

О.И. Алдошина, В.В. Бачериков, Е.Е. Лимар, В.А. Фабриков

РЕГРЕССИОННАЯ МОДЕЛЬ ВЕРХНЕЙ ГРАНИЦЫ ОБЛАЧНОСТИ НАД КОНТИНЕНТАМИ ЗЕМНОГО ШАРА ПО ДАННЫМ МЕЖДУНАРОДНОГО ЭКСПЕРИМЕНТА FGGE

Разработаны алгоритмы и пакет программ расчета моментов, плотностей и функций распределений, корреляционных матриц и линий регрессий для статистического анализа характеристик облачности по данным международного эксперимента FGGE. Построены сезонно-географические распределения балльности и высот верхней границы облачности над континентами земного шара. Установлено, что распределение балльности и высот верхней границы облачности над континентами может быть аппроксимировано логарифмически-нормальным законом. Аппроксимация проведена по критерию Колмогорова на уровне значимости 0,95.

Эти данные могут быть использованы для расшифровки спутниковой информации при дистанционном зондировании Земли в оптическом диапазоне.

При расшифровке спутниковой информации в оптическом диапазоне важнейшими факторами, от знания и умения прогнозировать которые зависит точность и надежность регистрации объектов из космоса, являются характеристики облачности. Анализ литературных данных показал, что наименьшей состоятельностью, а зачастую и противоречивыми данными до недавнего времени обладали оценки верхней границы облачности. Это объясняется различиями в методах и средствах наблюдений, недостаточностью выборки, различием масштабирования, неопределенностью погрешностей и локальностью измерений. Наблюдения за облаками с КА позволили поставить более корректно задачу, однако с одного КА можно осуществлять контроль за облачностью только с периодичностью, соответствующей периоду обращения КА вокруг Земли. Отсутствие глобальности по пространству и непрерывности во времени не позволяли составить объективную картину облачных покровов над всей поверхностью земного шара. В эксперименте FGGE [1] измерения высот верхней границы облаков и общего количества облачности с дискретом 00, 03, 06, 09, 12, 15, 18 и 21 час по Гринвичу за период с 01.12.78 по 10.11.79 г. были проведены в 262144 точках полярной стереографической сетки с разрешающей способностью 50 км. Ниже делается попытка на основе данных международного эксперимента FGGE построить модель распределения балльности и высот верхней границы облачности над континентами земного шара.

Для количественного исследования балльности и высот верхней границы облачности над поверхностью земного шара были использованы методы статистического анализа в районах размером $5 \times 5^\circ$ геодезической сетки помесачно и по сезонно с вычислением математических ожиданий M_x , дисперсий D_x и коэффициентов корреляции ρ [2]. Объединение результатов измерений в районы $5 \times 5^\circ$ привело к размеру обрабатываемой выборки ~ 184300 , помесачное объединение — к размеру выборки более 44 млн, что, по видимому, можно считать достаточным для получения надежных статистических выводов.

Не зная пока законов распределения исследуемых величин, мы использовали тот факт, что первая производная логарифма функции правдоподобия L при условии регулярности распределена асимптотически нормально с нулевым средним и

$$D \left(\frac{\partial \log L}{\partial \Theta} \right) = M \left\{ \left(\frac{\partial \log L}{\partial \Theta} \right)^2 \right\} = - M \left\{ \frac{\partial^2 \log L}{\partial \Theta^2} \right\},$$

где D и M соответственно дисперсия и математическое ожидание функции Θ . Тогда, обозначая

$$\Psi = \frac{\partial \log L}{\partial \Theta} / \left[M \left\{ \left(\frac{\partial \log L}{\partial \Theta} \right)^2 \right\} \right]^{1/2},$$

имеем, что Ψ при больших выборках является стандартной нормальной

случайной величиной [3]. Так как Ψ — монотонная функция от Θ , то из нормального интеграла были определены доверительные интервалы с коэффициентом доверия $1 - \alpha$ для балльности и высоты верхней границы облачности для квадратов с большой выборкой. Для районов, где большая часть была занята водным пространством и не обрабатывалась, а обрабатывалась только меньшая часть суши и поэтому выборка была мала, при построении доверительных интервалов был использован известный критерий Стьюдента. Переходя от Θ к $Z = (\bar{\Theta} - M/S)$ (где S^2 — выборочная дисперсия), рас-

пределенной по закону Стьюдента: $dF = \frac{\kappa dZ}{(1+Z^2)^{N/2}}$, при заданном коэффициенте доверия $1-\alpha$ можно найти Z_0 и Z_1 такие, что

$$\int_{-\infty}^{-Z_1} dF = \int_{Z_0}^{\infty} dF = \frac{\alpha}{2}.$$

Отсюда $P = (-Z_1 \leq Z \leq Z_0) = 1-\alpha$, которое эквивалентно соотношению $P(\bar{\Theta} - SZ_0 \leq M \leq \bar{\Theta} + SZ_1) = 1-\alpha$.

