

УДК 551.501; 551.581

# Климатическое районирование территории Сибири по режиму общей и нижней облачности как основа для построения локальных облачных моделей атмосферы.

## Часть 1. Методические основы

В.С. Комаров, Д.П. Нахтигалова, С.Н. Ильин,  
А.В. Лавриненко, Н.Я. Ломакина\*

*Институт оптики атмосферы им. В.Е. Зуева СО РАН  
634021, г. Томск, пл. Академика Зуева, 1*

Поступила в редакцию 1.04.2014 г.

Рассмотрена оригинальная методология объективной классификации метеорологических объектов и прикладного климатического районирования территории Сибирского региона по режиму общей и нижней облачности, основанная на использовании метода кластерного анализа с применением в качестве информативных признаков среднесезонных значений количества общей и нижней облачности и повторяемостей 5 градаций (0, 1–3, 4–6, 7–9 и 10 баллов), а в качестве меры однородности признаков – характеристик степени их коррелированности и специальных критериев сходства.

*Ключевые слова:* методика, объективная классификация объектов, режим общей и нижней облачности, районирование территорий; technique, objective classification, total and lower cloudiness conditions, climatic zoning.

Решение широкого круга научных и прикладных задач, связанных, в частности, с исследованием и моделированием климата, с разработкой и использованием спутниковых систем дистанционного зондирования окружающей среды, с оценкой условий распространения в атмосфере оптического и инфракрасного излучения и т.п., требует существенного углубления наших знаний о пространственной структуре полей облачности. При этом особенно важными при решении различных задач дистанционного зондирования атмосферы с земной поверхности и из космоса являются сведения о пространственном распределении облаков нижнего яруса, поскольку в реальных условиях наличие облачности нижнего яруса на линии визирования измерительной аппаратуры, работающей в оптическом и инфракрасном диапазонах спектра, может полностью исключить саму возможность получения какой-либо информации о состоянии земной поверхности и подоблачного слоя атмосферы при их дистанционном зондировании из космоса [1] или о состоянии надоблачного слоя при применении лидарных систем наземного базирования [2]. Здесь следует подчеркнуть, что априорные статистические данные о поле облачности должны

быть ограниченными по своему объему и охватывать значительные по площади территории. В связи с этим для соблюдения подобных условий необходимо использовать определенные принципы оптимального обобщения и преобразования климатической информации, получаемой для отдельных станций, которые бы учитывали естественную пространственную изменчивость облачного покрова.

Одним из самых перспективных подходов к решению такой проблемы является подход, основанный на модельном представлении обобщенных параметров состояния облачного покрова с учетом результатов объективной классификации метеорологических объектов и прикладного климатического районирования заданных территорий.

С учетом этого на практике (например, при разработке методов и средств лидарного зондирования атмосферы, имитационном моделировании фоновой обстановки из космоса, оперативном и текущем планировании работы систем дистанционного зондирования и т.п.) стали широко использовать глобально-региональные и региональные климатические облачные модели атмосферы (см., например, [3–8]), построенные с учетом результатов прикладного климатического районирования заданных территорий по режиму облачности (главным образом на основе гистограмм распределения количества общей облачности) и описывающие состояние облачного покрова в каждом из выявленных однородных облачных районов (ООР).

\* Валерий Сергеевич Комаров (gfm@iao.ru); Дарья Петровна Нахтигалова; Сергей Николаевич Ильин; Андрей Викторович Лавриненко; Наталья Яковлевна Ломакина (luya@iao.ru).

Однако из-за значительной изменчивости поля облачности в пространстве эти модели не отражают в должной мере его локальных особенностей в отдельных физико-географических районах, являющихся частью обширных регионов, таких, например, как Сибирь, и формирующихся не только под воздействием синоптических и мезомасштабных, но и микромасштабных атмосферных процессов. Кроме того, глобально-региональные и региональные облачные модели атмосферы обычно не включают в свой состав данных о климатических показателях облаков нижнего яруса, особенно сильно влияющих на эффективность дистанционного зондирования атмосферы, проводимого из космоса и с платформ наземного базирования.

