

ГОСУДАРСТВЕННЫЙ КОМИТЕТ СССР ПО ГИДРОМЕТЕОРОЛОГИИ
И КОНТРОЛЮ ПРИРОДНОЙ СРЕДЫ

ТРУДЫ
ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ
ГЛАВНОЙ ГЕОФИЗИЧЕСКОЙ ОБСЕРВАТОРИИ
им. А. И. ВОЕЙКОВА

Выпуск

437

АНАЛИЗ
ГИДРОМЕТЕОРОЛОГИЧЕСКИХ
ПРОЦЕССОВ
С ПОМОЩЬЮ ЭВМ

Под редакцией
канд. физ.-мат. наук Б. М. ИЛЬИНА,
д-ра физ.-мат. наук И. И. ПОЛЯКА



ЛЕНИНГРАД ГИДРОМЕТЕОИЗДАТ 1981

Сборник посвящен статистическому анализу гидрометеорологических наблюдений с помощью ЭВМ. Приведены алгоритмы построения моделей авторегрессии и дано описание соответствующего математического обеспечения, проводится анализ результатов сглаживания коррелированных наблюдений с помощью цифровых фильтров, исследуется корреляционная и спектральная структура длиннопериодных колебаний метеоэлементов и т. д.

Сборник рассчитан на научных работников, аспирантов и студентов метеорологических специальностей.

The publication deals with the statistical analysis of hydrometeorological observations with computer. The algorithms of constructing autoregression models are given as well as the description of proper software, the results of smoothing the correlated observations are analyzed with the aid of digital filters, the correlation and spectral structure of meteorological long-period fluctuations is studied.

The publication is meant for researchers, post-graduates and students of meteorological specialities.

343484

Ленинградский
Гидрометеорологический ин-т
БИБЛИОТЕКА
Л-д 193196, Малоохтенский пр., 98

Г. В. Груза, И. И. Поляк, В. А. Шахмейстер

К АНАЛИЗУ СТАТИСТИЧЕСКОЙ СТРУКТУРЫ ШИРОТНО-ВРЕМЕННЫХ ПОЛЕЙ ОТ $\frac{500}{1000}$ И ЗОНАЛЬНОЙ СОСТАВЛЯЮЩЕЙ СКОРОСТИ ГЕОСТРОФИЧЕСКОГО ВЕТРА ПОВЕРХНОСТИ 500 мбар СЕВЕРНОГО ПОЛУШАРИЯ

Спектральный и корреляционный анализ временных рядов и полей находит широкое применение в метеорологии [1, 4]. Получено много интересных результатов, характеризующих колебания самых различных масштабов.

В настоящей статье будут рассмотрены широтно-временные поля среднемесячных значений двух элементов: относительной топографии $OT_{\frac{500}{1000}}$ и скорости геострофического ветра поверхности 500 мбар, которые найдены осреднением по различным пятиградусным широтным зонам данных в узлах регулярной сетки. Осредненные значения соответствуют широтам 30, 35, ..., 80° (11 узлов по широте) северного полушария и периоду 1948—1975 гг. по времени (336 узлов по времени). Значения $OT_{\frac{500}{1000}}$ измерены в геопотенциальных метрах, а скорость ветра — в м/с.

При получении рассматриваемых среднемесячных значений объем исходной информации значительно варьировал во времени и пространстве. Проследить изменение числа радиозондовых сообщений за весь рассматриваемый период

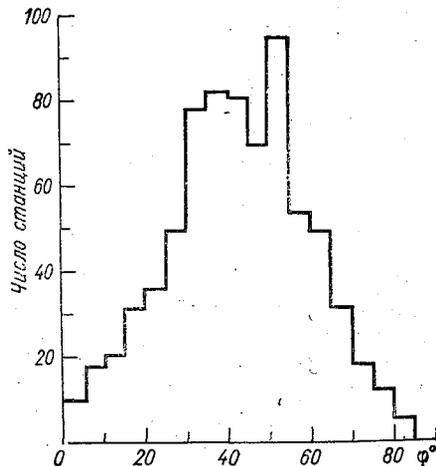


Рис. 1. Количество аэрологических станций в пятиградусных широтных зонах северного полушария за период 1968—1973 гг.

Общее количество станций 734.

Годовой ход норм

Широта, °	I	II	III	IV	V	От 500 1000										
						VI	VII	VIII	IX	X	XI	XII				
80	496,80	497,04	497,96	504,90	517,53	532,05	539,07	536,07	524,57	511,38	503,53	499,12				
75	499,99	500,20	501,50	508,49	520,98	534,70	542,33	539,12	528,30	515,71	506,81	502,49				
70	503,10	503,17	505,62	513,25	525,46	538,82	546,54	543,39	532,80	520,38	510,48	505,61				
65	506,49	506,88	510,54	518,96	530,43	543,07	550,36	547,48	537,33	525,21	514,56	509,25				
60	511,47	511,80	515,93	524,49	534,83	546,19	552,93	550,76	541,66	530,51	520,17	514,48				
55	516,57	516,67	521,06	529,76	539,22	549,49	555,75	554,44	546,20	535,97	525,74	519,69				
50	521,76	521,69	526,35	535,53	544,81	554,25	560,31	559,57	552,03	542,33	531,87	525,15				
45	528,97	528,87	533,36	542,37	551,42	560,17	566,25	565,98	559,02	549,83	539,57	532,52				
40	537,18	537,22	541,40	549,43	557,81	565,94	571,93	572,04	565,88	557,19	547,47	540,66				
35	546,33	546,53	550,26	556,82	563,98	571,37	576,27	576,47	571,71	564,08	555,79	549,57				
30	555,97	556,16	559,08	563,63	569,24	575,25	578,36	578,69	575,66	569,97	563,68	558,61				
Зональная составляющая скорости геострофического ветра на уровне 500 мбар																
80	2,82	3,27	2,63	2,16	2,76	2,39	4,49	3,33	3,45	3,60	1,69	3,21				
75	3,71	3,95	4,29	4,17	3,89	3,75	4,87	4,48	4,73	5,16	3,25	3,92				
70	4,48	4,57	5,76	6,00	4,80	4,74	4,81	5,14	5,71	6,36	4,68	4,54				
65	5,45	5,04	6,41	6,87	4,87	4,39	3,60	4,48	5,99	7,26	6,29	5,74				
60	7,36	6,50	7,30	7,69	5,77	5,09	4,41	5,73	7,71	9,50	8,71	7,94				
55	9,58	8,57	8,46	8,92	7,67	6,75	6,81	8,41	10,22	12,24	11,65	10,58				
50	12,78	12,03	10,76	10,93	10,38	9,22	9,80	11,04	12,72	14,55	14,80	13,88				
45	15,74	15,43	13,92	12,82	11,94	10,62	10,94	11,75	13,27	14,77	16,13	16,44				
40	17,94	18,06	17,06	14,70	12,71	10,93	9,69	9,79	11,79	13,53	16,38	17,89				
35	19,32	19,72	19,17	15,56	12,44	10,02	6,66	6,33	8,95	11,71	15,79	18,39				
30	19,45	20,69	21,25	16,18	11,75	7,76	2,77	1,99	4,80	9,41	14,17	17,93				

Годовой ход стандартов

Широта, °	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
80	3,79	2,55	3,56	1,86	2,53	1,86	1,47	1,37	1,76	2,20	3,38	3,02
75	2,47	1,68	2,51	1,90	1,83	1,45	1,16	1,25	1,21	1,75	2,69	2,59
70	1,71	1,62	2,10	1,68	1,17	1,26	1,28	1,16	1,25	1,60	1,86	2,22
65	1,63	1,90	2,06	1,38	1,01	1,01	1,15	1,01	1,38	1,58	1,39	1,85
60	1,92	1,98	1,99	1,45	1,06	1,00	0,91	0,92	1,23	1,46	1,25	1,43
55	2,04	1,77	1,89	1,38	1,22	0,87	0,94	0,93	0,98	1,39	1,39	1,30
50	2,01	1,66	1,73	1,36	1,19	1,02	1,17	0,98	0,97	1,37	1,39	1,34
45	1,57	1,60	1,42	1,23	1,14	1,22	1,14	1,11	0,98	1,11	1,23	1,07
40	1,24	1,38	1,00	1,10	1,08	1,13	1,16	1,22	1,04	0,80	0,91	0,87
35	1,02	1,15	0,94	1,07	1,03	1,03	1,13	1,21	1,05	0,87	0,74	0,86
30	1,05	1,28	1,22	1,28	0,94	0,78	0,85	0,95	0,93	1,06	0,97	1,06
Зональная составляющая скорости геострофического ветра на уровне 500 мбар												
80	2,45	3,40	3,63	2,68	2,47	3,26	2,88	2,90	2,25	2,15	2,52	2,85
75	2,30	2,81	2,90	1,84	1,63	2,10	1,95	2,01	1,74	1,30	1,88	2,15
70	2,32	2,32	2,33	1,46	1,68	1,63	1,52	1,51	1,30	1,30	2,02	1,77
65	2,25	1,79	2,18	1,66	1,72	1,25	1,65	1,34	1,17	1,58	2,21	2,00
60	2,16	1,69	2,16	1,52	1,67	0,99	1,45	1,26	1,18	1,71	1,82	2,11
55	2,47	2,22	1,98	1,72	1,78	1,41	1,35	1,27	1,33	1,55	1,60	2,03
50	2,47	2,69	1,85	1,83	1,92	1,13	0,91	1,12	1,34	1,07	1,37	1,85
45	2,38	2,15	1,92	1,66	1,82	1,03	1,17	0,97	1,01	1,68	1,58	1,73
40	2,40	1,68	1,91	1,55	1,43	1,03	1,20	1,04	0,98	1,80	1,76	1,94
35	2,08	1,57	1,80	1,62	1,25	0,86	0,61	0,90	0,90	1,24	1,46	1,66
30	2,56	2,72	2,38	2,02	1,46	1,00	1,08	1,25	1,43	1,49	1,79	2,39

достаточно трудно. Приведем лишь рисунок (рис. 1), заимствованный из [5], который показывает количество действующих аэрологических станций на отдельных широтах в период 1968—1973 гг. Рисунок 1 дает хотя бы приближенное представление о надежности результатов, полученных в данной статье.

Применяемые методы анализа подробно рассмотрены в работах [2—4]. Методика и последовательность вычислений в основном соответствуют статье [1].

Таблица 3

Значения корреляционной функции для запаздываний в один и два месяца и оценка 95%-ной статистической значимости отклонения от белого шума спектральных функций

Широта, °	OT ₅₀₀ 1000			Зональная составляющая скорости геострофического ветра на уровне 500 мбар		
	k(1)	k(2)	выход за доверительные интервалы	k(1)	k(2)	выход за доверительные интервалы
80	0,178	0,057	—	0,174	—0,038	—
75	0,211	0,044	—	0,151	—0,025	—
70	0,302	0,124	+	0,116	0,073	—
65	0,316	0,167	+	0,230	0,147	—
60	0,312	0,167	+	0,215	0,164	—
55	0,408	0,218	+	0,079	0,099	—
50	0,448	0,261	+	0,062	0,035	—
45	0,384	0,302	+	0,279	0,090	+
40	0,364	0,278	+	0,352	0,236	+
35	0,480	0,341	+	0,239	0,160	+
30	0,602	0,507	+	0,291	0,093	+
Среднее	0,364	0,224		0,198	0,094	

1. В табл. 1 приведены оценки среднемесячных норм различных широтных зон. Такие оценки, вычисленные с помощью значительно меньшего количества наблюдений, изучались в ряде работ, например в [6].