Для малых выборок границы доверительных интервалов строились на $\bar{\Theta} - SZ_0$ и $\bar{\Theta} + SZ_1$ с коэффициентом доверия $1-\alpha$. На рис. 1 представлены для иллюстрации месячные математические ожидания и селективные доверительные интервалы на уровне $M \pm \sigma$ балльности и высоты верхней границы облачности для территорий США (рис. 1, а, в) и Китая (рис. 1, б, г) в целом. Гистограммы математических ожиданий при переходе от квадрата к квадрату сглажены по формулам

$$M_i = \frac{1}{4}M_{i-1} + \frac{1}{2}M_{i+1}.$$

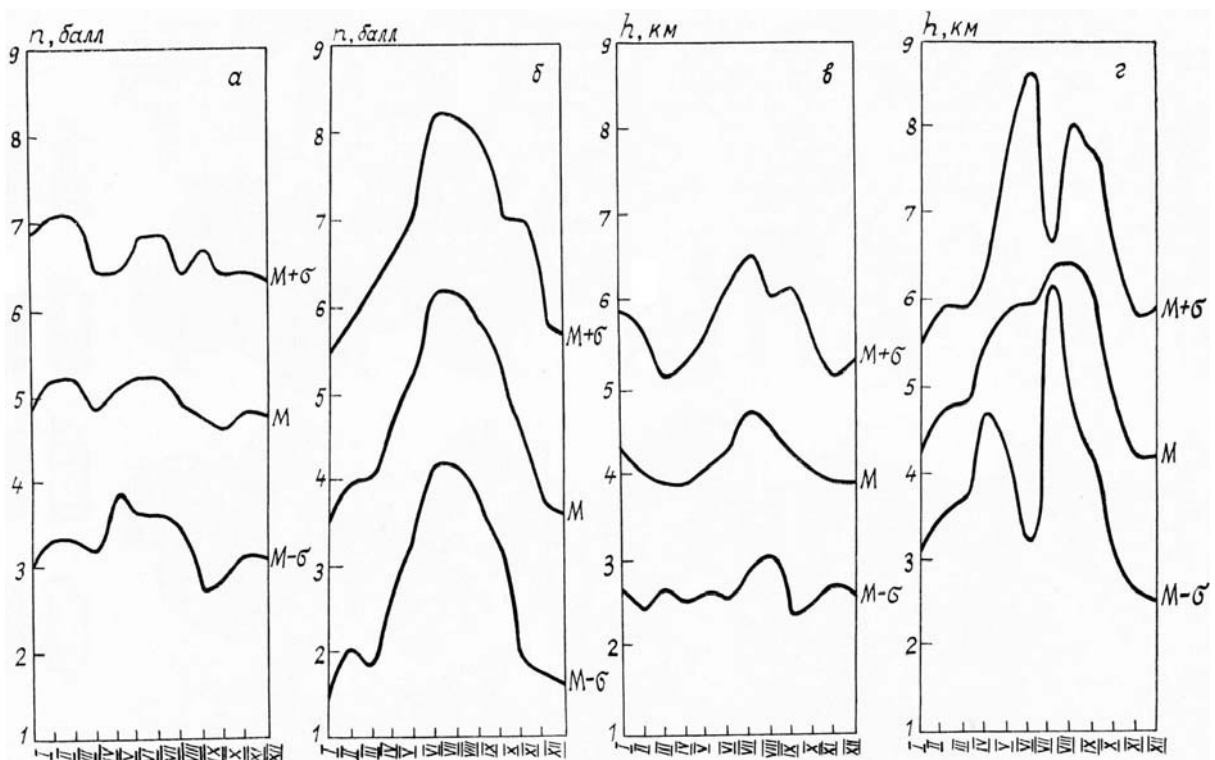


Рис. 1. Доверительные интервалы балльности (n) и высот (h) верхней границы облачности над территориями США (а, в) и Китая (б, г). Римскими цифрами обозначены месяцы

