

УДК [551.509.313+551.509.32]:551.589

*А.Н. Ефременко**, *Н.О. Моисеева**, *Е.А. Румянцева***, *В.В. Черный**

**РАЗРАБОТКА АВТОМАТИЗИРОВАННОГО МЕТОДА
КЛАССИФИКАЦИИ МАКРОСИНОПТИЧЕСКИХ ПРОЦЕССОВ
В ИНТЕРЕСАХ СРЕДНЕСРОЧНОГО ПРОГНОЗИРОВАНИЯ
ПОЛЕЙ МЕТЕОРОЛОГИЧЕСКИХ ВЕЛИЧИН**

* Военно-космическая академия имени А.Ф. Можайского; ** Российский государственный гидрометеорологический университет, rumkate@rambler.ru

A.N. Efremenko, N.O. Moiseeva, E.A. Rumiantceva, V.V. Chernyy

**DEVISING OF THE AUTOMATED METHOD
OF THE MACRO-SYNOPTICAL PROCESSES CLASSIFICATION
IN THE INTERESTS OF THE MEDIUM-TERM WEATHER
FORECASTING**

Статья посвящена исследованию методов среднесрочного прогнозирования полей метеорологических величин. Особое внимание уделяется проблеме метода аналогов, заключающейся в формировании перечня типовых макросиноптических процессов. Для решения этой проблемы предлагается при разработке метода среднесрочного прогнозирования полей метеорологических величин на определенной территории использовать разработанную авторами модель классификатора макросиноптических процессов. Результаты исследования предполагается использовать при автоматизации методов физико-статистической интерпретации результатов численного прогнозирования полей гидрометеорологических величин.

Ключевые слова: *среднесрочное прогнозирование полей метеорологических величин, статистические методы, метод аналогов, макросиноптические процессы, коэффициент корреляции.*

This article is devoted to medium-range weather forecasting methods. Special attention is paid to the problem of the method of analogues. This problem lies in formation of a list of typical macro-synoptic processes. In order to solve this problem the authors offer to use the model of the typical macro-synoptical processes, designed by them, for the development of the medium-range forecasting methods of the meteorological conditions on the defined territory. Findings of the investigation are intended to be used in automation of physical-statistical methods of the interpretation results of the numerical weather forecasting.

Keywords: *medium-term weather forecasting, physical statistical methods, analog method, macro-synoptical processes, correlation coefficient.*

Введение

В настоящее время в интересах различных отраслей экономики получили развитие технологии выпуска среднесрочных прогнозов на сроки от двух до десяти дней. Так, в сфере разработки методов средне- и долгосрочного прогнозирования отчетливо выделились несколько направлений, использующих существенно отличающиеся по своему характеру предположения об объекте прогнозирования. Можно выделить следующие подходы к методам прогнозирования: *синоптический, гидродинамический и статистический.*

Синоптические методы позволяют получить качественные характеристики закономерностей пространственного распределения погоды на основе анализа синоптических карт. Такие методы до сих пор применяются в оперативной практике прогнозирования, хотя и имеют определенные недостатки в виде зависимости от квалификации и интуиции прогнозиста. Гидродинамические методы позволяют производить предвычисление будущих значений тех или иных атмосферных параметров в синоптическом масштабе путем численного интегрирования по времени уравнений гидродинамики и термодинамики атмосферы при наблюдаемых начальных и выбранных граничных условиях.

Несмотря на значительные успехи в области развития гидродинамических методов прогнозирования, остается целый ряд нерешенных проблем:

а) модели общей циркуляции атмосферы применимы лишь для установившегося движения воздушных масс;

б) при заданных краевых условиях возможно существование нескольких стационарных решений системы уравнений гидротермодинамики;

в) параметризация процессов подсеточного масштаба вносит субъективные ошибки в модель.

Статистические методы основаны на выявлении вероятностных связей между параметрами атмосферы в различные моменты времени на основе обработки результатов наблюдений за погодой. К достоинствам этих методов следует отнести объективность установленных статистических связей при соблюдении условия репрезентативности выборки исходных данных, широкий перечень прогнозируемых параметров атмосферы и возможность статистической обработки результатов гидродинамического прогноза полей метеорологических величин. В зависимости от применяемого статистического аппарата выделяют следующие методы среднесрочного и долгосрочного прогнозирования: метод аналогов, регрессионный, корреляционный, факторный, экстраполяционный и дискриминантный [1, 2, 7].