Попытка построения локальных моделей атмосферы для Сибирского региона с учетом результатов его климатического районирования, проведенного по комплексу «температура, влажность, ветер», была предпринята в последние годы в Институте оптики атмосферы им. В.Е. Зуева СО РАН [9–11].

В настоящей статье излагаются результаты дальнейших исследований по рассматриваемой проблеме, связанной с разработкой методики объективной классификации и климатического районирования территорий по режиму общей и нижней облачности и построением локальных облачных моделей атмосферы.

## 1. Некоторые замечания по рассматриваемой проблеме

Современные исследования, проводимые в области прикладного климатического районирования с целью построения региональных и локальных физико-статистических моделей атмосферы, показывают, что эффективность такого районирования зависит в значительной степени от выбора метода объективной классификации метеорологических объектов, которая, согласно [12], заключается в разбиении совокупности объектов, характеризующихся большим числом информативных признаков, на группы (классы), в каждый из которых должны входить объекты, в определенном смысле однородные или обладающие схожими свойствами. При этом обычно классифицируются не сами объекты, а их выбранные информативные признаки.

Из анализа имеющихся публикаций (см., например, [3, 5–8, 13, 14]) следует, что для решения задачи климатического районирования территорий по режиму облачности широко используются методы многомерного кластерного анализа. Выбор подобных методов обусловлен тем, что они позволяют решать задачу объективной классификации метеорологических объектов при отсутствии какой-либо априорной информации о наблюдениях внутри классов [12].

Сущность методов кластерного анализа, согласно [12, 15], состоит в том, чтобы на основе множества векторов измерений  $X_{(n)}$  (оно образует пространство признаков  $E$ , описывающих заданные объекты  $A_{(n)}$ ) разбить эти объекты с помощью некоторой меры сходства на ограниченное число классов (под-

множеств), причем таким образом, чтобы объекты  $a_i$  принадлежали лишь к одному классу и были между собой схожими.

В метеорологических классификациях в качестве меры сходства для двух признаков, обозначаемых через  $k$ -мерные векторы  $X_i = (x_{1i}, x_{2i}, \dots, x_{ki})^T$  и  $X_j = (x_{1j}, x_{2j}, \dots, x_{kj})^T$  (здесь  $T$  – операция транспонирования), обычно используются меры расстояний, среди которых чаще всего берется евклидово расстояние [8, 12, 16]:

$$d_{E_y}(X_i, X_j) = \left[ \sum_{l=1}^r (x_{li} - x_{lj})^2 \right]^{1/2}. \quad (1)$$

В этом случае при метеорологической классификации объектов необходимо проводить нормировку информативных признаков (из-за различий в единицах их измерения) и иметь некоторый оптимальный уровень разбиения смешанной выборки на классы, являющийся пределом возможного объединения объектов в группы. Но такой уровень неизвестен [17].

Однако, как показано в [12], задача классификации объектов методом кластерного анализа может быть существенно упрощена, если в качестве меры сходства (однородности) признаков использовать характеристики степени их коррелированности (например, коэффициенты корреляции). В этом случае отпадает необходимость нормировки отдельных признаков, имеющих различные единицы измерения. Кроме того, можно будет воспользоваться результатами работы [10] и в качестве уровня разбиения множества объектов на классы использовать критическое значение коэффициента корреляции. Подобным образом поступили и авторы настоящей статьи.

## 2. Методика объективной классификации объектов и прикладного климатического районирования территорий по режиму облачности

Пусть задано некоторое множество объектов  $A_{(n)} = \{a_1, a_2, \dots, a_n\} = \{a_i\}$  при  $i = 1, 2, \dots, n$ , образующих смешанную выборку и характеризующих совокупностью информативных признаков, т.е. каждому объекту может соответствовать  $k$ -мерный вектор наблюдаемых характеристик  $X_{(k)} = \langle x_1, x_2, \dots, x_k \rangle^T$ , а всему множеству объектов  $A_{(n)}$  – множество векторов  $X_{(n)} = \{X_1, X_2, \dots, X_n\}^T$ , образующее пространство информативных признаков  $E$ .