В табл. 2 даны оценки стандартов рассматриваемых данных. Так же как и нормы, стандарты имеют годовой ход. Его минимум в основном соответствует теплomu времени года, а максимум — холодному.

В низких широтах (35...40°) амплитуды годового хода стандартов OT₅₀₀¹⁰⁰⁰ невелики и отмеченная закономерность иногда нарушается, по-видимому, из-за недостатка данных в этих районах. По этой же причине на широте 80° стандарты скорости ветра не имеют определенной закономерности. Высоким широтам со-

ответствуют несколько большие стандарты, чем низким (это заметно особенно в летние месяцы), т. е., чем ниже температура воздуха, тем больше изменчивость рассматриваемых метеоэлементов.

2. Проведем спектральный и корреляционный анализ изучаемых данных. Прежде всего отфильтруем годовой ход путем нахождения отклонений каждого среднемесячного значения от соответствующей оценки среднего и деления этих отклонений на оценки стандартов. Полученные таким образом стандартные аномалии для каждой широтной зоны в отдельности представляют собой временные ряды, которые можно анализировать методами теории стационарных случайных процессов.

Не приводя графиков спектральных функций, укажем (см. последние столбцы табл. 3), что отклонения рассчитанных значений от теоретической спектральной функции белого шума являются статистически значимыми с 95 %-ным уровнем у скорости ветра в четырех и у OT_{1000}^{500} в девяти случаях.

В табл. 3 даны оценки корреляционных функций для запаздываний в один и два месяца. Значения k (1) ветра варьируют в пределах 0,06...0,36, OT_{1000}^{500} — 0,17...0,60.

3. Рассмотрим далее широтно-временные поля стандартных аномалий анализируемых видов наблюдений и найдем их двумерные спектральные и корреляционные характеристики. Центральные части двумерных оценок спектров даны на рис. 2, оценок корреляционных функций — на рис. 3. Участки поля, не приводимые на графиках, близки к нулю и не содержат каких-либо существенных особенностей.

Не приводя оценок периодограмм, укажем, что основная мощность соответствует изменениям во времени. На первую гармонику вдоль оси ординат приходится 0,52 % дисперсии изменчивости скорости ветра, 2,42 % — OT_{1000}^{500} . На каждой из периодограмм имеется несколько небольших максимумов, в основном соответствующих точкам, близким к началу координат.

При сглаживании отмеченные особенности периодограмм формируют максимумы оценок спектров в окрестности начала координат (см. рис. 2). Для ветра они незначительны, максимум OT_{1000}^{500} более заметен. Однако ввиду очень редкой сетки точек сглаживание, по-видимому, несколько исказило реальную картину распределения дисперсии по частотам вследствие расположения мощности первых гармоник.

Центральные части двумерных корреляционных функций даны на рис. 3. Корреляция от одной широтной зоны к другой стандартных аномалий средних значений скорости ветра — 0,46, OT_{1000}^{500} — 0,53. Корреляция этих же величин от месяца к месяцу соответственно равна 0,21 и 0,37.

4. Для временных рядов среднемесячных величин отдельных широтных зон наряду с одномерными спектрами можно получать двумерные спектральные и корреляционные характеристики.

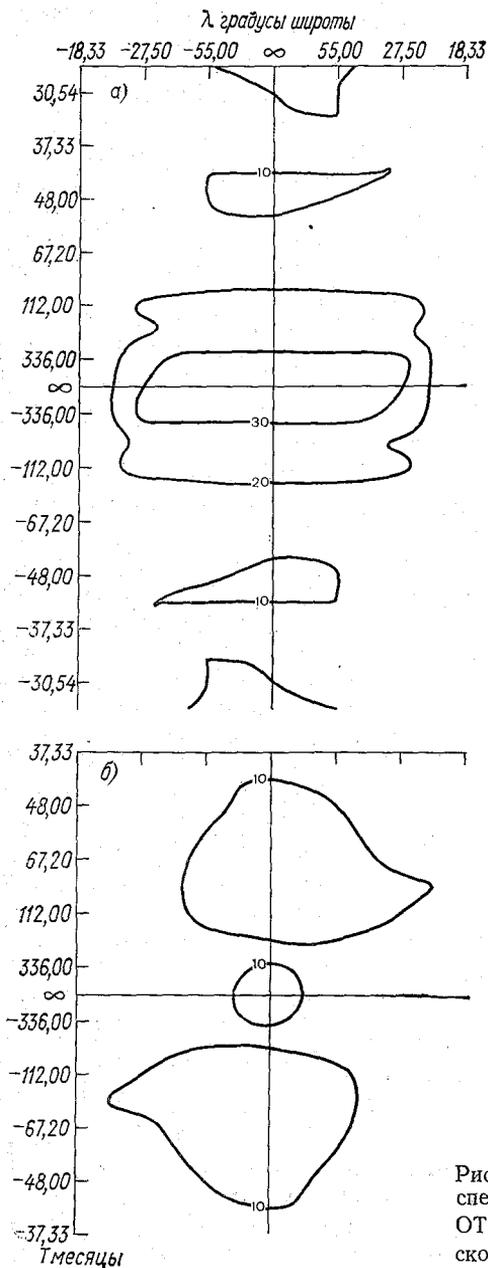


Рис. 2. Центральные части двумерных спектров широтно-временных полей OT_{1000}^{500} (а) и зональной составляющей скорости геострофического ветра (б) поверхности 500 мбар.

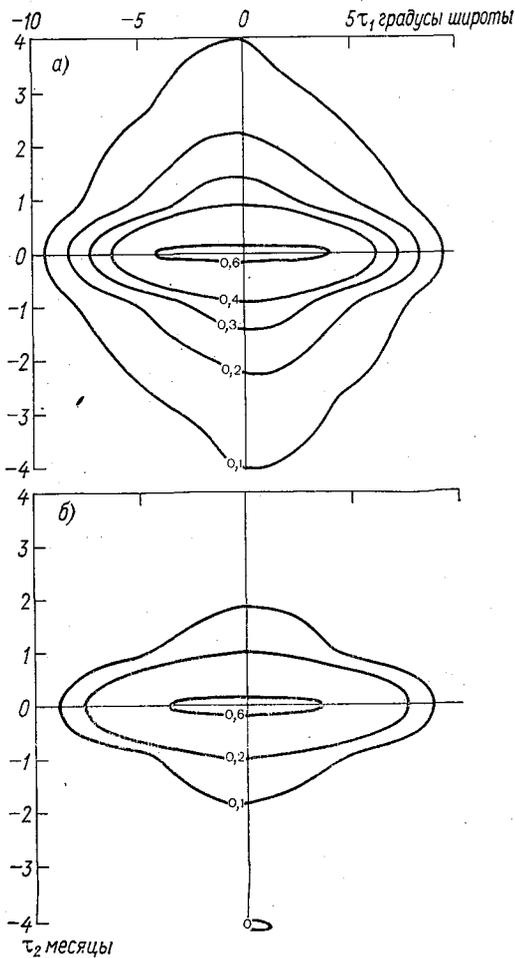


Рис. 3. Центральные части двумерных корреляционных функций широтно-временных полей OT_{1000}^{500} (а) и зональной составляющей скорости геострофического ветра (б) поверхности 500 мбар.

В этом случае каждый такой временной ряд рассматривается как двумерное поле, где изменения вдоль оси абсцисс идут с шагом в один месяц, а по оси ординат — с шагом в один год. Анализируя такие поля, мы как бы разделяем внутригодовые и межгодовые колебания и имеем возможность исследовать их взаимодействие. Для каждого из рассматриваемых элементов приведем результаты, полученные с помощью анализа временных рядов, соответствующих одной из широт, например 55° .

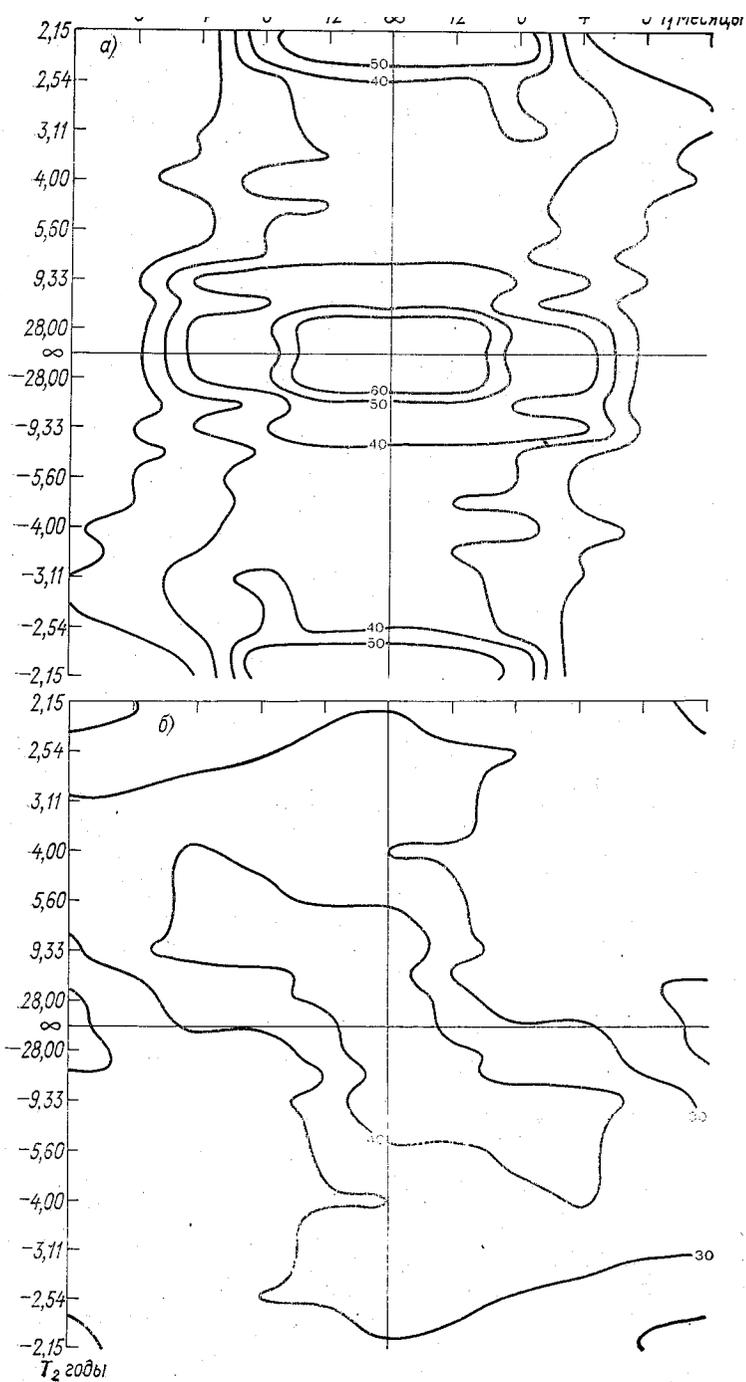


Рис. 4. Центральные части двумерных спектров временных рядов среднемесячных значений от 1000^{500} (а) и зональной составляющей скорости геострофического ветра (б) поверхности 500 мбар для широты 55° .