Одним из самых распространенных методов долгосрочного прогноза погоды является метод аналогов. Метод заключается в принятии допущения о сходстве протекания синоптических процессов при их совпадении в некоторый момент времени в прошлом. Для оценки аналогичности полей разработан ряд объективных показателей аналогичности, которые позволяют устанавливать степень сходства между сравниваемыми полями, или макропроцессами, при помощи компьютерных алгоритмов.

В общем виде схема метода аналогов представлена на рис. 1.

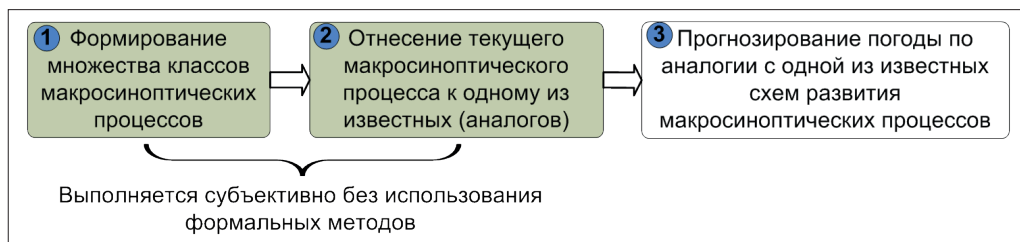


Рис. 1. Схема метода аналогов

Постановка задачи разработки модели классификатора макросиноптических процессов

Основная проблема метода аналогов — это проблема подбора самих аналогов, которые авторами статьи названы типовыми макросиноптическими процессами. В этой связи представляется чрезвычайно важным разработка модели классификатора макросиноптических процессов, позволяющей адекватно и корректно осуществить классификацию макросиноптических процессов для заданного географического региона с целью использования результатов работы данной модели при разработке метода среднесрочного прогнозирования гидрометеорологических условий на определенной территории.

В связи с тем, что процесс отнесения процесса к тому или иному аналогу связан с теорией распознавания образов, модель классификатора макросиноптических процессов будем разрабатывать, исходя из математического аппарата данной теории. Под моделью классификатора макросиноптических процессов будем понимать модель, осуществляющую формирование классов реализаций макросиноптических процессов на основе математического аппарата теории распознавания образов [6, 8]. Под классификацией понимается процесс отнесения реализации макросиноптического процесса к тому или иному классу реализаций этих процессов.

В общем виде этапы разработки модели классификатора макросиноптических процессов представлены на рис. 2.

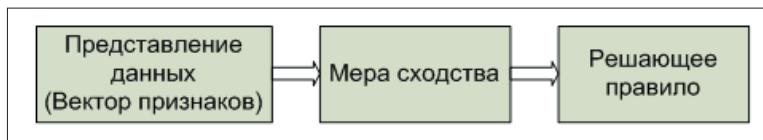


Рис. 2. Этапы разработки модели классификатора макросиноптических процессов

Как видно из рис. 2, для осуществления классификации макросиноптических процессов модель классификатора макросиноптических процессов должна включать элементы, которые выполняют следующие операции:

- кодирование (описание) макросиноптического процесса наборами наиболее информативных признаков, формирование вектора признаков;
- сравнение двух векторов из пространства признаков между собой на основе меры сходства;
- принятие решения об отнесении на основе решающего правила реализации макросиноптического процесса к одной из областей пространства признаков и указание в качестве решения номера, соответствующего этой области класса макросиноптических процессов.

Для того чтобы определить на множестве реализаций макросиноптических процессов, представленных в виде полей давления и геопотенциала, данный класс, необходимо в первую очередь ввести меру сходства (подобия), которая может быть положена в основу правила отнесения конкретной реализации макросиноптических процессов к определенному классу реализаций.

Конкретный метод представления метеорологической информации накладывает определенные ограничения на показатели схожести векторов — наборов признаков. Исходя из того что поля давления у земли и геопотенциала на изобарической поверхности 500 гПа представляются в узлах регулярной сетки, а также на основании результатов исследований, представленных в [3, 4], то меру сходства между ними будем определять с помощью коэффициента корреляции, который характеризует степень тесноты взаимосвязи двух анализируемых барических полей.