Тогда задача объективной классификации заключается в том, чтобы на основе данных, содержащихся во множестве  $X_{(n)}$ , провести разбиение множества объектов  $A_{(n)}$  на  $m$  классов, причем таким образом, чтобы объект  $a_i$  принадлежал только одному из них и чтобы объекты, принадлежащие одному и тому же классу, были схожими, в то время как объекты, принадлежащие разным классам, были разнородными. При этом само разбиение множества объектов  $A_{(n)}$  на классы осуществляется в два этапа [18]:

вначале, исходя из требований решаемой задачи, проводится отбор совокупности информативных признаков, характеризующих объекты (в нашем случае режим облачности в районе заданной станции), а затем осуществляется построение разделяющего или решающего правила, позволяющего объективным образом выделить границы между классами в многомерном пространстве информативных признаков, т.е. провести климатическое районирование заданной территории.

В настоящей статье для объективной классификации режимов облачности и климатического районирования территории Сибирского региона в качестве информативных признаков использованы:

– среднесезонные значения количества общей  $\bar{N}$  и нижней  $\bar{N}_n$  облачности (в баллах), предварительно рассчитанные по данным 8-срочных наблюдений для каждого года взятого 44-летнего периода (1969–2012 гг.);

– эмпирические вероятности (повторяемости) пяти градаций общей  $P_l^{(N)}$  и нижней  $P_l^{(N_n)}$  облачности: 0, 1–3, 4–6, 7–9 и 10 баллов, представленные в процентах и определенные для четырех календарных сезонов по данным каждого года того же многолетнего периода.

Указанные информативные признаки для каждого объекта могут быть представлены с помощью некоторого  $k$ -мерного вектора (в нашем случае  $k = 12$ ), записываемого в виде

$$X_{(k)} = \langle x_1 = \bar{N}, x_2 = P_1^{(N)}, \dots, x_6 = P_5^{(N)}, x_7 = \bar{N}_n, x_8 = P_1^{(N_n)}, \dots, x_{12} = P_5^{(N_n)} \rangle^T. \quad (2)$$

В качестве меры однородности признаков  $x_i$  и  $x_j$ , т.е. компонент векторов  $X_i$  и  $X_j$ , взята характеристика степени коррелированности, представляющая собой критерий сходства (подобия)  $r_{ij}^{(x)}$ , который можно определить с помощью выражения вида

$$r_{ij}^{(x)} = r_x(\rho_i, \rho_j) = \frac{1}{\sigma_x(\rho_i)\sigma_x(\rho_j)} \times \left[ \frac{1}{m} \sum_{l=1}^m x_l(\rho_i)x_l(\rho_j) - \bar{x}(\rho_i)\bar{x}(\rho_j) \right], \quad (3)$$

где  $\bar{x}(\rho_i) = \frac{1}{m} \sum_{l=1}^m x_l(\rho_i)$  – многолетнее среднесезонное

значение признака в точке  $\rho_i$  (здесь  $\rho$  – радиус-вектор точки (станции), а в качестве ее пространственных координат взяты широта  $\varphi$  и долгота  $\lambda$ ), рассчитанное по данным  $m$ -го числа лет;  $x_l(\rho_i)$  – среднесезонное значение того же признака за  $l$ -й год;

$\sigma_x(\rho_i) = \left[ \frac{1}{m} \sum_{l=1}^m x_l(\rho_i) - \bar{x}(\rho_i) \right]^{1/2}$  – среднеквадратическое отклонение признака  $x$  в точке  $\rho_i$ .

Критерии подобия  $r_{ij}^{(x)}$  рассчитываются отдельно для каждой компоненты вектора (2), поскольку для объективной классификации объектов на осно-

ве информативных признаков  $x$  берется их среднее значение, определяемое по формуле

$$\bar{r}_{ij} = \frac{1}{k} \sum_1^k r_{ij}^{(x_k)} \quad \text{при } k = 1, 2, \dots, 12, \quad (4)$$

где  $r_{ij}^{(x_k)}$  – критерий подобия, рассчитанный для  $k$ -го признака.

Используя осредненные значения критериев подобия  $\bar{r}_{ij}$ , полученных по формуле (4) для каждой пары векторов  $X_i$  и  $X_j$ , описывающих объекты  $A_i$  и  $A_j$ , можно сформировать симметричную матрицу соответствующих критериев.