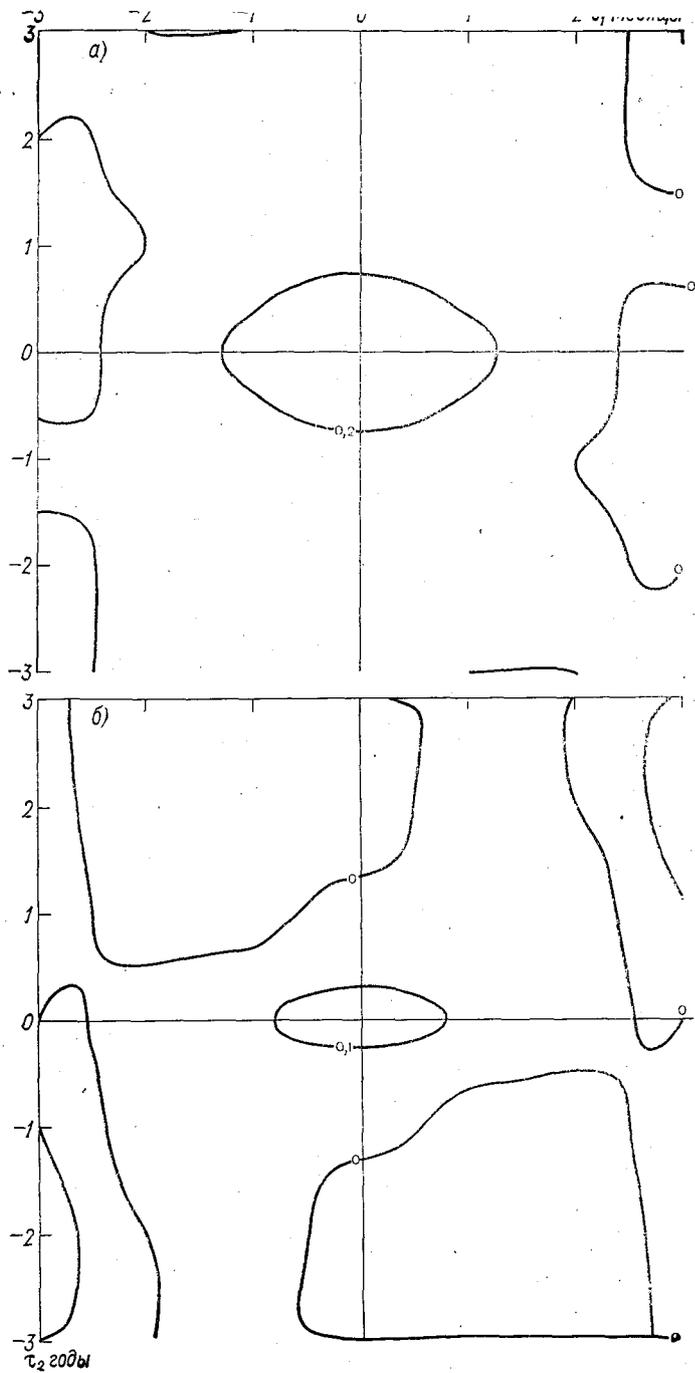


Рис. 5. Центральные части двумерных корреляционных функций временных рядов среднемесячных значений OT_{1000}^{500} (а) и зональной составляющей скорости геострофического ветра (б) поверхности 500 мбар для широты 55° .

На рис. 4 даны оценки спектральных плотностей, на рис. 5 — корреляционных функций.

Спектр OT_{1000}^{500} имеет размытый максимум с центром в начале координат; распределение дисперсии скорости ветра по частотной плоскости сравнительно равномерно. Корреляционная функция скорости ветра показывает, что для широты 55° статистическая зависимость данных близка к нулю как от года к году, так и от месяца к месяцу.

Корреляционная функция OT_{1000}^{500} показывает наличие коррелированности данных соседних месяцев около 0,3 и практическое отсутствие зависимости с шагом в один год и более.

Таблица 4

Оценки коэффициентов корреляции для запаздываний в один месяц и в один год при двумерном анализе временных рядов OT_{1000}^{500} и зональной составляющей скорости геострофического ветра

Широта, °	Коэффициент корреляции			
	OT_{1000}^{500}		зональная составляющая скорости геострофического ветра на уровне 500 мбар	
	←месяц	год	месяц	год
80	0,11	-0,02	0,13	0,04
75	0,15	0,01	0,10	0,01
70	0,23	0,10	0,07	-0,08
65	0,23	0,09	0,18	0,00
60	0,21	0,03	0,19	0,03
55	0,28	0,04	0,09	0,07
50	0,31	0,09	0,07	-0,05
45	0,29	0,16	0,14	-0,08
40	0,29	0,22	0,27	0,00
35	0,36	0,24	0,23	0,08
30	0,45	0,30	0,22	0,12

Аналогичные оценки были получены для каждого из рассматриваемых временных рядов. Однако представление всех результатов заняло бы слишком много места. Поэтому в табл. 4 даны только оценки коэффициентов корреляции для запаздываний в один год и один месяц. По сравнению с данными табл. 3 оценки коэффициента корреляции при запаздывании в один месяц стали несколько меньше. Это обусловлено, по-видимому, в основном неоднородностью данных. Оценки табл. 4 более реальны, так как они ближе к результатам, получаемым при анализе временных рядов отдельных станций [4].

5. Проведем анализ тенденций крупномасштабных изменений данных в течение изучаемого периода времени. С этой целью

среднегодовые величины каждой из широтных зон аппроксимируются прямой линией $\bar{Y} + \beta_1 t$ и оценивается статистическая значимость средней скорости (β_1) изменения элемента.

Результаты, приводимые в табл. 5, показывают, что за рассматриваемый промежуток времени на всех широтах в среднем происходило небольшое уменьшение OT_{1000}^{500} ; в пяти случаях (из 11) уменьшение оказалось статистически значимым с 95 %-ным уровнем. Среднегодовые значения ветра за этот период в основном увеличивались, и для трех широтных зон (35, 40 и 45°) это увеличение превысило 95 %-ные доверительные границы.

Таблица 5

Оценки статистических характеристик линейного тренда для геопотенциала 500 мбар различных широтных зон за 28 лет (1948—1975)

Широта, °	σ_t	β_t	σ_{β_t}	$t(k-1)$	$\sigma_{\bar{Y}}$	σ_0	\bar{Y}
80	1,727	-0,074	0,0404	1,8	0,340	1,801	524,368
75	1,360	-0,052	0,0318	1,6	0,265	1,401	527,193
70	1,003	-0,046	0,0235	2,0	0,199	1,054	530,893
65	0,631	-0,027	0,0148	1,8	0,124	0,657	534,918
60	0,497	-0,018	0,0116	1,6	0,097	0,511	539,314
55	0,643	-0,013	0,0150	0,9	0,121	0,641	544,771
50	0,673	-0,063	0,0157	4,0	0,158	0,838	551,825
45	0,617	-0,081	0,0144	5,6	0,170	0,902	559,889
40	0,466	-0,035	0,0109	3,2	0,102	0,539	567,429
35	0,423	-0,014	0,0099	1,4	0,081	0,430	574,546
30	0,537	-0,031	0,0126	2,5	0,110	0,584	580,532

Более детальную характеристику изменений среднегодовых значений дают оценки параметра β_1 , найденные по данным трех различных временных интервалов (1948—1955, 1956—1965 и 1966—1975 гг.). Эти оценки приведены в табл. 6, где знаком «+» отмечены статистически значимые величины с 95 %-ным уровнем.

Как показывают результаты, уменьшение OT_{1000}^{500} особенно интенсивно происходило за десятилетие 1956—1965, где большинство оценок β_1 статистически значимо. За период наиболее надежных аэрологических данных (1966—1975 гг.) оценки β_1 только для трех широтных зон превышают 95 %-ный уровень в случае OT_{1000}^{500} . Большинство оценок β_1 скорости ветра как за десятилетие 1956—1965 гг., так и за 1966—1975 гг. не являются статистически значимыми.

Найденные одномерные и двумерные (пространственно-временные) статистические характеристики крупномасштабных колебаний аэрологических наблюдений имеют большое значение как в задачах изучения изменений климата и возможности его прогнозирования, так и при сравнении результатов моделирования с реальными данными.

Проведенные исследования позволяют сделать следующие выводы.

1. Основную долю изменчивости стандартных аномалий временных рядов анализируемых данных составляет белый шум,

Таблица 6

Оценки параметра линейного тренда (β_1) и его статистическая значимость (+), найденные по данным за различные периоды наблюдений

Широта, °	OT $_{1000}^{500}$			Зональная составляющая скорости геострофического ветра на уровне 500 мбар		
	1948—1955	1956—1965	1966—1975	1948—1955	1956—1965	1966—1975
80	-0,175	-0,216+	-0,047	0,327	0,156+	-0,038
75	-0,105	-0,179+	-0,067	0,205	0,057	-0,035
70	-0,090	-0,184+	-0,080	0,043	0,036	-0,022
65	-0,125+	-0,184+	-0,063	-0,104	0,100	0,005
60	-0,124+	-0,145+	-0,091	-0,106	0,098	-0,035
55	-0,163+	-0,093	-0,102	-0,137+	0,084	0,015
50	-0,243+	-0,083	-0,056	-0,068	0,026	0,056
45	-0,193+	-0,113+	-0,005	0,101	0,032	0,102+
40	-0,096+	-0,117+	0,082+	0,245+	0,061	0,109
35	0,064	-0,135+	0,122+	0,255+	0,064	0,045
30	0,094	-0,176+	0,180+	0,185	0,085	-0,055

уровень которого примерно 75...80 % для скорости ветра и примерно 65 % для OT $_{1000}^{500}$ от соответствующих дисперсий.

2. Спектры скорости ветра имеют небольшие максимумы в низких частотах с периодами в несколько лет; аналогичный максимум OT $_{1000}^{500}$ значительно мощнее. Этому участку частот соответствуют отдельные флуктуации периодограмм относительно большой интенсивности, показывающие возможность существования длиннопериодных колебаний наблюдений.

Ввиду того что за рассматриваемый промежуток времени количество аэрологических станций существенно увеличивалось, точность анализируемых среднеширотных значений возрастала. Мы не можем делать категорических утверждений о том, что об-

наруженные максимумы вызваны только естественными причинами. Рост числа станций неизбежно приводит к уточнению среднеширотных оценок и, следовательно, к появлению длиннопериодных тенденций, обусловленных постоянным уменьшением систематической ошибки среднего. Эти тенденции могут быть одной из причин длиннопериодных колебаний, обнаруженных как при анализе спектров, так и при рассмотрении линейного тренда. Подобные колебания являются следствием методических, а не естественных, физических причин.

Для того чтобы выяснить, какая часть обнаруженных особенностей вызвана ростом сети станций, необходимо провести аналогичный анализ хотя бы отдельных среднеширотных временных рядов, вычисленных с помощью наблюдений сети станций, существовавших в течение всего рассматриваемого периода времени.