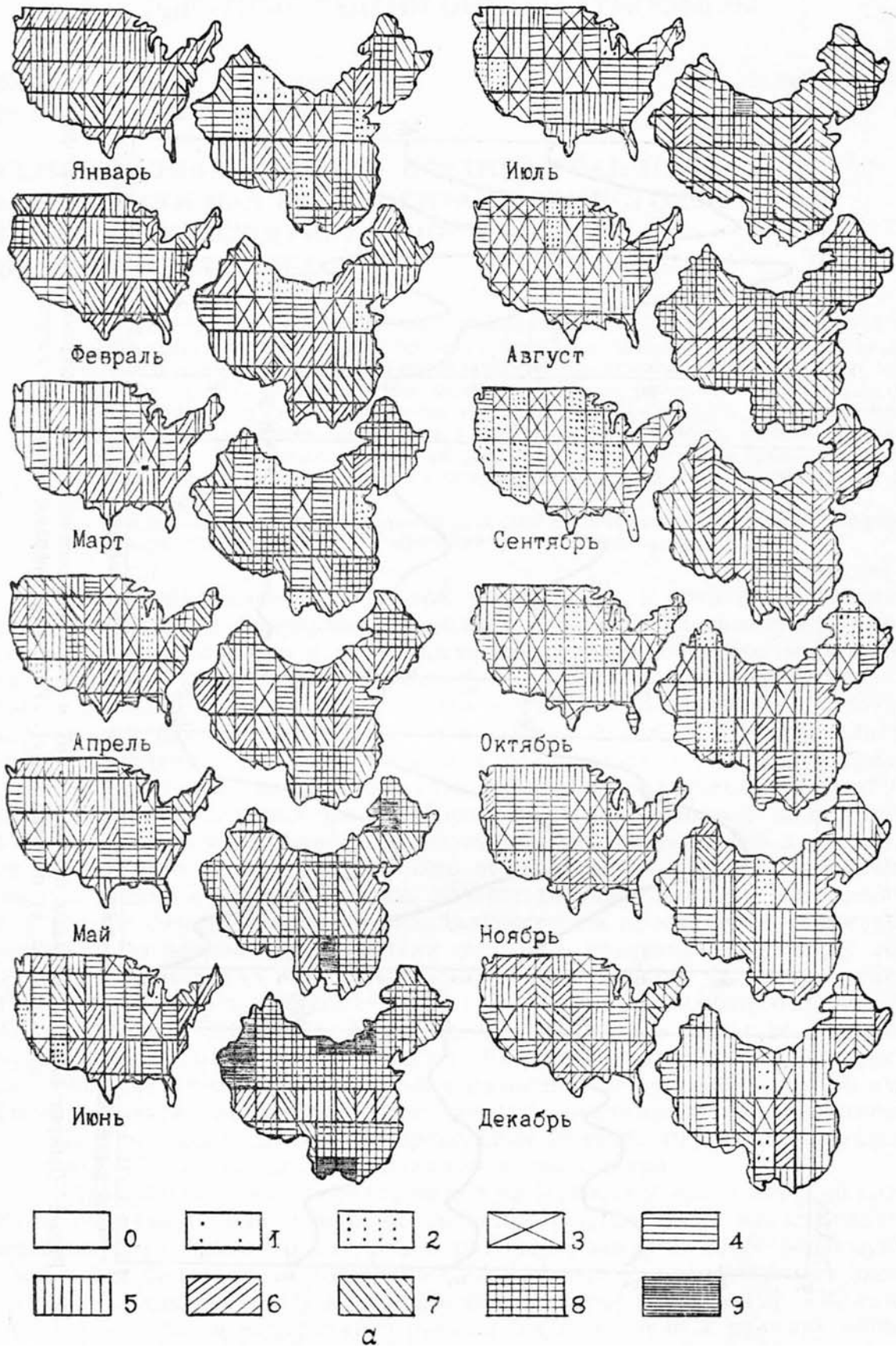
На картах США (левые столбцы рис. 2, а, б) и Китая (правые столбцы рис. 2 а, б) представлены для наглядности средние значения балльности (2, а) и высот верхней границы облачности (2, б) с разбиением $5 \times 5^\circ$ геодезической меркаторовской системы координат помесечно. Обозначения на рис. 2, а соответствуют балльности облачности от 0 до 9 баллов с дискретом 1 балл, на рис. 2, б — высотам верхней границы облачности от 0 (безоблачно) до 12 км с дискретом 1 км.

На рис. 3 представлены месячные плотности распределения балльности и высоты верхней границы облачности над территорией США и Китая соответственно, полученные описанным выше путем.