Сравнение двух полей будет производиться по величине коэффициента корреляции, который рассчитывается по формуле (1):

$$\tilde{r}_{\hat{P}_{(n)}^u, \hat{P}_{(n)}^c} = \frac{\frac{1}{n-1} \sum_{i=1}^n (p_i^u - \tilde{M}_{\hat{P}_{(n)}^u})(p_i^c - \tilde{M}_{\hat{P}_{(n)}^c})}{\tilde{\sigma}_{\hat{P}_{(n)}^u} \tilde{\sigma}_{\hat{P}_{(n)}^c}}, \quad (1)$$

где $\tilde{M}_{\hat{P}_{(n)}^u} = \frac{1}{n} \sum_{i=1}^n p_i^u$ — оценка математического ожидания исходного поля приземного давления, или геопотенциала, в качестве которой выступает пространственная средняя;

$\tilde{M}_{\hat{P}_{(n)}^c} = \frac{1}{n} \sum_{i=1}^n p_i^c$ — среднее значение по сравниваемому полю приземного давления, или геопотенциала;

$\tilde{\sigma}_{\hat{P}_{(n)}^u} = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i^u - \tilde{M}_{\hat{P}_{(n)}^u})^2}$ — оценка среднего квадратического отклонения исходного поля приземного давления, или геопотенциала;

$\tilde{\sigma}_{\hat{P}_{(n)}^c} = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i^c - \tilde{M}_{\hat{P}_{(n)}^c})^2}$ — оценка среднего квадратического отклонения

сравниваемого поля приземного давления, или геопотенциала; $p_i^u = p^u(\varphi_i, \lambda_i)$ — исходное значение приземного давления, или геопотенциала, в точке с координатами φ_i, λ_i ; $p_i^c = p^c(\varphi_i, \lambda_i)$ — сравниваемое значение приземного давления, или геопотенциала, в точке с координатами φ_i, λ_i ; n — объем анализируемой выборки.

Сравнение двух полей происходит путем сравнения значений приземного давления, или геопотенциала, в каждой точке одного поля со значениями в каждой точке другого поля в узлах регулярной сетки, на которой они представлены. Рассмотрим правило, по которому сравниваются два поля между собой.

После расчета значений коэффициентов между двумя полями давления (геопотенциала) $\tilde{r}_{\hat{P}_{(n)}^u, \hat{P}_{(n)}^c}$ по формуле (1) для коэффициента корреляции необходимо задать пороговое значение r_{por} , с которым будут сравниваться полученные значения $\tilde{r}_{\hat{P}_{(n)}^u, \hat{P}_{(n)}^c}$. Сравнение происходит по критерию (2):

$$\tilde{r}_{\hat{P}_{(n)}^u, \hat{P}_{(n)}^c} \geq r_{por}, P_{(n)}^u \in \{S_k\} \Rightarrow P_{(n)}^c \in \{S_k\}, \quad (2)$$

В соответствии с этим критерием поле $P_{(n)}^c$ считается принадлежащим к множеству $\{S_k\}$ реализаций k -го макросиноптического процесса, если значение коэффициента корреляции превышает некоторое пороговое значение r_{por} . Необходимо задать оптимальное пороговое значение, которое может быть задано опытным оператором (инженером-синоптиком). Изменяя пороговое значение коэффициента корреляции r_{por_eksp} , можно получить три разные ситуации при классификации реализаций макросиноптических процессов:

- первая классификация — при $r_{por} > r_{por_eksp}$ (рис. 3);
- вторая классификация — при $r_{por} = r_{por_eksp}$ (рис. 4);
- третья классификация — при $r_{por} < r_{por_eksp}$ (рис. 5).

В первом случае, когда $r_{por} > r_{por_eksp}$, реализации одного и того же макросиноптического процесса могут попадать в разные классы. Таким образом, возникнет ситуация, когда несколько классов различных макросиноптических процессов будут содержать реализации одного и того же макросиноптического процесса. В этом случае число классов реализаций макросиноптических процессов будет неоправданно большим.

Во втором случае, когда $r_{por} = r_{por_eksp}$, реализации макросиноптических процессов будут попадать в «свои» классы.

