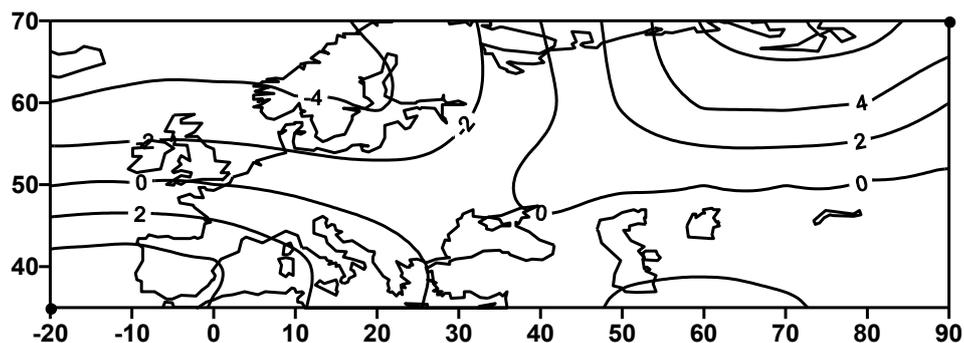


Рисунок 4 – Композиционное поле аномалий  $H_{500}$  для типичного зонального процесса.

Отличия 2 и 3 типа полей  $H_{500}$  определяются особенностями взаимного расположения гребней и ложбин и географическим положением ПВФЗ. Это оказывает определённое влияние на характер потенциала загрязнения атмосферы над теми или иными регионами.

Наибольшая повторяемость отмечается для типов 1 и 4 и составляет соответственно 31 и 36 % всех рассмотренных случаев. Средняя продолжительность существования колеблется от 3 до 6 суток, что соответствует естественным синоптическим периодам, установленным в процессе изучения крупномасштабных атмосферных процессов. Типы 2 и 3, как правило, менее устойчивые – длительность их существования не превышает 3 дней.

Таким образом, результаты выполненного исследования показывают, что в процессе изучения потенциала загрязнения атмосферы основной информационной базой служит синоптическое отображение действия физических процессов в атмосфере. Это давно подтверждено самой практикой изучения общециркуляционных процессов воздушной среды.

В этих условиях становится целесообразным проведение классификации поведения синоптических объектов, как способа повышения объективности проводимых работ. Во-первых, это методически синхронизируется с последовательностью действий в процессе интерпретации результатов исследований (которые всегда начинаются с общей оценки синоптической ситуации) и позволяет формализовать знания о динамике атмосферных процессов в конкретном регионе. Во-вторых, движение от общей объективной оценки синоптической ситуации к локальному загрязнению атмосферы позволяет в качестве математического аппарата успешно применять синоптико-статистический инструментарий, широко применяемый в прикладных исследованиях. В частности, данный подход предоставляет возможность перейти от формального субъективного описания синоптической ситуации (как правило, предшествующей количественному анализу изучаемого явления) к ее объективной типизации и уже на этой основе к формированию однородных, статистически более устойчивых выборок исходных данных. В третьих, объективная классификация синоптических ситуаций, их каталогизация позволяет создать систему накопления знаний об их региональных особенностях в форме, удобной для использования.

Полученные на первом этапе исследований результаты, а именно, отработанные приемы распознавания синоптических ситуаций применительно к казахстанскому региону и их типизация, позволяют решать перечисленные выше задачи.

#### СПИСОК ЛИТЕРАТУРЫ

1. Багров Н.А., Мякишева Н.Н. Некоторые характеристики аномалий средних месячных температур воздуха // Труды ММЦ. – 1966. – Вып.9. – С. 3-17.
2. Бурцев А.А. Вероятностный подход к задаче объективной классификации // Метеорология и гидрология. – 1980. – №3, С. 39 - 44.
3. Бурцев А.А. Эксперименты по объективной классификации метеорологических полей // Труды Гидрометцентра СССР. – 1980. – Вып. 226. – С. 11-22.
4. Вильфанд Р.М. Применение объективной типизации метеорологических полей к месячному прогнозу аномалии температуры воздуха на Европейской территории СССР // Метеорология и гидрология. – 1977. – №10. – С. 28-36.
5. Вильфанд Р.М. Применение объективной типизации полей Н - 500 к прогнозу месячного количества осадков // Труды Гидрометцентра СССР. – 1977. – Вып. 188. – С. 3-11.

6. Головкин Б.А. Машинное распознавание и линейное программирование. – М.: Советское радио, 1973. – 176 с.
7. Груза Г.В., Раньков Е.Я. О принципах автоматической классификации метеорологических объектов // Метеорология и гидрология. – 1972. – №2. – С. 12-22.
8. Davis R.E. and Kalkstein L.S. Development of an automated spatial synoptic climatological classification // Int. J. Climatol. – 1990. – Vol.10, №8, P. 769-794.
9. Molteni F., Tibaldi S. and Palmer T.N. Regimes in the wintertime circulation over northern extratropics. P. I.: Observational evidence // Quart. J. Roy. Met. Soc. – 1990. – Vol. 116. – No 491. – P. 31-67.
10. Rubin J. Optimal classification into groups: an approach for solving the taxonomy problem // J. Theor. Biol. – 1967. – Vol. 15, №1.
11. Ruosteemoja K. Factors affecting the occurrence and lifetime of 500 mb height analogues: A study based on a large amount of data // Mon. Weath Rev. – 1988. – Vol. 116. – №2. – P. 368-376.
12. Toth Z. Preferred and unpreferred circulation patterns in the Northern Hemisphere extratropical winter. – Extended abstracts submitted to the ICTP/WMO International technical conference on long-range weather forecasting research. Trieste, Italy, 8-12 April 1991 // WMO/TD. – №395. – 1991. – P. 277-280.

Казахский национальный университет им. аль-Фараби

**ҚАЗАҚСТАННЫҢ АТМОСФЕРАЛЫҚ АУА САПАСЫН БАСҚАРУ  
ЖҮЙЕСІНДЕ ПАРАМЕТРЛЕНГЕН ІРІ МАСШТАБЫ АТМОСФЕРАЛЫҚ  
АЙНАЛЫМНЫҢ КЕЙБІР ҚОЛДАНЫЛУ МҮМКІНДІКТЕРІ**

Геогр. ғылымд.                      В.Г. Сальников  
канд.

*Қазақстанның атмосфералық ауа сапасын басқару жүйесінде қолдану мүмкіндіктері және ірі масштабтағы айналымның параметрленген формасы әртүрлі бағытта зерттелген. Головкин алгоритмін қолдану мүмкіндігі көрсетілген. Республикадағы ластану деңгейінің қалыптасуын процестің бөлінген типтері анықтайды. Мұндай жинақты жасау жүйесінде анықтаудың жүйелілігі ауаны қорғау шарасында көрсетілген.*