3. При выбранном разрешении корреляции в пространстве значительно выше корреляции во времени.

4. Большой интерес представляют дальнейшие исследования аналогичных статистик всего комплекса аэрологических наблюдений на нескольких различных уровнях. Сравнительное изучение получаемых оценок, а также их взаимных статистических характеристик позволит выявить основные закономерности статистической структуры длиннопериодных колебаний метеоэлементов свободной атмосферы.

СПИСОК ЛИТЕРАТУРЫ

1. Некоторые пространственно-временные характеристики длиннопериодных колебаний атмосферного давления/Б. М. Ильин, И. И. Поляк, Л. В. Руховец, В. А. Шахмейстер.—Изв. АН СССР. Физика атмосферы и океана, 1978, т. 14, № 7.
2. Поляк И. И. Оценивание линейного тренда временных метеорологических рядов.—Труды ГГО, 1975, вып. 364, с. 51—55.
3. Поляк И. И. Оценивание спектральных плотностей и ковариационных функций двумерных случайных полей.—Труды ГГО, 1975, вып. 364, с. 45—50.
4. Поляк И. И. Численные методы анализа наблюдений.—Л.: Гидрометеиздат, 1975.—211 с.
5. Report of the JOC ad hoc Working Group on the global data base for climate research (Princeton, 21—23 April 1975). Geneva, 1976.
6. Oort A. H., Rasmusson E. M. Atmospheric circulation statistics. Rockville, Md, 1971.

СГЛАЖИВАНИЕ КОРРЕЛИРОВАННЫХ НАБЛЮДЕНИЙ

При изучении линейной фильтрации обычно предполагается, что анализируемые ряды наблюдений представляют собой равно- точные статистически независимые случайные величины [1].

Такое предположение позволяет получить интересные резуль- таты о статистической структуре точечных оценок, вычисляемых с помощью фильтров различного вида. Однако обработка реаль- ных данных часто приводит к необходимости сглаживания кор- релированных наблюдений. Статистическая структура точечных оценок в этом случае определяется видом ковариационной мат- рицы исходных данных. Наибольший интерес представляют ре- зультаты для стационарных корреляционных связей.

Рассмотрим реализацию

$$Y_0, Y_1, \dots, Y_N \quad (1)$$

стационарного случайного процесса $Y(t)$ с ковариационной функ- цией $M(\tau)$. Ковариационная матрица \mathbf{M} ряда (1) имеет теплице- ву форму

$$\mathbf{M} = \begin{pmatrix} M_0 & M_1 & M_2 & \dots & M_{N-1} & M_N \\ M_1 & M_0 & M_1 & \dots & M_{N-2} & M_{N-1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ M_N & M_{N-1} & M_{N-2} & \dots & M_1 & M_0 \end{pmatrix}, \quad (2)$$

где $M_\tau = M(\tau)$ ($\tau=0, 1, \dots, N$).

Найдем ковариационную структуру точечных оценок

$$\bar{Y}_t = \sum_{j=-r}^r \alpha_j Y_{t+j} \quad (t=r, r+1, \dots, N-r), \quad (3)$$

где α_j суть весовые коэффициенты произвольного цифрового фильтра, причем $\alpha_j = \alpha_{-j}$.

Рассмотрим матрицу

$$F = \{F_{i,j}\}_{i=r, j=0}^{N-r, N} = \begin{pmatrix} \alpha_{-r} & \alpha_{-(r-1)} & \dots & \alpha_r & 0 & \dots & 0 \\ 0 & \alpha_{-r} & \dots & \alpha_{r-1} & \alpha_r & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots & \alpha_r \end{pmatrix}. \quad (4)$$

Так как формула (3) осуществляет линейное преобразование наблюдений Y_t в \bar{Y}_t , то ковариационная матрица $R = \{R_{i,j}\}_{i,j=r}^{N-r}$ оценок \bar{Y}_t может быть представлена в виде

$$R = F M F^T, \quad (5)$$

где T — знак транспонирования.

Подставляя в правую часть (5) выражения элементов матриц M и F и осуществляя необходимые вычисления, получаем следующее.

1. Матрица R имеет теплицеву форму (элементы, стоящие на диагоналях, параллельных главной, равны между собой), т. е. величины \bar{Y}_t являются реализацией стационарного случайного процесса.

2. Ковариационная функция $R(\tau)$ такого процесса в точках $\tau=0, 1, \dots, N-2r$ равна

$$R_\tau = \sum_{v=-r}^r \sum_{j=-r}^r \alpha_j \alpha_v M_{j-v-\tau}. \quad (6)$$

Часто рассматривается корреляционная функция K_τ , имеющая вид

$$K_\tau = \frac{R_\tau}{R_0} = \frac{\sum_{v=-r}^r \sum_{j=-r}^r \alpha_j \alpha_v M_{j-v-\tau}}{\sum_{v=-r}^r \sum_{j=-r}^r \alpha_j \alpha_v M_{j-v}}. \quad (7)$$

В наиболее распространенном случае, когда сглаживание представляет собой простое осреднение

$$\bar{Y}_t = \frac{1}{2r+1} \sum_{j=-r}^r Y_{t+j}, \quad (8)$$

формулы (6) и (7) принимают вид

$$\begin{aligned} R_\tau &= \frac{1}{(2r+1)^2} \sum_{v=-r}^r \sum_{j=-r}^r M_{j-v-\tau} = \\ &= \frac{1}{2r+1} \left[M_\tau + \sum_{i=1}^{2r} \left(1 - \frac{i}{2r+1}\right) (M_{\tau+i} + M_{\tau-i}) \right], \quad (9) \end{aligned}$$

$$K_{\tau} = \frac{R_{\tau}}{R_0} \frac{M_{\tau} + \sum_{i=1}^{2r} \left(1 - \frac{i}{2r+1}\right) (M_{\tau+i} + M_{\tau-i})}{M_0 + 2 \sum_{i=1}^{2r} \left(1 - \frac{i}{2r+1}\right) M_i}. \quad (10)$$

Выпишем также выражение для R_0 :

$$R_0 = \frac{1}{2r+1} \left[M_0 + 2 \sum_{i=1}^{2r} \left(1 - \frac{i}{2r+1}\right) M_i \right], \quad (11)$$

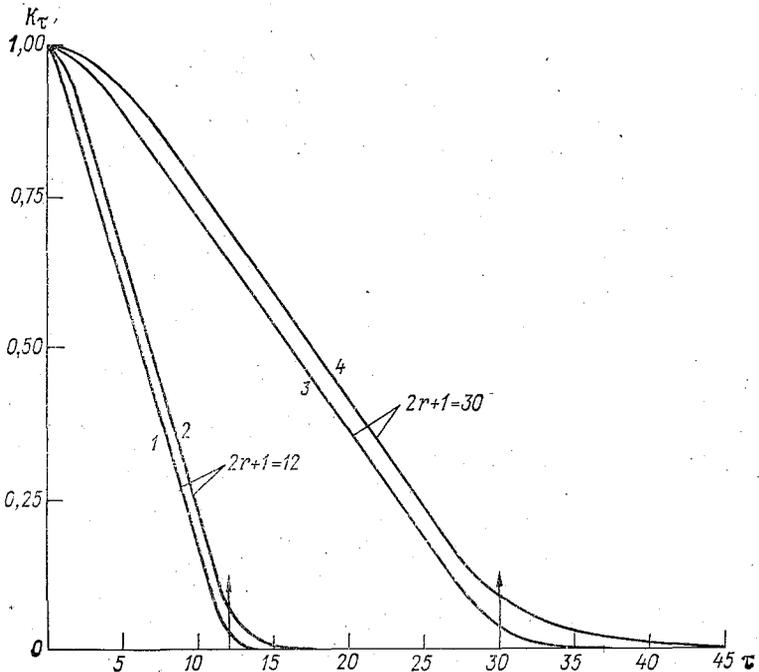
которое используется для нахождения оценок дисперсии среднего коррелированных случайных величин.

Пример. Предположим, что ковариационная функция наблюдений имеет вид

$$M(\tau) = \sigma^2 e^{-\alpha|\tau|}. \quad (12)$$

Этот пример имеет большое значение в метеорологии, поскольку часто выражение (12) хорошо аппроксимирует ковариационную функцию значений метеоэлементов [2].

Подставляя (12) в (10), для различных значений α можно вычислить корреляционную функцию K_{τ} .



Корреляционные функции для среднегодовых (1, 2) и среднемесячных (3, 4) данных.

1) $e^{-\alpha}=0.15$; 2) $e^{-\alpha}=0.5$; 3) $e^{-\alpha}=0.6$; 4) $e^{-\alpha}=0.8$.

Рассмотрим два случая.

1. Оценим возможную корреляцию соседних среднемесячных значений метеозаэлементов, найденных осреднением ежедневных наблюдений с ковариационной функцией вида (12) и с коэффициентом корреляции от одного дня к другому, заключенным в интервале от 0,6 до 0,8. Такая ситуация вполне реальна для температуры воздуха, атмосферного давления и ряда других метеозаэлементов [2].

Считая в месяце 30 дней ($2r+1=30$), построим корреляционную функцию сглаженных величин, подставляя в формулу (10) $2r+1=30$ и $M_\tau = \sigma^2 \cdot 0,6^\tau$ ($M_\tau = \sigma^2 \cdot 0,8^\tau$). Результаты даны на рисунке, искомый коэффициент корреляции соответствует значению $\tau=30$.

Как следует из графика, при сделанных предположениях значение этого коэффициента не превосходит 0,1, т. е. корреляция среднемесячных значений, обусловленная синоптическими флуктуациями погоды, меньше 0,1. Конечно, для получения этого результата график можно было и не строить, а положить в формуле (10) $\tau=30$.

При произвольном виде ковариационной функции $M(\tau)$ ежедневных значений коэффициент корреляции соседних значений среднемесячных величин вычисляется по формуле

$$K_{30} = \frac{M_{30} + \sum_{i=1}^{29} \left(1 - \frac{i}{30}\right) (M_{30+i} + M_{30-i})}{M_0 + 2 \sum_{i=1}^{29} \left(1 - \frac{i}{30}\right) M_i} \quad (13)$$

2. Оценим коэффициент корреляции соседних значений среднегодовых значений метеозаэлементов, полученных осреднением среднемесячных данных, коэффициент корреляции которых при запаздывании в один месяц заключен в интервале от 0,15 до 0,5, а ковариационная функция имеет вид (12). Ситуация также вполне реальна [2].

Проведя вычисления, аналогичные рассмотренному случаю, с $2r+1=12$, $M_\tau = \sigma^2 \cdot 0,15^\tau$ ($M_\tau = \sigma^2 \cdot 0,5^\tau$), получаем корреляционную функцию сглаженных значений (см. рисунок). При $\tau=12$ находим, что коэффициент корреляции среднегодовых величин не превосходит 0,07.