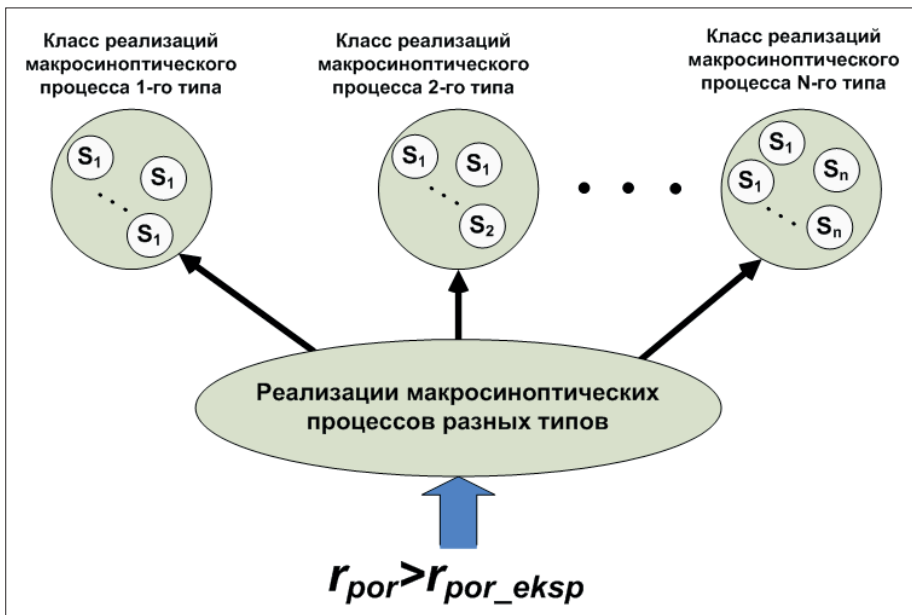


Рис. 3. Классификация реализаций макросиноптических процессов при пороговом значении коэффициента корреляции $r_{por} > r_{por_eksp}$

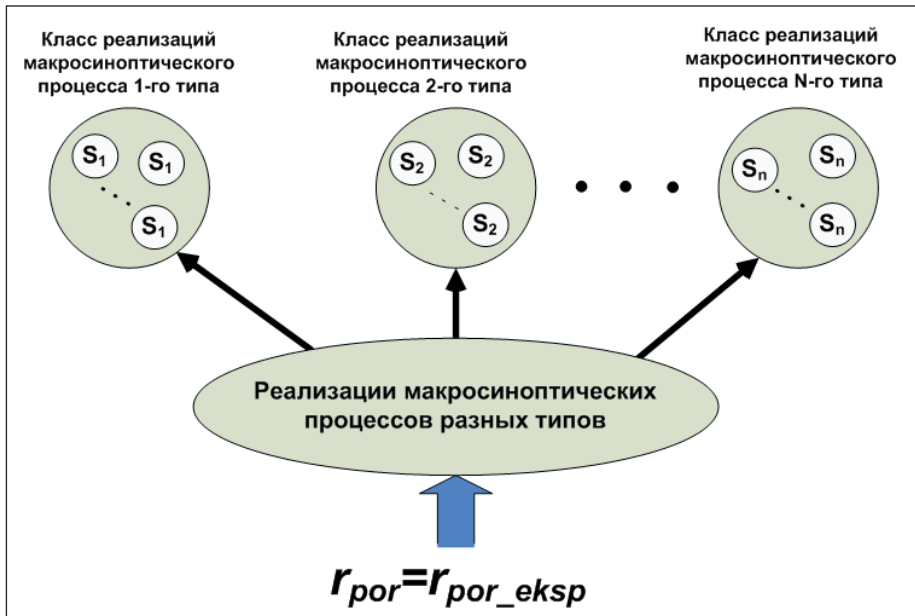


Рис. 4. Классификация реализаций макросиноптических процессов при пороговом значении коэффициента корреляции $r_{por} = r_{por_eksp}$

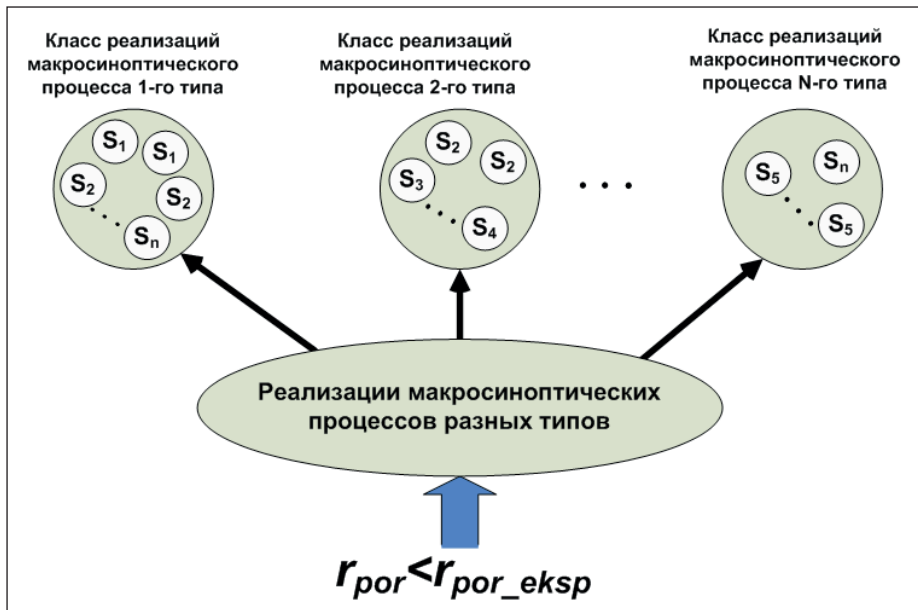


Рис. 5. Классификация реализаций макросиноптических процессов при пороговом значении коэффициента корреляции $r_{por} < r_{por_eksp}$