Если считать, что кроме месячного и годового хода есть и более длинные периоды, то следует предположить, что сделанные оценки моментов имеют смещение. Но несмотря на это по моментам можно получить математическое выражение распределений, которое удовлетворительно описывает выборку наблюдений. Сравнение между разными законами [4, 5, 6]: нормальным, логнормальным, показательным, γ -распределением, β -распределением, Вейбулла, Фишера, Типпета (1-й и 2-й тип), проведено на основе критерия Колмогорова:

$$\max / F_{\text{теор}}(\Theta) - F_{\text{эмп}}(\Theta) / = \Delta_m,$$

где $F_{\text{теор}}(\Theta)$, $F_{\text{эмп}}(\Theta)$ — теоретическое и эмпирическое распределения соответственно; Δm — мера расхождения.



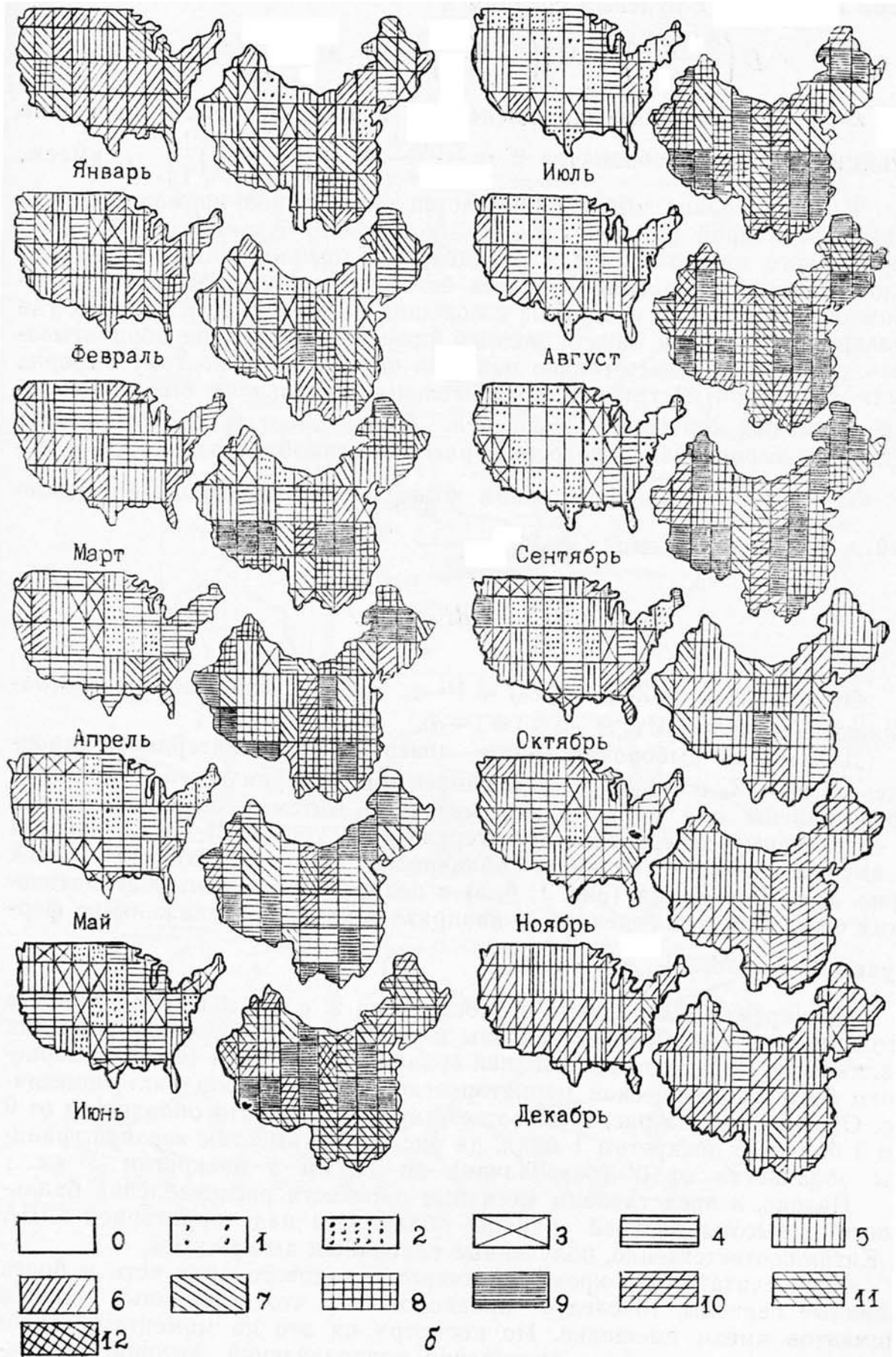


Рис. 2. Балльность облачности (а); высота верхней границы облачности (б)