При произвольной ковариационной функции $M(\tau)$ среднемесячных значений формула, аналогичная (13), для подсчета корреляции среднегодовых величин имеет вид

$$K_{12} = \frac{M_{12} + \sum_{i=1}^{11} \left(1 - \frac{i}{12}\right) (M_{12+i} + M_{12-i})}{M_0 + 2 \sum_{i=1}^{11} \left(1 - \frac{i}{12}\right) M_i} \quad (14)$$

СПИСОК ЛИТЕРАТУРЫ

1. Поляк И. И. Численные методы анализа наблюдений.— Л.: Гидрометеониздат, 1975.— 212 с.
2. Madden R. A. Estimates of the autocorrelations and spectra of seasonal mean temperatures over North America.— Mon. Wea. Rev., 1977, vol. 105, N 1, p. 9—18.

А. П. Соколов

ПОСТРОЕНИЕ МОДЕЛИ АВТОРЕГРЕССИИ — СКОЛЬЗЯЩЕГО СРЕДНЕГО ДЛЯ СТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА

В последнее время модели авторегрессии — скользящего среднего (АРСС) — находят широкое применение при решении различных задач, особенно в экономике. Следует ожидать, что в ближайшее время этот метод будет использоваться и в метеорологии. В связи с этим возникает задача создания математического обеспечения для решения этих задач.

В настоящей статье дается краткое описание алгоритма и программы, реализующей схему нахождения модели АРСС для стационарного временного ряда. Подробное описание моделей АРСС и алгоритма их построения даны в [1]. Модель авторегрессии — скользящего среднего порядка (p, q) — для стационарного временного ряда имеет вид

$$W_t = b_0 + b_1 w_{t-1} + \dots + b_p w_{t-p} + a_t - b_{p+1} a_{t-1} - \dots - b_{p+q} a_{t-q},$$

где $\{W_t\}_{t=0, 1, \dots, m}$ — данный временной ряд; b_0 — оценка общей константы; b_1, \dots, b_p — параметры авторегрессии; b_{p+1}, \dots, b_{p+q} — параметры скользящего среднего; $\{a_t\}_{t=0, 1, \dots, m}$ — ряд остаточных ошибок. При адекватном построении модели остаточные ошибки должны быть белым шумом.

С помощью имеющейся модели АРСС можно проанализировать систему, генерирующую данный временной ряд, прогнозировать будущие значения ряда, получить выражение для корреляционной функции и спектра процесса через параметры модели.

Построение модели состоит из следующих этапов.

I. Получение начальных оценок параметров. Начальные оценки получаются из соотношений, связывающих параметры модели с корреляционной функцией временного ряда.

II. Получение оценок методом наименьших квадратов. Оценки

В качестве начальных значений принимаем

$$\tau_0 = \sqrt{c'_0}, \quad \tau_1 = \tau_2 = \dots = \tau_q = 0.$$

Вычисления заканчиваются, когда $\max |f_j| < \varepsilon$ ($j=0, 1, \dots, q$).
Оценки параметров скользящего среднего находятся по формуле

$$b_{j+p} = -\tau_j/\tau_0 \quad (j=1, 2, \dots, q).$$

4. Оценка общей константы b_0 имеет вид

$$b_0 = \begin{cases} \omega_0 \left(1 - \sum_{i=1}^p b_i \right) & p > 0, \\ \omega_0 & p = 0, \end{cases}$$

где ω_0 — среднее значение ряда ω_t .

II. Получение оценок максимального правдоподобия и критериев проверки адекватности модели.

В качестве оценок максимального правдоподобия параметров модели принимаются значения параметров, минимизирующие сумму квадратов остаточных ошибок:

$$S(\mathbf{b}) = \sum_{i=-Q}^M a_i^2,$$

где $(\mathbf{b} = b_0, b_1, \dots, b_p, b_{p+1}, \dots, b_{p+q})$.

Для вычисления $S(\mathbf{b})$ необходимо прогнозированием назад получить значения $\omega_{-1}, \omega_{-2}, \dots, \omega_{-Q}$, где Q определяется из условия $(\omega_{-j} - \omega_0) \approx 0$ для $j > Q$ [1, с. 240].

Минимизация $S(\mathbf{b})$ производится при помощи алгоритма Марквардта [1, с. 376]. Необходимо задать значения $p_i, p_i M$ и f , ограничивающих поиск. Во время поиска необходимо вычислять величины $a_t(\mathbf{b})$ и $x_{it} = -\partial a_t / \partial b_i$, где $t = -Q, \dots, M, i = 1, \dots, p+q$ на каждом шаге итерации.

Этап 1. С помощью величин a_t и x_{it} , найденных по текущим значениям параметров, находим следующие величины:

1) матрицу размером $(p+q) \times (p+q)$

$$\mathbf{V} = \{v_{ij}\},$$

$$\text{где } v_{ij} = \sum_{t=-Q}^M x_{it} x_{jt};$$

2) вектор

$$\mathbf{g} = (g_1, g_2, \dots, g_{p+q}),$$

$$\text{где } g_i = \sum_{t=-Q}^M x_{it} a_t;$$

3) нормирующие величины $d_i = \sqrt{v_{ii}}$.

Этап 2.

1) вычисляем нормированные величины

$$v_{ij}^* = v_{ij}/d_i d_j \quad i \neq j,$$

$$v_{ii}^* = 1 + p_i,$$

$$g_i^* = g_i/d_i;$$

2) решаем матричное уравнение

$$\mathbf{V}^* \mathbf{h}^* = \mathbf{g}^*$$

относительно \mathbf{h}^* ;

3) находим новые значения параметров по формуле

$$b_i^{j+1} = b_i^j + h_i,$$

где $h_i = h_i^*/d_i$,
и вычисляем $S(\mathbf{b}^{j+1})$.

Этап 3.

1) если $S(\mathbf{b}^{j+1}) < S(\mathbf{b}^j)$, исследуются поправки параметров \mathbf{h} . Если $\sum |h_i| < \epsilon$, то вычисления заканчиваются и в качестве оценок параметров принимается \mathbf{b}^j . Если $\sum |h_i| > \epsilon$, то \mathbf{b}^j принимает значения \mathbf{b}^{j+1} , p_i уменьшается в f раз и вычисления повторяются, начиная с этапа 1;

2) если $S(\mathbf{b}^{j+1}) > S(\mathbf{b}^j)$, то p_i увеличивается в f раз и вычисления повторяются, начиная с этапа 2. Параметр p_i имеет максимальное значение $p_i M$, и если $p_i > p_i M$, то вычисления заканчиваются.

После нахождения оценок наименьших квадратов параметров модели вычисляются следующие величины:

оценка остаточной дисперсии ряда остаточных ошибок

$$S_a = S(\mathbf{b})/(M + 1);$$

оценка общей константы

$$b_0 = \left(1 - \sum_{i=1}^p b_i\right) \omega_0 \quad p \neq 0;$$

стандартные ошибки параметров

$$\sigma_i = \sqrt{S_a v_{ii}^{-1}}.$$

Для проверки адекватности полученной модели вычисляются следующие величины:

1) остаточные автокорреляции по ряду остаточных ошибок

$$\text{cov}(j) = \text{cov}(j)/\text{cov}(0) \quad j = 1, 2, \dots, k,$$

где

$$\text{cov}(j) = \frac{1}{M+1} \sum_{t=0}^{M-j} (a_t - a_0)(a_{t+j} - a_0) \quad j = 0, 1, \dots, k,$$

$$a_0 = \frac{1}{M+1} \sum_{i=0}^M a_i.$$

2) совокупный критерий согласия (χ^2 — статистика)

$$Q = (M+1) \sum_{j=1}^k (\text{cог}(j))^2.$$

Если модель адекватна данному временному ряду, то Q распределено приближенно как $\chi^2(k-p-q)$.

Программа, в которой реализован описанный выше алгоритм, составлена на языке ГДР-АЛГОЛ для ЭВМ БЭСМ-6. Рассматриваемый временной ряд может вводиться как с перфокарт, так и читаться с магнитной ленты. В программе предусмотрена возможность проведения счета по части ряда.

Исходные данные состоят из нескольких последовательностей. Каждая последовательность набивается на отдельной перфокарте. Порядок описания последовательностей совпадает с порядком расположения соответствующих операторов ввода.

Первая последовательность состоит из 7 элементов. Порядок описания элементов совпадает с порядком их расположения в операторе ввода:

- 1) p — число параметров авторегрессии;
- 2) q — число параметров скользящего среднего;
- 3) N — величина, на единицу меньшая длины вводимого временного ряда;
- 4) k — сдвиг при расчете по части ряда;
- 5) M — величина, на единицу меньшая длины рассматриваемой части ряда;
- 6) L — число точек на период гармонического тренда. Например, при обработке временного ряда среднемесячных величин и необходимости фильтрации годового хода $L=12$. При $L=0$ фильтрация тренда не производится;
- 7) NN — параметр, определяющий вид кодировки чисел временного ряда:

$NN=1$ числа в кодировке ЭВМ М-220,

$NN=2$ числа в кодировке ЭВМ БЭСМ-6 для бесформатного ввода,

$NN>2$ — числа записаны на МЛ.

Вторая последовательность состоит из одного числа:

eps — точность, используемая при нахождении начальных оценок параметров скользящего среднего в алгоритме Ньютона — Рафсона, при нахождении оценок наименьших квадратов параметров модели в алгоритме Марквардта и в стандартных процедурах обращения матрицы. В программе использовано значение 10^{-4} .

Третья последовательность состоит из трех чисел:

d — величина, определяющая число шагов при прогнозировании назад в процедуре вычисления остаточных ошибок;

dl — шаг, используемый при вычислении производных остаточных ошибок;

k — длина ряда корреляций остаточных ошибок.

Четвертая последовательность состоит из трех чисел:

pi , f , piM — параметры, ограничивающие поиск в алгоритме Марквардта.

В программе использованы значения: $pi=1$; $f=2$; $piM=100$.

Последняя последовательность состоит из одного числа:

L — максимальное упреждение при прогнозировании по полученной модели.

При вводе временного ряда с перфокарт массив чисел подкладывается после второй перфокарты исходных данных (ввод eps).

Все получаемые оценки выдаются на печать с соответствующими надписями, как правило не требующими пояснений.

СПИСОК ЛИТЕРАТУРЫ

1. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Вып. 1. — М.: Мир, 1974.

В. А. Шахмейстер

ПРОГРАММА СПЕКТРАЛЬНОГО АНАЛИЗА ОДНОРОДНОГО СЛУЧАЙНОГО ПОЛЯ

Приведем описание алгоритма и программы двумерного спектрального анализа однородного случайного поля. В основу программы положен алгоритм оценивания периодограммы, спектральной плотности и корреляционной функции случайного поля по его конечной реализации (1), описанный в [4],

$$Y_{\theta t} \begin{pmatrix} \theta = 0, 1, \dots, N-1 \\ t = 0, 1, \dots, K-1 \end{pmatrix}. \quad (1)$$

Программа написана для транслятора с Фортрана для ЭВМ БЭСМ-6. В ней предусмотрены различные процедуры первоначальной обработки поля $Y_{\theta t}$ с целью исключения детерминированной составляющей. Применение той или иной процедуры зависит от природы анализируемого случайного поля. Рассмотрим кратко эти процедуры.

1. Исключение широтного хода в рассматриваемом поле и его дисперсии с помощью перехода к отклонениям от выборочных средних, рассчитанных по широтным кругам, и нормировки на выборочный стандарт.

2. Исключение широтного хода в анализируемом поле и его дисперсии с помощью климатических данных о широтном ходе норм и дисперсий этого поля.