То есть реализации одного и того же макросиноптического процесса не будут попадать в другие классы, а будут относиться именно к «своему» классу. В этом случае число классов реализаций макросиноптических процессов будет соответствовать действительности.

В третьем случае, когда $r_{por} < r_{por_eksp}$, реализации разных макросиноптических процессов могут попадать в один класс. Таким образом, из анализа последствий вариации порогового значения коэффициента корреляции r_{por} следует то, что выбор наиболее оптимального его значения должен осуществлять оператор (инженер-синоптик), т. к. именно он проверяет результаты работы модели классификатора макросиноптических процессов путем просмотра сформированных классов их реализаций. Варьируя величиной порогового значения, он добьется такого результата, когда $r_{por} = r_{por_eksp}$, т. е. все реализации будут отнесены каждый к «своему» классу.

На рис. 6 представлена структура модели классификатора макросиноптических процессов. В блоке расчета коэффициентов корреляции между двумя полями метеорологических параметров происходит вычисление значений коэффициентов корреляции $\tilde{r}_{\hat{P}_{(n)}^u, \hat{P}_{(n)}^c}$ всех векторов $\hat{P}_{(n)}$ метеорологических параметров, входящих в состав выборки, и формируется матрица значений коэффициентов корреляции

$$\left\| \tilde{r}_{\hat{P}_{(n)}^u, \hat{P}_{(n)}^c} \right\|_h^h.$$

После расчета матрицы значений коэффициентов корреляции происходит формирование классов реализаций макросиноптических процессов следующим образом:

– сравнивается рассчитанное значение коэффициента корреляции $\tilde{r}_{\hat{P}_{(n)}^u, \hat{P}_{(n)}^c}$ с пороговым значением r_{por} , введенным экспертом вручную. Если

$$\tilde{r}_{\hat{P}_{(n)}^u, \hat{P}_{(n)}^c} \geq r_{por}, \quad (3)$$

то все поля данных метеорологических параметров образуют первый класс реализаций макросиноптического процесса первого типа;

– после завершения формирования первого класса реализаций макросиноптического процесса первого типа ищется поле, не вошедшее в первый сформированный класс реализаций, и повторяется описанная выше процедура формирования последующих классов до полного окончания выборки;

– результатом станут сформированные классы реализаций макросиноптических процессов.

Для проверки адекватности результатов работы классификатора макросиноптических процессов результатам ручной классификации проводились численные эксперименты по проверке зависимости числа классов от порогового значения коэффициента корреляции r_{por} и даты начала проведения классификации (даты начала отсчета).

Для проведения численных экспериментов использовался архив данных метеорологических наблюдений за 2013–2015 гг. с дискретностью по времени