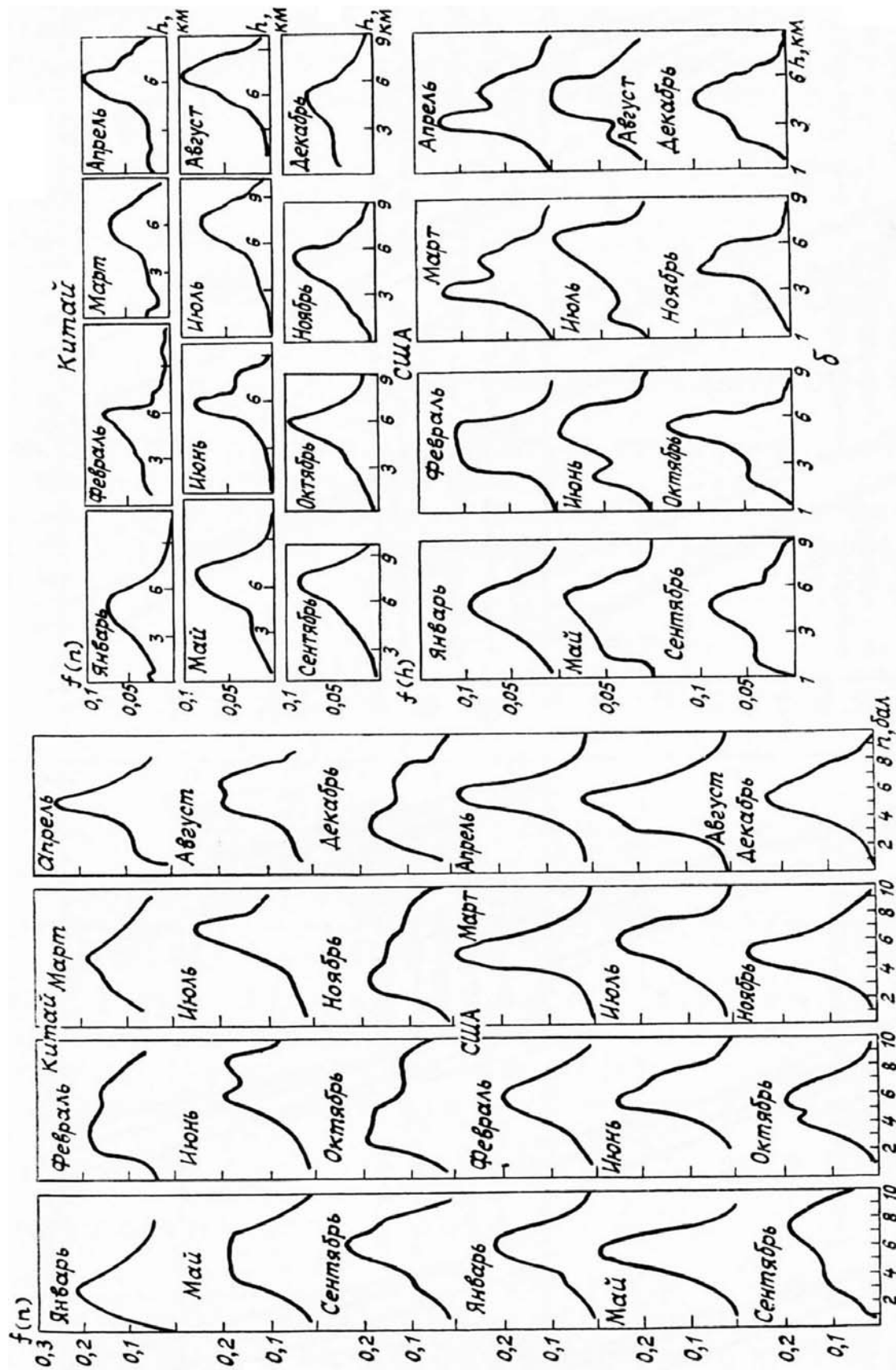


Рис. 3. Месячные плотности распределения облачности (а) и высот верхней границы облачности (б)

Описание эмпирических распределений нормальным законом привело на уровне 95%-й значимости к существенному расхождению: $\Delta m = 17\%$. Кроме того, недостатком такой аппроксимации является симметричная форма, а также бесконечный интервал сосредоточения распределения: в зависимости от величины дисперсии такая аппроксимация может придать слишком большой вес нереальным значениям искомых величин.

Остановимся подробнее на анализе возможности применения бета-распределения.