Процедуры 1 и 2 могут быть использованы при анализе пространственных полей метеозлементов, содержащих широтный тренд.

3. Переход к отклонениям от поля, восстановленного по первым nn (по оси θ), kk (по оси t) двумерным гармоникам, разложения $Y_{\theta t}$ в ряд Фурье.

4. Исключение годового хода из полей среднемесячных величин $Y_{\theta t}$, где θ представляет собой изменение от года к году, а t — от месяца к месяцу. Алгоритм такого исключения описан в [2].

Последующая обработка измерений состоит в переходе к отклонениям от среднего по полю $Y_{\theta t}$ и нормировке на стандарт, рассчитанный по всему полю.

В программе используется алгоритм преобразования Фурье [1], который был расширен на двумерный случай. Преобразование производится сначала по строкам, затем по столбцам исходного поля.

Двумерное дискретное разложение Фурье реализации $Y_{\theta t}$ может быть записано в виде

$$\begin{aligned} A_{pq} &= \frac{1}{Nk} \sum_{\theta=0}^{N-1} \sum_{t=0}^{k-1} Y_{\theta t} e^{-i(\Omega_p \theta + \omega_q t)} = \\ &= \frac{1}{N} \sum_{\theta=0}^{N-1} c_{\theta q} (\cos \Omega_p \theta - i \sin \Omega_p \theta) \end{aligned} \quad (2)$$

где

$$(\Omega_p = 2\pi p/N, \quad \omega_q = 2\pi q/k)$$

$$c_{\theta q} = \frac{1}{k} \sum_{t=0}^{k-1} Y_{\theta t} (\cos \omega_q t - i \sin \omega_q t). \quad (3)$$

Обозначим через $a_{\theta q}$ и $b_{\theta q}$ соответственно вещественные коэффициенты косинус и синус преобразования Фурье исходного поля по строкам. Тогда

$$\begin{aligned} c_{\theta q} &= \frac{1}{2} (a_{\theta q} - i b_{\theta q}) \quad (q = 1, 2, \dots, k1, k1 = k+2), \\ c_{\theta 0} &= a_{\theta 0}/2, \\ c_{\theta k1} &= a_{\theta k1}/2 \quad (\text{при четном } k). \end{aligned} \quad (4)$$

Вследствие симметрии коэффициентов Фурье

$$c_{\theta q} = c_{\theta k-q}^* \quad (q = k1 + 1, \dots, k). \quad (5)$$

Пусть

$$c_{\theta q} = c'_{\theta q} + i c''_{\theta q}. \quad (6)$$

Тогда формула (2) примет вид

$$\begin{aligned} A_{pq} &= A'_{pq} + i A''_{pq} = \frac{1}{N} \sum_{\theta=0}^{N-1} (c'_{\theta q} \cos \Omega_p \theta + c''_{\theta q} \sin \Omega_p \theta) + \\ &+ \frac{i}{N} \sum_{\theta=0}^{N-1} (c''_{\theta q} \cos \Omega_p \theta - c'_{\theta q} \sin \Omega_p \theta) \\ &\left(\begin{array}{l} p = 1, 2, \dots, N1, N1 = N + 2 \\ q = 1, 2, \dots, k - 1 \end{array} \right). \end{aligned} \quad (7)$$

Из симметрии двумерных коэффициентов Фурье следует, что

$$\begin{aligned} A_{pq}^* &= A_{N-p, k-q} & (p = N1 + 1, \dots, N - 1) \\ A_{p0} &= A_{N-p, 0} & (q = 1, \dots, k - 1) \end{aligned} \quad (8)$$

Вычисление периодограммы, спектральной плотности и корреляционной функции поля $Y_{\theta t}$ производилось по формулам, приведенным в [5].

Так же как и в [5], в программе производится преобразование нормированных периодограмм и спектров в область неотрицательных частот ($0 \leq p \leq N1$, $0 \leq q \leq k1$).

Равенство Парсеваля [4] для нормированной периодограммы имеет вид

$$\frac{1}{Nk} \sum_{\theta=0}^{N-1} \sum_{t=0}^{k-1} Y_{\theta t}^2 = \sum_{p=0}^{N-1} \sum_{q=0}^{k-1} S_{pq}^2, \quad (9)$$

или

$$\frac{1}{Nk} \sum_{\theta=0}^{N-1} \sum_{t=0}^{k-1} Y_{\theta t}^2 = \sum_{p=0}^{N1} \sum_{q=0}^{k1} S'_{pq}, \quad (10)$$

где S'_{pq} связано с S_{pq} соотношениями:

$$\begin{aligned} S'_{00} &= S_{00}, \\ S'_{0q} &= 2S_{0q} \quad (q = 1, \dots, k1 - 1), \\ S'_{p0} &= 2S_{p0} \quad (p = 1, \dots, N1 - 1), \\ S'_{pq} &= 2(S_{pq} + S_{N-p, q}) \quad \begin{cases} (p = 1, \dots, N1 - 1) \\ (q = 1, \dots, k1 - 1) \end{cases}; \end{aligned} \quad (11)$$

$$\begin{aligned} S'_{N10} &= \begin{cases} S_{N10} & N - \text{четное,} \\ 2S_{N10} & N - \text{нечетное,} \end{cases} \\ S'_{0k1} &= \begin{cases} S_{0k1} & k - \text{четное,} \\ 2S_{0k1} & k - \text{нечетное;} \end{cases} \end{aligned} \quad (12)$$

$$S'_{N1q} = \begin{cases} 2S_{N1q} & N - \text{четное,} \\ 2(S_{N1q} + S_{N-N1q}) & N - \text{нечетное,} \end{cases} \quad (q = 1, \dots, k1 - 1); \quad (13)$$

$$S'_{pk1} = \begin{cases} 2S_{pk1} & k - \text{четное} \\ 2(S_{pk1} + S_{N-p, k1}) & k - \text{нечетное} \end{cases} \quad (p = 1, \dots, N1 - 1); \quad (14)$$

$$S'_{N1k1} = \begin{cases} S_{N1k1} & N, k - \text{четные,} \\ 2(S_{N1k1} + S_{N-N1k1}) & N, k - \text{нечетные,} \\ 2S_{N1k1} & N - \text{четное, } k - \text{нечетное,} \\ 2S_{N1k1} & N - \text{нечетное, } k - \text{четное.} \end{cases} \quad (15)$$

Если исходное поле задано с одинаковой дискретностью по осям θ и t , то после вычисления двумерных характеристик рассчитываются их одномерные аналоги путем осреднения по градациям. Переход от двумерной корреляционной функции $K(\mu, \tau)$ к ее одномерному аналогу производится осреднением значений, находящихся на равном расстоянии от начала координат в пространстве запаздываний. Одномерные характеристики для периодограмм и спектральных плотностей определяются суммированием в пространстве частот амплитуд двумерных гармоник, попадающих в заданные градации осреднения, причем суммирование производится по кругам и прямоугольникам.

Исходные данные к программе должны быть подготовлены в виде шести массивов. Ввод массивов I—III обязателен, а массивы IV—VI вводятся по условию, в зависимости от значений соответствующих элементов информационного массива I.

1. Информационный массив из 13 элементов. Он состоит из целых чисел, набитых в формате (13i5):

1. N — количество строк исходного поля;
2. k — количество столбцов исходного поля;
3. $m1$ определяет вид сглаживания периодограммы:

$$m1 = \begin{cases} -1 & \text{— счет без сглаживания,} \\ 0 & \text{— счет со сглаживанием по формулам,} \\ & \text{описанным в [3],} \\ 1 & \text{— счет со сглаживанием, в качестве сгла-} \\ & \text{женного значения принимается среднее} \\ & \text{значение по прямоугольнику, опреде-} \\ & \text{ляемому параметрами } s \text{ и } r, \text{ где} \end{cases}$$

4. s — полуширина фильтра вдоль оси θ ;
5. r — полуширина фильтра вдоль оси t ;
6. $m2$ определяет вид выдачи полей результатов на АЦПУ:

$$m2 = \begin{cases} -1 & \text{— печать производится через 1 интервал АЦПУ,} \\ 1 & \text{— печать производится через 2 интервала АЦПУ;} \end{cases}$$

7. $m3$ — свободен.

8. $m4 = \begin{cases} -1 & \text{— ввод исходных данных с перфокарт в формате М-220,} \\ 0 & \text{— ввод с перфокарт в фортрановском формате 12f6.1,} \\ 1 & \text{— чтение с магнитной ленты с математическим} \end{cases}$

номером 43;

$$9. m5 = \begin{cases} -1 & \text{— не делается исключение широтного тренда,} \\ & 0 \text{— исключение тренда производится по широт-} \\ & \text{ным кругам (с помощью процедуры 1, опи-} \\ & \text{санной выше),} \\ & 1 \text{— исключение широтного тренда производится} \\ & \text{с помощью процедуры 2;} \end{cases}$$

$$10. m6 = \begin{cases} 1 & \text{— печать периодограммы и спектра (в } 10^2 \% \text{)} \\ & \text{производится во всей области расчета и в об-} \\ & \text{ласти неотрицательных частот,} \\ 0 & \text{— периодограммы и спектры печатаются только} \\ & \text{в области неотрицательных частот;} \end{cases}$$

11. $m7$ — количество исключаемых гармоник по оси θ ;

12. $m8$ — количество исключаемых гармоник по оси t . Если нет исключения гармоник, то параметры $m7$ и $m8$ полагаются равными нулю;

13. $m9 = \begin{cases} 0 & \text{— не производится исключение годового хода в ряду и дисперсии,} \\ -12 & \text{— исключается годовой ход, поля результатов печатаются в два столбца на странице,} \\ 12 & \text{— годовой ход не исключается, печать результатов производится в два столбца на странице.} \end{cases}$

II. Вводятся два числа $d1$ и $d2$ — шаги отсчета по осям t и θ соответственно (в формате 2f12.7).

III. При $m4 \neq 1$ вводится поле $Y_{\theta t}$, подготовленное по столбцам в формате (12f6.1) или в формате М-220. При $m4 = 1$ вводится $m10$ — трехзначное восьмеричное число (в формате i5) — номер зоны на ленте с математическим номером 43, начиная с которой расположено исходное поле.

IV. При $m5 = 0$ производится исключение широтного хода в рассматриваемом поле и его дисперсии с помощью расчета выборочных средних и дисперсий. При этом вводится информационный массив из 5 чисел (в формате (415, f i2.7)):

1. $l1$ — координата Северного полюса вдоль оси θ ;
 2. $l2$ — координата Северного полюса вдоль оси t (параметры $l1, l2$ — это номера точек сетки);
 3. $l3$ — ширина полосы вдоль оси θ , в которой рассчитываются средние, дисперсии;
 4. $l4$ — ширина полосы расчета средних и дисперсий вдоль оси t (параметры $l3, l4$ задаются в точках сетки);
5. $bf = \begin{cases} 1 & \text{— производится вычитание в каждом широтном круге соответствующего среднего,} \\ 2 & \text{— делается нормировка на стандарт,} \\ 3 & \text{— в каждом широтном круге производится переход к нормированным аномалиям.} \end{cases}$

V. При $m5 = 1$ исключается широтный ход в поле и его дисперсии с помощью климатических данных. При этом должны быть введены следующие пять массивов:

- a) $i\theta$ — координата Северного полюса по оси θ ,
 - б) it — координата Северного полюса по оси t ,
 - в) $j1$ — размерность массива климатических норм,
 - г) $j2$ — размерность массива климатических стандартов,
- д) $bf = \begin{cases} 1 & \text{— имеют тот же смысл,} \\ 2 & \text{что и в информационном} \\ 3 & \text{массиве IV.} \end{cases}$

Массив 1 вводится в формате (5i5);

2. массив координат норм (в градусах), размерности $j1$;
3. массив норм;
4. массив координат для стандартов, размерности $j2$;

5. массив стандартов.