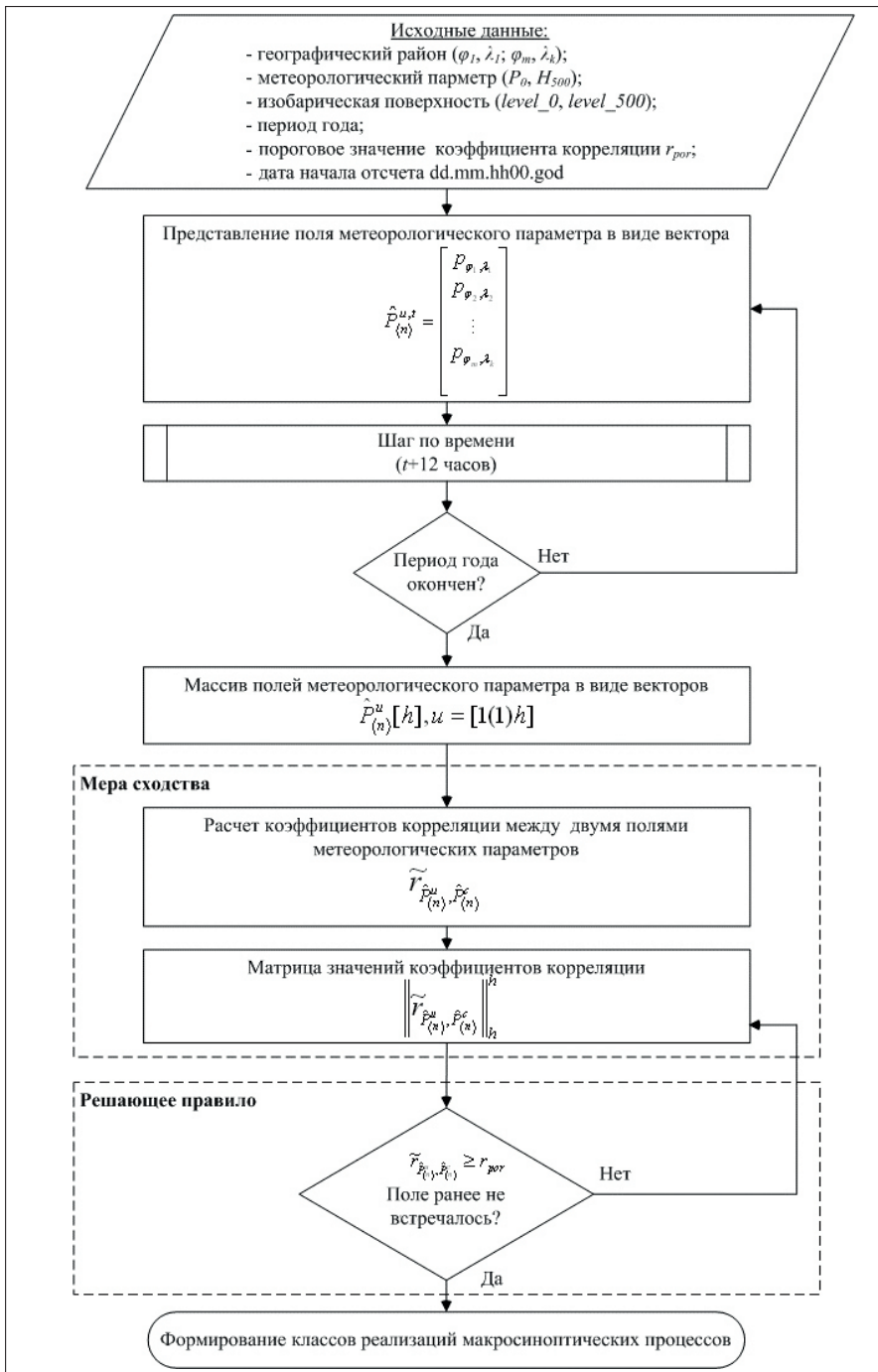


Рис. 6. Структура модели классификатора макросиноптических процессов

в 12 ч в сроки 00 ч и 12 ч СГВ на изобарической поверхности 500 гПа и у поверхности земли для района 30–70° с. ш. и 25–65° в. д. (Европейская территория РФ) с шагом 1° по широте и 1° по долготе. Общее количество полей равно 2192 для каждой поверхности. Исходные данные для эксперимента были получены с помощью автоматизированного рабочего места военного гидрометеоролога, где представление данных в узлах регулярной сетки точек обеспечивает процедура сплайн-интерполяции. Расчет производился для теплого и холодного периодов года.

Для получения устойчивых оценок параметров статистических связей между соответствующими аналогичными полями приземного давления и геопотенциала изобарической поверхности 500 гПа необходимо, чтобы используемая выборка исходных данных полей метеорологических величин за 2013–2015 гг. была однородной, а ряды элементов выборки — стационарными, т. е. их средние значения и дисперсии оставались постоянными на весь период их оценивания.

Проверка выборки на однородность осуществлялась с помощью непараметрического рангового показателя Вилкоксона–Манна–Уитни [5]. Нулевая гипотеза H_0 формулировалась в виде равенства функций распределений F :

$$H_0: F(x_1) = F(x_2), \quad (4)$$

где x_1 — меньшая выборка; x_2 — большая выборка из объединенной выборки.

Критическая статистика показателя Вилкоксона–Манна–Уитни представляет собой сумму рангов n_0 элементов упорядоченной выборки объемом $n_0 = n_1 + n_2$:

$$S_R = \sum_{i=1}^{n_0} R_i, \quad (5)$$

где R_i — ранги элементов упорядоченной выборки; n_1 — число элементов в меньшей выборке; n_2 — число элементов в большей выборке.

Нулевая гипотеза подтвердилась при уровне значимости $\alpha = 5\%$ как для полей приземного давления, так и геопотенциала АТ-500.