Параметры β -распределения:

$$p = \frac{m'q}{1-m'}; \quad q = \frac{1-m'}{(\sigma')^2} [m'(1-m') - (\sigma')^2],$$

где $m' = \frac{m}{x_{\max}}$, $\sigma' = \frac{\sigma}{x_{\max}}$ – нормированные эмпирические математическое ожидание и среднеквадратическое отклонение; x_{\max} – максимальное значение случайной величины x .

Далее по [2] и [7] рассчитывались начальные a_i и центральные m_i ($i = 1, 2, 3, 4$) моменты соответствующих порядков (до 4-го включительно):

$$\begin{aligned} a_2 &= \frac{\Gamma(p+2)\Gamma(p+q)}{\Gamma(p)\Gamma(p+q+2)}; \quad m_2 = a_2 - a_1^2; \\ a_3 &= \frac{\Gamma(p+3)\Gamma(p+q)}{\Gamma(p)\Gamma(p+q+3)}; \quad m_3 = a_3 - 3a_1a_2 + 2a_1^3; \\ a_4 &= \frac{\Gamma(p+4)\Gamma(p+q)}{\Gamma(p)\Gamma(p+q+4)}; \quad m_4 = a_4 - 4a_1a_3 + 6a_1^2a_2 - 3a_1^4, \end{aligned}$$

где $\Gamma(p)$, $\Gamma(q)$, $\Gamma(p+q)$, $\Gamma(p+i)$, $\Gamma(p+q+i)$ – значения Γ -функции в соответствующих точках. При $x > 2$ использовалось следующее свойство Γ -функции: $\Gamma(x+1) = x \cdot \Gamma(x)$.

При больших значениях x ($x > 10$) применялась формула Стирлинга $\Gamma(x+1) \approx \left(\frac{x}{e}\right)^x \cdot \sqrt{2\pi x}$ (погрешность не более 1%).

Критерий применимости β -распределения [2]:

$$k = \frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)} < 0; \quad \text{где } \beta_1 = \frac{m_3^2}{m_2^3}; \quad \beta_2 = \frac{m_4}{m_2^2};$$

Расчет показал, что все $k < 0$. Следовательно, применимо распределение Пирсона I типа (β -распределение):

$$F(x, p, q) = \begin{cases} 0 & \text{при } x \leq 0 \\ \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x x^{p-1} (1-x)^{q-1} dx & \text{при } 0 < x < 1; \\ 1 & \text{при } x \geq 1 \end{cases}$$

Для тех же эмпирических зависимостей балльности и высоты верхней границы облачности на логарифмически-вероятностной координатной сетке были построены графики теоретических функций логнормального распределения (рис. 4).

По принципу минимизации максимальной разности между эмпирическими и соответствующими теоретическими значениями Δm :

$$F(x, x_0, \sigma_0) = \begin{cases} 0 & \text{при } x \leq 0 \\ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-t^2/2} dt & \text{при } x > 0, \end{cases}$$

где

$$u = \frac{\lg x - \lg x_0}{\lg \sigma_0}; \quad \kappa = \frac{1}{\lg e} = 2,303.$$