Массивы 2—5 вводятся в формате (12f7.3).

VI. При $N=k$ вводятся два числа (в формате (f12.7, i5)):

1. aa — размер градации при осреднении периодограмм, спектров и корреляционных функций (в точках сетки).

2. $l5$ — размерность одномерных массивов осредненных периодограмм, спектров и корреляционных функций (должно быть $l5 \leq N/2 + 1$). Все печатаемые характеристики, кроме информационных массивов, снабжены соответствующими заголовками. Печать производится в следующем порядке:

1. Информационные массивы I и II.

2. При исключении детерминированной составляющей печатаются вводимые при этом информационные массивы.

3. Печатаются частоты, периоды и сдвиги по осям t и θ . (В заголовках ось t названа « x axis», а ось θ — « y axis»).

4. Печатаются среднее и дисперсия поля под заголовками «mean value» и «dispersion».

5. Под заголовком «periodograms» печатается представление периодограммы в I квадранте — s'_{pq} , ее одномерные аналоги (если они рассчитываются) и затем общее поле периодограммы s_{pq} (в 10²%).

6 В случае счета со сглаживанием печатаются стандарты периодограммы и спектра с соответствующими заголовками.

7. При счете со сглаживанием под заголовком «spectrum» печатаются все представления спектральной плотности (аналогично п. 5).

8. Печатается одномерный аналог корреляционной функции (если он рассчитывается).

9. Под заголовком «correlation function» печатается двумерная корреляционная функция. Перед выдачей на печать значения корреляционной функции умножаются на 100.

Программа предоставляет пользователю широкие возможности для изучения статистических характеристик рассматриваемых полей и может быть использована при анализе полей метеоэлементов самого различного вида: пространственных, временных, широтно-временных и т. д.

СПИСОК ЛИТЕРАТУРЫ

1. Алгоритмы (151—200)/М. И. Агеев, М. Г. Грюнберг, Ю. И. Марков, Г. М. Швакова. — М.: Изд-во ВЦ АН СССР, 1970. — 218 с.

2. Гриб Н. К. Об одной схеме гармонического анализа временных метеорологических рядов. — Труды ГГО, 1975, вып. 364, с. 62—65.

3. Поляк И. И. Об одном классе двумерных фильтров. — Труды ГГО, 1975, вып. 364, с. 56—59.

4. Поляк И. И. Оценка спектральных плотностей и ковариационных функций двумерных случайных полей. — Труды ГГО, 1975, вып. 364, с. 45—50.

5. Шахмейстер В. А. Программа взаимного гармонического анализа двух однородных случайных полей. — Труды ГГО, 1978, вып. 409, с. 63—70.

Б. Г. Вагер

ИСПОЛЬЗОВАНИЕ ОДНОМЕРНЫХ СПЛАЙНОВ ПРИ ОБРАБОТКЕ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

В настоящее время интенсивно развивается и широко применяется сравнительно новый аппарат вычислительной математики — сплайны [1—4]. Они используются при решении следующих задач: 1) восстановление функций по известным значениям в узлах сетки, 2) нахождение функционалов (в частности, производных и интегралов) от таблично заданных функций, 3) численное интегрирование дифференциальных уравнений. Если первые два направления приложений сплайнов уже хорошо исследованы, то в третьем направлении делаются еще только первые пробные шаги. Успешное применение сплайнов при решении задач теории приближений объясняется тем, что сплайны имеют ряд преимуществ перед такими классическими аппаратами теории приближения, как алгебраические и тригонометрические полиномы. Во-первых, по своему построению они непрерывны вместе со своими производными до $(m-1)$ -го порядка в узлах сетки (m — степень сплайна); во-вторых, сплайны обладают экстремальными свойствами, и, наконец, они оказались удобным аппаратом в связи с простой реализацией алгоритмов на ЭВМ.

В настоящей работе изучается возможность применения кубических сплайнов при решении задач первого типа на материале экспедиционных данных о распределении метеоэлементов в пограничном слое атмосферы. Пример использования сплайнов при решении задач второго типа по тем же материалам рассмотрен в [5].

Перейдем теперь к математической постановке задачи. По известным значениям аргумента $a = x_0 < x_1 < \dots < x_n = b$ и соответствующим значениям функции y_0, y_1, \dots, y_n ($y_i = f(x_i)$) требуется построить функцию $S(x)$, заданную для всех x из отрезка $[a, b]$ и удовлетворяющую условию

$$S(x_i) = y_i \quad (i = 0, 1, \dots, n).$$

Сплайном $S_m(f, x)$ порядка m называется функция, являющаяся многочленом степени m на каждом из отрезков $[x_{k-1}, x_k]$

и удовлетворяющая условию непрерывности производных до $(m-1)$ -го порядка в точках x_k :

$$S_m(f, x) \equiv S_{m, k}(x) = a_{k0} + a_{k1}x + a_{k2}x^2 + \dots + a_{km}x^m \quad (x_{k-1} \leq x \leq x_k),$$

$$S_{m, k}^{(i)}(x_k) = S_{m, k+1}^{(i)}(x_k) \quad (i = 1, 2, \dots, m-1). \quad (1)$$

Введем общепринятые в математическом анализе обозначения: $C^2(a, b)$ — пространство функций, непрерывных вместе с двумя первыми производными на (a, b) ; $W_2^2(a, b) \in C^1(a, b)$ — пространство функций, имеющих суммируемые с квадратом вторые производные.

Будем искать решение сформулированной задачи интерполирования в виде функции $g(x)$, удовлетворяющей следующим условиям:

$$g(x) \in C^2(a, b), \quad (2)$$

$$g(x) \equiv g_k(x) = \sum_{i=0}^3 a_i^{(k)}(x_k - x)^i \quad (k = 1, 2, \dots, n), \quad (3)$$

$$g''(a) = g''(b) = 0, \quad (4)$$

$$g(x_k) = f_k \quad (k = 0, 1, \dots, n). \quad (5)$$

Равенство (3) означает, что $g(x)$ на каждом интервале $[x_{k-1}, x_k]$ является кубическим полиномом с неизвестными пока коэффициентами $a_i^{(k)}$, причем верхний индекс k указывает принадлежность данного коэффициента интервалу $[x_{k-1}, x_k]$.

Как следует из (3), для определения функции $g(x)$ надо найти $4n$ неизвестных коэффициентов $a_i^{(k)}$ ($i=0, 1, 2, 3$; $k=1, 2, \dots, n$). Из условия непрерывности функции и ее первых двух производных (2) имеем $(3n-3)$ уравнений. Условие прохождения сплайна через заданные точки (5) дает еще $(n+1)$ уравнение. Замыкают систему уравнений два граничных условия [4]. После несложных преобразований удается выразить коэффициенты $(a_0^{(k)}, a_1^{(k)}, a_3^{(k)})$ через коэффициенты $a_2^{(k)}$. Матрица, соответствующая системе линейных алгебраических уравнений для определения коэффициентов $a_2^{(k)}$, оказывается трехдиагональной и для ее численного решения можно применить метод прогонки [4]. Подробное описание алгоритма нахождения коэффициентов сплайна (3) приведено в [3]. В этой же монографии доказывается существование и единственность кубического сплайна и отмечается его очень важное экстремальное свойство, а именно $g(x)$ является решением следующей вариационной задачи: среди всех допустимых функций $u(x) \in W_2^2(a, b)$ и таких, что $u(x_k) = f_k$ ($k=0, 1, \dots, n$), функция $g(x)$ минимизирует функционал:

$$\Phi(u) \equiv \int_a^b \left(\frac{d^2u}{dx^2} \right)^2 dx, \quad \Phi(g) = \min_{u \in W_2^2(a, b)} \Phi(u). \quad (6)$$

До сих пор мы требовали выполнения условия (5). Однако такая постановка задачи интерполяции является не единственно возможной и даже часто не самой целесообразной. В ряде задач искомая функция может не точно проходить через заданные точки, а лишь достаточно близко от них. Необходимость в таких условиях возникает, когда функция $f(x)$ найдена из опыта. Точность получения экспериментальных значений часто бывает невысокой. В таких задачах излишне требовать прохождения функции через экспериментальные точки, достаточно, чтобы она отклонялась от этих точек по ординате на величину, не большую погрешности измерения.

Для решения такой задачи интерполяции используется кубический сплайн со сглаживанием. Искомая интерполяционно-сглаживающая функция $g_1(x)$ подчиняется первым трем условиям кубического сплайна (2)–(4), а условие (5) заменяется следующим требованием: $g_1(x)$ должно быть решением вариационной задачи:

$$\Phi_1(u) \equiv \int_a^b \left(\frac{d^2 u}{dx^2} \right)^2 dx + \sum_{k=0}^n p_k [u(x_k) - f_k]^2,$$

$$\Phi_1(g_1) = \min_{u \in \mathcal{W}_2^2(a, b)} \Phi_1(u). \quad (7)$$

Так же как и в предыдущем случае, коэффициенты сплайна $g_1(x) - a_0^{(k)}, a_1^{(k)}, a_3^{(k)}$ удается выразить через $a_2^{(k)}$. Матрица систе-

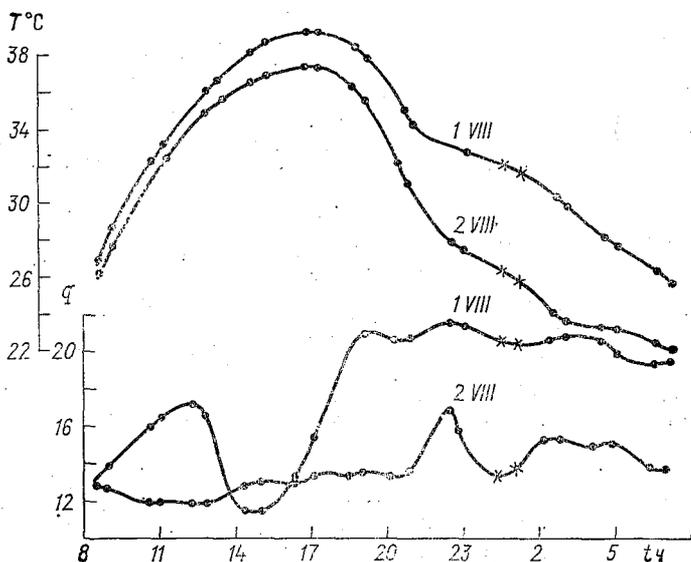


Рис. 1. Суточный ход температуры и влажности воздуха в августе.

мы уравнений сохраняет ленточную структуру, однако будет уже пятидиагональной и для ее решения можно применить метод факторизации [3, 6]. Формулы нахождения коэффициентов сглаживающего сплайна приведены в [3].