Стационарность распределений проверялась с помощью непараметрического одновыборочного показателя Вилкоксона, основанного на использовании абсолютных рангов [5]. Из исходных выборок переменных x_i получались выборки преобразованных переменных $x_i = x_j - x_{j_2}$, где x_i — среднее значение переменной i . Одновыборочный критерий Вилкоксона использует статистику в виде суммы рангов R_i для положительных преобразованных переменных:

$$S_+ = \sum_{i=1}^l R_i. \quad (6)$$

Критическая область отвержения нулевой гипотезы формировалась для уровня значимости $\alpha = 5\%$ и позволила подтвердить гипотезу стационарности полей приземного давления и геопотенциала АТ-500.

Значимость коэффициентов корреляции между значениями приземного давления и геопотенциала АТ-500 и аналогами, которые выступали в роли предикторов, проверялась с помощью критериев Стьюдента t при уровне значимости гипотезы $\alpha = 5\%$ [5]:

$$t = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}}, \quad (7)$$

где r_{xy} — коэффициент корреляции между переменными x и y ; n — объем выборки. Коэффициенты корреляции были значимы, т. к. для всех предикторов оказалось, что $t > t_{кр}(\alpha, V = n - k - 1)$, где α — уровень значимости; $V = n - k - 1$ — число степеней свободы; $t_{кр}$ — критическое значение статистики из распределения Стьюдента.

Результаты проведения численных экспериментов представлены в табл. 1 и на рис. 7. При проведении численных экспериментов варьировалась дата начала отсчета и пороговое значение коэффициента корреляции.

Таблица 1

Результаты численных экспериментов с моделью классификатора макросиноптических процессов для изобарической поверхности 500 гПа

Дата начала отсчета	Значение коэффициента корреляции r для изобарической поверхности 500 гПа									
	0,68	0,70	0,71	0,73	0,75	0,78	0,80	0,85	0,90	0,95
	Рассчитанное число классов, ед									
01.01.2014	11	13	16	18	20	23	29	47	93	253
01.03.2015	11	15	15	17	18	26	30	50	99	252
01.06.2013	12	14	14	14	16	21	26	48	98	255
01.08.2014	11	13	14	17	20	22	27	51	96	257
01.10.2015	12	14	14	17	20	25	26	49	101	253
01.12.2013	12	13	15	16	19	23	28	50	100	254

Из анализа табл. 1 следует сделать вывод, что дата начала отсчета не играет роли при определении количества классов реализаций макросиноптических процессов. При задании произвольной даты начала отсчета количество определенных типов макросиноптических процессов меняется незначительно, в пределах 5 %, что является допустимым и говорит о независимости количества классов реализаций макросиноптических процессов от даты начала отсчета.

Анализ результатов проведенных экспериментов

Анализ изменения количества классов реализаций макросиноптических процессов в зависимости от порогового значения коэффициента корреляции показывает уменьшение количества классов реализаций макросиноптических процессов с уменьшением порогового значения коэффициента корреляции и наоборот: увеличение количества классов реализаций макросиноптических процессов при увеличении порогового значения коэффициента корреляции (рис. 7).

Было выявлено, что увеличение количества классов имеет нелинейный характер. В диапазоне коэффициентов корреляции 0,68...0,70 количество классов увеличивается на 2...4 для разных дат начала отсчета, в диапазоне 0,70...0,75 — на 2...9, в диапазоне 0,75...0,80 — на 6...12, в диапазоне 0,80...0,85 — на 18...24,