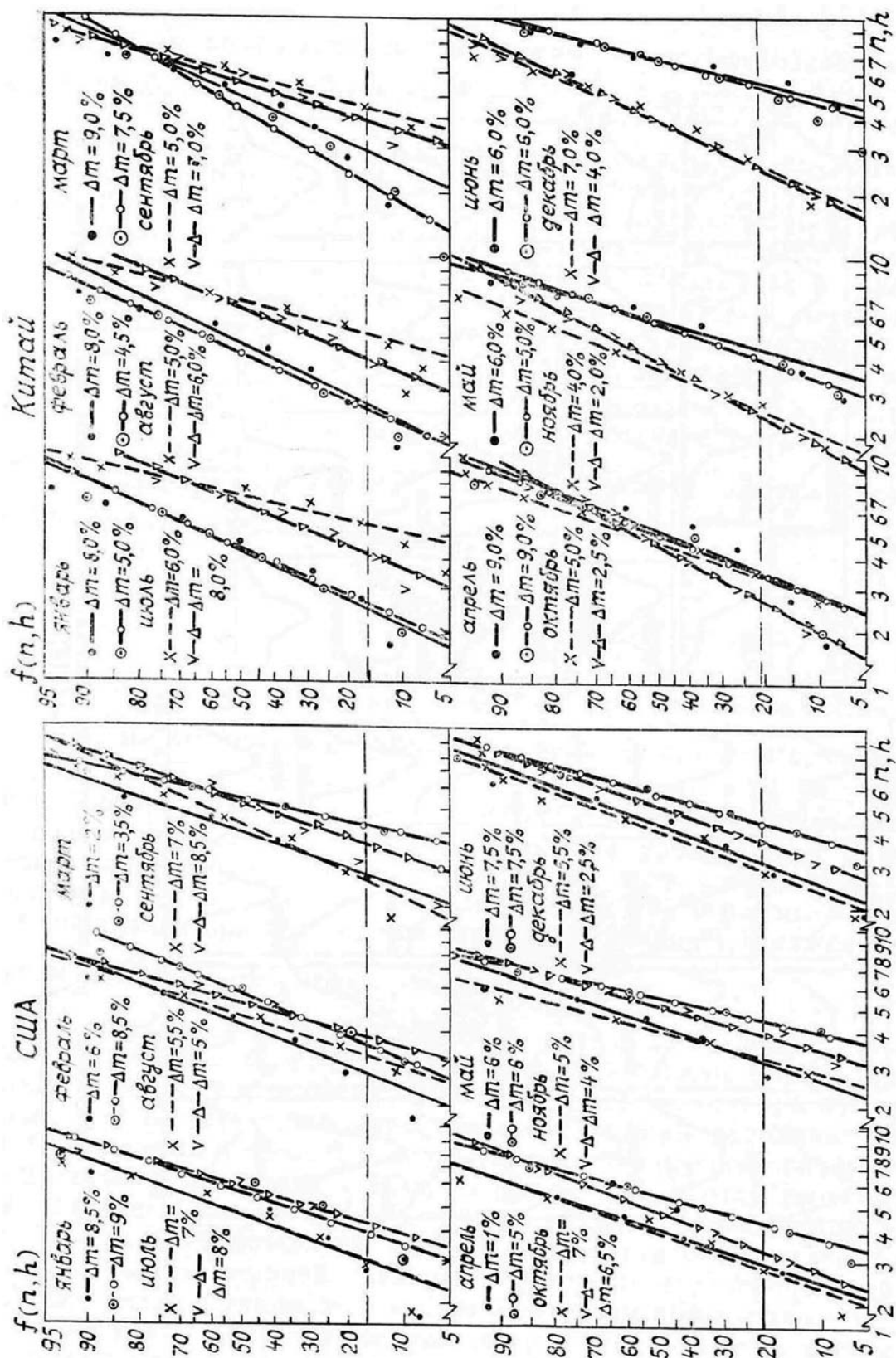


Рис. 4. Функции распределения бальности (— Δ —, — \circ —) и высот верхней границы облачности (— — — — —)

Величина $m_y = \lg x_0$ представляет собой математическое ожидание случайной величины (СВ) $Y = \lg x$, а $\sigma_y = \lg \sigma_0$ — ее среднеквадратическое отклонение. Значения x_0 и σ_0 определяются графически [8]: $x_0 = x_{50\%}$ (медиана СВ x) $\sigma = x_{50\%}/x_{15,9\%}$.

Математическое ожидание m_x и среднеквадратическое отклонение σ_x случайной величины x равны [9]:

$$m_x = x_0 e^{\frac{\kappa^2 \lg^2 \sigma_0}{2}} = x_0 e^{2,651 \lg^2 \sigma_0}; \quad \sigma_x = m_x \sqrt{\left(\frac{m_x}{x_0}\right)^2 - 1}.$$

Анализ полученных данных показывает, что максимальная величина Δm для β -распределения составляет 13,8%, а для логнормального распределения — 9%. В данном случае несмотря на универсальность β -распределения логнормальное распределение лучше описывает эмпирические зависимости. По Р. Фишеру [4] были найдены доверительные интервалы для проверки близости коэффициентов корреляции нулю. Для двух выборок из нормального распределения с объемом выборки N , коэффициентом корреляции совокупности ρ и выборочным коэффициентом корреляции r , величина $\delta = \frac{1}{2} \log \frac{1+r}{1-r}$ распределена уже при небольших r приблизительно нормально со средним $\bar{\delta} = \frac{1}{2} \log \frac{1+\rho}{1-\rho} + \frac{\rho}{2(N-1)}$ и дисперсией $1/(N-3)$. Используя распределение δ были найдены 95%-е интервалы для r в предположении, что $\rho = 0$. Таким образом, выделялись некоррелированные зоны.