Сравнивая элементы этой матрицы с критическим значением  $r_{кр}$ , можно с помощью выражения вида

$$r_{ij} \geq r_{кр} = \text{th}z_{кр} \quad (5)$$

(здесь  $z_{кр} = 3\sigma_z$  – критическое значение функции Фишера [19], где  $\sigma_z = 1/\sqrt{m-3}$ , а  $m$  – объем выборки, определяемый числом взятых лет (в нашем случае  $m = 44$  годам)) оценить степень сходства (однородности) для каждой пары векторов  $X_i$  и  $X_j$ . Поскольку для поля облачности характерна большая пространственная изменчивость, то для оценки сходства векторов  $X_i$  и  $X_j$  может быть использован менее жесткий критерий  $z_{кр} = 2,5\sigma_z$  [20].

С учетом сказанного рассмотрим алгоритм сопоставления векторов  $X_i$  и  $X_j$ , используемый в настоящей статье.

Вначале по данным выборки объема  $m$  (в нашем случае  $m = 44$ ) определяется величина  $\sigma_z = 1/\sqrt{m-3} = 1/\sqrt{44-3} = 0,156$ , далее вычисляется  $z_{кр} = 2,5\sigma_z = 2,5 \cdot 0,156 = 0,390$  и в заключение оценивается (например, с помощью математических таблиц [21]) значение  $r_{кр} = \text{th}z_{кр} = \text{th}0,390 = 0,371$ . Сравнение значения  $r_{кр}$  с эмпирическими величинами критериев подобия  $\bar{r}_{ij}$  позволяет провести объективную классификацию объектов и определить границы однородных облачных районов по режиму общей и нижней облачности. При этом если соблюдается условие

$$\bar{r}_{ij} \geq r_{кр} = 0,371, \quad (6)$$

то обе сравниваемые станции относятся к одному и тому же классу (в нашем случае к одному однородному облачному району), а если

$$\bar{r}_{ij} < r_{кр} = 0,371, \quad (7)$$

то сравниваемые станции принадлежат к двум различным однородным облачным районам.

Поскольку распределение метеорологических станций по территории Сибири является неоднородным, нами применен метод пространственной интерполяции данных отдельных станций в узлы регулярной сетки с шагом 100 км по широте и 100 км по долготе, основанный на использовании алгоритма фильтра Калмана и учете трех влияющих станций, расположенных наиболее близко к узлу сетки. При

этом величина, определяемая в узле сетки, складывается из суммы средневзвешенных значений данных трех влияющих станций обратно пропорционально расстоянию до каждой из них и слагаемого, полученного в результате работы алгоритма фильтра Калмана, оценивающего методом наименьших квадратов вклад каждой из 60 станций в рассматриваемый узел сетки. Данный подход является в чем-то близким к методу локальных полиномов, который используется в современных ГИС [22].

Таким образом, путем сопоставления с помощью выражения (5) всех элементов матрицы критериев подобия  $\bar{r}_{ij}$  с критическим значением  $r_{кр}$  можно объективным образом выделить на рассматриваемой территории (в нашем случае на территории Сибирского региона) отдельные однородные облачные районы, в каждом из которых состояние облачного покрова идентично в пределах всего ООР и может быть описано одной локальной облачной моделью атмосферы.

О результатах такого районирования и физико-статистического моделирования и пойдет речь во второй части настоящей статьи.

1. Кондратьев К.Я., Тимофеев Ю.М. Метеорологическое зондирование атмосферы из космоса. Л.: Гидрометеоздат, 1978. 280 с.
2. Зуев В.Е., Зуев В.В. Дистанционное оптическое зондирование атмосферы. СПб.: Гидрометеоздат, 1992. 232 с.
3. Ременсон В.А. К вопросу о классификации метеорологических объектов: Межвуз. сборник. Л.: Изд. ЛПИ им. М.И. Калинина, 1983. Вып. 82. С. 78–84.
4. Justus C.G., Woodrum A.W., Roper R.G., Smith O.E. Four-D Global Reference Atmosphere. Technical Description. Part 1. NASA Technical Memorandum, NASA TMX-64871. 1974.
5. Орлов А.И., Фадеев В.С. Климатическое районирование северного полушария на основе статистических распределений количества облаков: Межвуз. сборник. Л.: Изд. ЛПИ им. М.И. Калинина, 1983. Вып. 82. С. 86–97.
6. Комаров В.С., Ременсон В.А. Состояние и развитие моделей атмосферы для решения задач атмосферной оптики // Оптика атмосф. 1988. Т. 1, № 7. С. 3–16.
7. Бобровников В.Г. Математические модели облачности для априорного планирования наблюдений Земли из космоса // Исслед. Земли из космоса. 1981. № 1. С. 83–89.