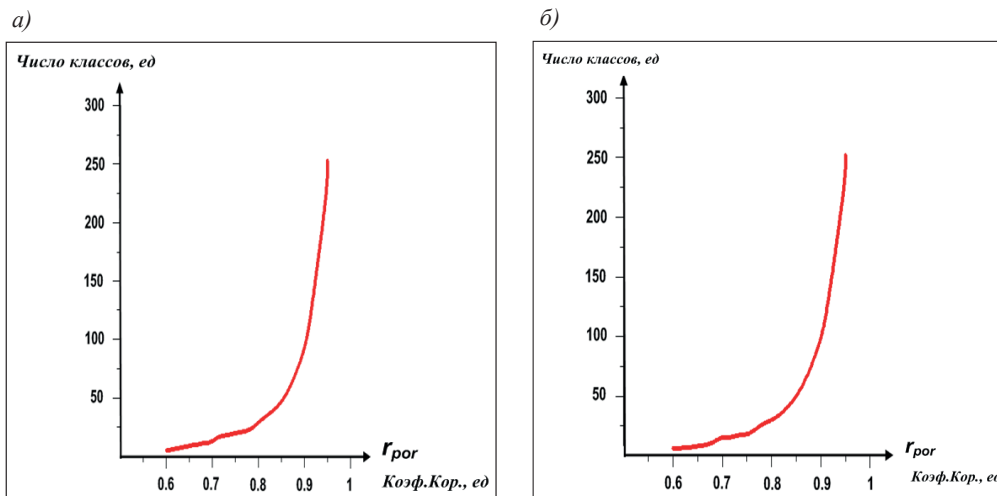


Рис. 7. Зависимость числа классов реализаций макросиноптических процессов от порогового значения коэффициента корреляции для изобарической поверхности 500 гПа: а — для даты начала отсчета 01.01.2014 г.; б — для даты начала отсчета 01.03.2015 г.

в диапазоне 0,85...0,9 — на 48...55, в диапазоне 0,90...0,95 — на 152...160 классов, т. е. в каждой последующей градации количество классов возрастает в 2 и более раз. Значительное увеличение числа классов наблюдается уже со значения коэффициента корреляции, равного 0,70, которое выбирается в качестве порогового.

Выбранное пороговое значение коэффициента корреляции соответствует требованиям корреляционного анализа (табл. 2 [8]) и определяет связь между анализируемыми полями изобарической поверхности 500 гПа как сильную.

Таблица 2

Критерии оценки тесноты связи между переменными

Значение коэффициента корреляции r	Характер связи между переменными
$r < \pm 0.3 $	связь практически отсутствует
$ \pm 0.3 \leq r < \pm 0.5 $	связь слабая
$ \pm 0.5 \leq r < \pm 0.7 $	связь умеренная
$ \pm 0.7 \leq r < \pm 1.0 $	связь сильная

Проверка адекватности модели осуществлялась путем сравнения числа полученных классов с помощью модели классификатора и полученных вручную специалистами-метеорологами для Европейской территории страны. Расчеты показали, что при пороговом значении коэффициента корреляции 0,70 количество классов и входящие в их состав реализации макросиноптических процессов соответствуют тем, что ранее получены специалистами-метеорологами вручную.

Заключение

Рассмотрение методов среднесрочного прогнозирования полей метеорологических величин показало, что среди физико-статистических методов наиболее

перспективным является метод аналогов. Введение типовых макросиноптических процессов в качестве аналогов позволило разработать модель классификатора, которая корректно определяет тип макросиноптического процесса в заданном географическом регионе в интересах среднесрочного прогнозирования полей приземного давления и геопотенциала изобарической поверхности 500 гПа. В качестве количественной меры аналогичности использовался коэффициент корреляции между исходными и сравниваемыми полями в узлах регулярной сетки точек.

Проверка выборок исходных данных за 2013–2015 гг. на однородность и стационарность производилась с помощью непараметрического рангового показателя Вилкоксона–Манна–Уитни и одновыборочного показателя Вилкоксона. Значимость коэффициентов корреляции проверялась с помощью критерия Стьюдента при уровне значимости гипотезы 5 %. Проверка адекватности результатов модели классификации макросиноптических процессов осуществлялась путем сравнения с результатами классификации, полученными специалистами вручную.

Научная новизна выполненного исследования заключается в возможности использования разработанной модели классификатора при автоматизации методов статистической интерпретации результатов численного прогнозирования гидрометеорологических полей.

Литература

1. Вальд А. Последовательный анализ / пер. с англ. – М.: Мир, 1960. — 375 с.
2. Вероятностные методы в прикладной кибернетике / под ред. Р.М. Юсупова. — Б. М.: МО СССР, 1976. — 426 с.
3. Груза Г.В., Ранькова Э.Я. Оценка различий некоторых метеорологических объектов и их предсказуемости методом аналогов / Тр. ВНИИГМИ-МЦД, 1977. Вып. 53. — С. 3–12.
4. Груза Г.В., Рейтенбах Р.Г. Статистика и анализ гидрометеорологических данных. — Л.: Гидрометеоиздат, 1982. — 215 с.
5. Малинин В.Н. Статистические методы анализа гидрометеорологической информации. — СПб.: РГГМУ, 2008. — 407 с.
6. Цыпкин Я.З. Информационная теория идентификации. — М.: Наука, 1995. — 336 с.
7. Чичасов Г.Н. Технология долгосрочных прогнозов погоды. — СПб.: Гидрометеоиздат, 1991. — 180 с.
8. Ширяев А.И. Статистический исследовательский анализ. — М.: Наука, 1976. — 229 с.