Представляло интерес найти сезонные регрессии для зон с большими коэффициентами корреляции. Была сделана попытка подбора множественной линейной регрессии сезонной балльности методом наименьших квадратов. Для модели $y = a + b_1 x + b_2 x^2 + \dots + \epsilon_1$, где y — зависимая переменная (в нашем случае балльность), x — месяцы, b_i — коэффициенты регрессии, a — свободный член, ϵ — ошибка измерения с нулевым средним, находились остатки $(y - \hat{y})^2$, где $\hat{y} = \hat{a} + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \dots + \hat{b}_p x_p$, где \hat{b}_i — оценка коэффициента регрессии; \hat{y} — предсказанное значение. Так как дисперсия балльности менялась от реализации к реализации, был применен взвешенный метод наименьших квадратов, причем вес реализации вводился как величина, обратно пропорциональная дисперсии. Серийная корреляция остатков вычислялась как

$$\sum_{j=2}^N \frac{(W_j W_{j-1})^{1/2} (y_j - \hat{y}_j) (y_{j-1} - \hat{y}_{j-1})}{\left\{ \sum_{j=2}^N W_j (y_j - \hat{y}_j)^2 \sum_{j=2}^N W_{j-1} (y_{j-1} - \hat{y}_{j-1})^2 \right\}^{1/2}},$$

где W_j — вес j -й реализации. Принцип минимизации сумм квадратов регрессии $\sum (y - \hat{y})^2$ выявил наилучшую сходимость криволинейной (или полиномиальной) зависимости y от x . Подбор коэффициентов по *Biometrike Tables* [5] привел к хорошей подгонке ортогональной полиномиальной регрессии третьей и четвертой степени $y = a + \hat{b}_1 x + \hat{b}_2 x^2 + \hat{b}_3 x^3 + \hat{b}_4 x^4$. Значения a и b_i для балльности облачности на 87%-й площади Китая в качестве иллюстрации представлены ниже. Средняя величина квадрата ошибки составляет 0,0053626.

a	b_1	b_2	b_3	b_4
5,4249201	-1,1451431	0,2431453	-0,0091621	0,0000001

Выводы

1. Разработаны алгоритмы и пакет программ для вычисления математических ожиданий, дисперсий, корреляционных матриц, регрессий, плотностей и законов распределений характеристик облачности по данным международного эксперимента ПГЭП.
2. Построены сезонно-географические распределения верхней границы облачности над континентами земного шара. Эти данные могут быть использованы для расшифровки спутниковой информации при дистанционном зондировании Земли в оптическом диапазоне.
3. Установлено, что распределение балльности и высоты верхней границы облачности может быть аппроксимировано логнормальным законом. Аппроксимация проведена по критерию Колмогорова на уровне значимости 0,95.

1. FGGE Data Catalogue, World Data CENTER-B, М., 1982.

2. Кэндалл М., Стьюарт А. Теория распределений. М.: Наука, 1966. 587 с.
3. Кэндалл М., Стьюарт А. Статистические выводы и связи. М.: Мир, 1973. 899 с.
4. Авасте О.А., Кярнер О.Ю., Ламден К.С., Шифрин К.С. // Оптика океана и атмосферы. Статистические характеристики облачности и суммарной радиации над различными акваториями мирового океана. М., 1981. С. 194–229.
5. Falles L.W. // J. Geophys. Res. 1974. V. 79. № 9. P. 1261–1264.
6. Глобальное поле облачности / Под ред. Л.Т. Матвеева. Л.: Гидрометеоиздат, 1986. 278 с.
7. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. М.: Наука, 1984. 831 с.
8. Райст П. Аэрозоли. Введение в теорию. М.: Мир, 1987. 278 с.
9. Абезгауз Г.Т., Тронь А.П., Коненкин Ю.П., Коровина И.А. Справочник по вероятностным расчетам. М.: Воениздат, 1970. 407 с.
10. Крамер Г. Математические методы статистики. М.: Мир, 1975. 648 с.

Всесоюзный научно-исследовательский институт оптико-физических измерений
Госстандарта СССР

Поступила в редакцию
30 октября 1989 г.

O.I. Aldoshina, V.V. Bacherikov, E.E. Limar, V.A. Fabrikov. Regression Model for Cloud Top Heights Distribution over the Continents Based on the Data of Fgge International Experiment.

The algorithm and computer programs for calculating the statistical momenta, density and distribution functions as well as the correlation matrices and regression lines for statistical analysis of the FGGE experimental data on the cloudiness characteristics are developed. Season-geographic distributions of expectation value for the maximum height of cloud tops and total cloud cover over the Earth's continents are constructed based on the data of the First GARP Global Experiment (FGGE). Maximum heights of cloud tops distribution and distribution of total cloud cover over the continents can be approximated by lognormal law. The approximation of the distributions was made following the Kolmogorov's criterion at the level of significance of 0,95.

The results of these calculations can be used for interpreting the data of spaceborne sensing of the Earth in the optical wavelength range.