8. Комаров В.С. Статистика в приложении к задачам прикладной метеорологии. Томск: Изд-во СО РАН, 1997. 256 с.
9. Комаров В.С., Ломакина Н.Я. Статистические модели пограничного слоя атмосферы Западной Сибири // Оптика атмосф. и океана. 2007. Т. 20, № 11. С. 1009–1012.
10. Комаров В.С., Ломакина Н.Я. Статистические модели пограничного слоя атмосферы Западной Сибири / Под общ. ред. Г.Г. Матвиевко. Томск: Изд-во ИОА СО РАН, 2008. 222 с.
11. Ломакина Н.Я., Комаров В.С., Ильин С.Н. Статистические модели высотного распределения температуры, влажности и ветра в пограничном слое атмосферы Восточной Сибири // Оптика атмосф. и океана. 2012. Т. 25, № 7. С. 629–632.
12. Калинина В.Н., Соловьев В.И. Введение в многомерный статистический анализ. М.: Изд-во Гос. ун-та управления, 2003. 66 с.
13. Лаухин С.В., Ременсон В.А., Шустер Л.Г. Использование метода последовательной кластеризации для классификации метеорологических объектов. М.: ВНИИГМИ-МЦД. Сер. Метеорол. и климатол. 1981. Вып. 9(117). С. 170–178.
14. Hughes N.A., Henderson-Sellers A. Analysis of cloud characteristics derived from archived satellite data // J. Remote Sens. 1983. V. 4, N 1. P. 159–173.
15. Дюран Б., Одел П. Кластерный анализ. М.: Статистика, 1977. 128 с.
16. Груза Г.В., Рейтенбах Р.Г. Статистический анализ гидрометеорологических данных. Л.: Гидрометеоздат, 1982. 216 с.
17. Ременсон В.А., Аникин А.Н. Об одном алгоритме последовательной кластеризации: Сб. алгоритмов и программ типовых задач. Л.: ВИКИ, 1983. Вып. 6. С. 255–263.
18. Кобышева Н.В., Наровлянский Г.Я. Климатическая обработка метеорологической информации. Л.: Гидрометеоздат, 1978. 295 с.
19. Математическая статистика / Под ред. А.М. Длинна. М.: Высш. школа, 1975. 398 с.
20. Зуев В.Е., Комаров В.С. Статистические модели температуры и газовых компонент атмосферы. Л.: Гидрометеоздат, 1986. 264 с.
21. Милн-Томсон Л.М., Комри Л.Дж. Четырехзначные математические таблицы. М.: Наука, 1964. 246 с.
22. Heywood D.I., Cornelius S., Carver S. An introduction to Geographical Information Systems (3rd-ed.). Essex. England: Pearson Prentice Hall, 2006. 426 p.

**V.S. Komarov, D.P. Nakhtigalova, S.N. Il'in, A.V. Lavrinenko, N.Ya. Lomakina. Climatic zoning of the Siberia territory according to the total and lower cloudiness conditions as a basis for construction of local cloud atmosphere models. Part 1. Methodical bases.**

The original methodology of objective classification of the meteorological objects and applied climatic zoning of the Siberian region according to the total and lower cloudiness conditions was considered. This methodology is based on the method of cluster analysis with the use of the average annual values of total and lower cloudiness amount and frequencies for five gradations 0, 1–3, 4–6, 7–9, and 10 balls as the informative signs, as well as characteristics of their correlation and specific similarity criteria as a measure of homogeneity signs.