Рассмотрим теперь некоторые результаты обработки экспериментальных данных с помощью сплайнов. Оба алгоритма построения кубического сплайна и сглаживающего сплайна были реализованы в виде программы на Фортране. В экспедиции отдела физики пограничного слоя ГГО (1977 г., южные районы страны) проводились наблюдения за температурой и влажностью воздуха в течение суток с интервалом в полчаса и полтора часа.

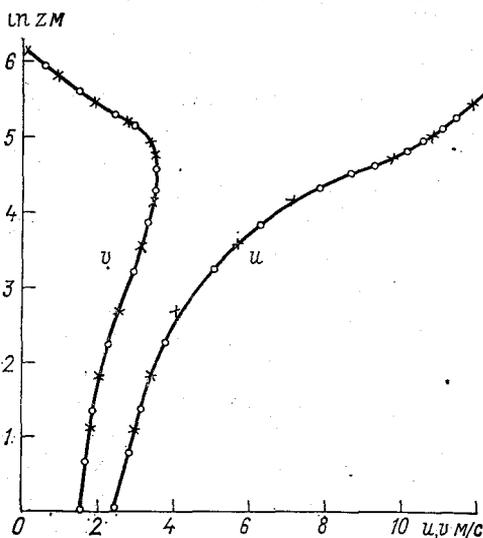


Рис. 2. Вертикальные профили компонент скорости ветра.

В ночное время перерыв между наблюдениями составлял 4 ч. Следовало восстановить значения температуры и влажности в этом промежутке времени, что и было сделано с помощью кубического сплайна.

На рис. 1 приведены кривые суточного хода температуры и влажности на высоте 2 м. Точками отмечены данные наблюдений, а крестиками — значения сплайна в искомым точках. Результаты применения кубического сплайна при интерполяции вертикальных профилей скорости ветра показаны на рис. 2. На этом рисунке точками обозначены экспериментальные значения компонент скорости ветра

(данные заимствованы из работы Орленко [7]), а крестиками — значения, рассчитанные по сплайну.

На практике во многих случаях условия (4) не выполняются. В то же время обычно имеется какая-либо информация о поведении второй производной на концах интервала. Поэтому приведенный в [3] алгоритм построения сплайна $g(x)$ был обобщен на случай ненулевых граничных условий (4):

$$g''(a) = A, \quad g''(b) = B, \quad (4')$$

где A и B — заданные константы.

Для исследования влияния условий (4') нами была проведена серия численных экспериментов. Все полученные в упомянутой выше экспедиции данные по суточному ходу температуры и влажности на различных высотах рассчитывались с граничными усло-

виями (4) и (4'), а константы A и B определялись из конечно-разностного аналога вторых производных по четырем точкам [6]. Анализ расчетов этих вариантов показал, что при интерполировании суточного хода метеозлементов влияние граничных условий (4') сказывается только в трех-четыре ближайших от концов интервала точках и, в частности, значения температуры и влажности в искомый промежуток времени остаются практически одинаковыми. Однако для функций с большой второй производной хотя бы на одном из концов интервала влияние граничных условий (4') распространяется на большее число узловых точек. Этот вывод был нами проверен на тестовом примере для функции e^x , $x \in [0; 5]$.

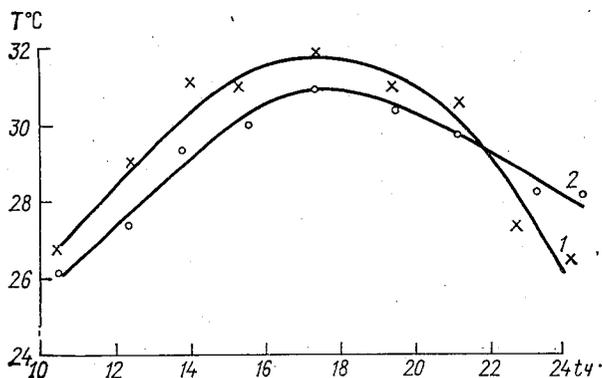


Рис. 3. Суточный ход температуры воздуха на высотах 25 м (1) и 100 м (2).

В том случае, когда экспериментальные данные получены с большим разбросом, целесообразно использовать сглаживающий сплайн $g_1(x)$. Согласно (7), процесс сглаживания существенным образом зависит от значений весовых коэффициентов p_k в выражении (7). Очевидно, что чем больше значение p_k , тем ближе сглаживающая функция подходит к экспериментальным значениям. Поэтому для точек, измеренных с маленькой погрешностью, соответствующие весовые коэффициенты должны быть большими. Однако конкретных и четких рекомендаций относительно выбора значений p_k в литературе по сплайнам в настоящее время нет.

Сглаживающий сплайн был нами использован для расчета кривых суточного хода температуры воздуха по данным вертолетного зондирования. Исходной информацией служили экспериментальные значения температуры на различных высотах за девять временных отсчетов. В данном случае при построении сглаживающего сплайна целесообразно брать постоянное значение весовых коэффициентов ($p_k = p = 0,8$), так как измерения производились на одной высоте. Результаты построения кривой

суточного хода температуры при помощи сглаживающего сплайна приведены на рис. 3. Крестиками и кружками на этом рисунке обозначены данные наблюдений, а сплошная кривая — сглаживающий сплайн $g_1(x)$.

Анализ приведенных выше расчетов и других численных экспериментов показал, что интерполяционный сплайн и сглаживающий сплайн являются хорошим аппаратом приближений и могут успешно применяться при обработке метеорологической информации в пограничном слое атмосферы.

СПИСОК ЛИТЕРАТУРЫ

1. Алберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения. — М.: Мир, 1972. — 316 с.
2. Стечкин С. Б., Субботин Ю. Н. Сплайны в вычислительной математике. — М.: Наука, 1976. — 215 с.
3. Марчук Г. И. Методы вычислительной математики. — Новосибирск: Наука, 1973. — 350 с.
4. Марчук Г. И. Методы вычислительной математики. — М.: Мир, 1977. — 454 с.
5. Вагер Б. Г. Применение сплайнов для приближенного вычисления коэффициента турбулентности. — Труды ГГО, 1977, вып. 382, с. 99—106.
6. Вагер Б. Г. Применение метода матричной факторизации к решению системы дифференциальных уравнений четвертого порядка. — Труды ГГО, 1977, вып. 382, с. 88—98.
7. Орленко Л. Р., Шкляревич О. Б. Некоторые особенности в распределении ветра и температуры в пограничном слое в холодный период. — Труды ГГО, 1973, вып. 297, с. 184—197.

Б. Г. Вагер, Е. Д. Надежина

ОБ ИЗМЕНЕНИИ ВЕРХНЕЙ ГРАНИЦЫ ПЛАНЕТАРНОГО ПОГРАНИЧНОГО СЛОЯ

Толщина турбулентного пограничного слоя и, следовательно, положение его верхней границы тесно связаны со структурой самого слоя. В большинстве модельных задач, посвященных описанию режима течений и теплопередачи в атмосферном пограничном слое, положение верхней границы (h) считается заданным внешним параметром модели. Намечающиеся в настоящее время пути параметризации пограничного слоя требуют решения обратной задачи, а именно определения верхней границы по характеристикам пограничного слоя. Несмотря на объективное наличие перемежаемости вблизи верхней границы пограничного слоя, при обычных обстоятельствах можно считать, что граница между свободным потоком и турбулентным пограничным слоем представляет собой непрерывную поверхность. Мгновенное расстояние от границы свободного потока до подстилающей поверхности является случайной функцией координат и времени. Обычно предполагается, что в условиях квазистационарности и горизонтальной однородности существует некоторый средний (установившийся) режим пограничного слоя и соответствующее ему установившееся положение верхней границы. Именно это положение границы ищется методами теории подобия [2].

Обработка экспериментальных данных по структуре планетарного пограничного слоя (ППС) обычно производится с целью подтверждения (или опровержения) связей, установленных методами теории подобия [3, 4, 6, 10]. Наличие большого количества взаимосвязанных влияний приводит к тому, что даже средний режим ППС не всегда поддается однозначному описанию [3, 6]. Если учесть при этом неполноту и весьма часто встречающуюся статистическую неоднородность экспериментальных данных, становится ясно, что в области обобщения экспериментов трудно добиться получения надежной информации о характеристиках пограничного слоя. Тем не менее опубликованные недавно экспериментальные работы [3, 4, 13] по структуре погранич-

ного слоя содержат интересный материал по распределению турбулентных потоков и по поводу существующих связей между положением верхней границы пограничного слоя и значениями турбулентных потоков у земли.

Современные модели нестационарного пограничного слоя в большинстве своем строятся с учетом изменений верхней границы, вызванных вовлечением не турбулентного устойчиво стратифицированного слоя воздуха, расположенного выше ППС, в область турбулизованного течения [9, 11, 12]. Скорость вовлечения (а тем самым и положение границы) определяется при этом через турбулентные потоки у поверхности на основании соотношений между составляющими баланса энергии турбулентности во всем пограничном слое [11, 12] или вблизи его верхней границы [9, 14, 15].

Существуют попытки отыскания соотношения, определяющего положение верхней границы ППС и в квазистационарном случае [7, 8]. Нам представляется, что эмпирические связи вида $h = f(u_{*0}^2; H_0/\rho c_p)$, где u_* — динамическая скорость, $H_0/\rho c_p$ — приземный поток тепла, могут служить замыкающим звеном в системе уравнений пограничного слоя, позволяющим находить h из решения в числе других неизвестных задачи. Если эти связи достаточно надежны и подкреплены выводами теории подобия то использование их позволит, по-видимому, учесть основные процессы, влияющие на положение границы ППС. С другой стороны, проверяя при моделировании ППС пригодность тех или иных зависимостей $h = f(u_{*0}^2; H_0/\rho c_p)$, мы сможем установить адекватность принятой в модели величины h и экспериментально определяемого уровня, отождествляемого с положением границы слоя.

Рассмотрим квазистационарный пограничный слой над однородной поверхностью. Режим течений и распределение турбулентных потоков в таком слое могут быть рассчитаны в результате решения следующей системы уравнений:

$$\frac{d}{dz} k \frac{du}{dz} + \bar{f}v = 0, \quad (1)$$

$$\frac{d}{dz} k \frac{dv}{dz} - \bar{f}(u - G) = 0, \quad (2)$$

$$\frac{d}{dz} \alpha_\tau k \frac{d\theta}{dz} + \frac{dR}{dz} = 0, \quad (3)$$

$$\frac{d}{dz} \alpha_b k \frac{db}{dz} - \epsilon + k \left[\left(\frac{du}{dz} \right)^2 + \left(\frac{dv}{dz} \right)^2 - \alpha_\tau \frac{g}{\theta} \frac{d\bar{\theta}}{dz} \right] = 0 \quad (4)$$

$$k = l\sqrt{b}, \quad \epsilon = cb^2/k, \quad c = \text{const.} \quad (5)$$

Здесь использованы традиционные обозначения для составляющих средней скорости (u , v), кинетической энергии турбулентности (b), диссипации энергии турбулентности (ϵ), потенциальной

температуры (Θ), отношения коэффициентов турбулентного обмена (α_r, α_b).

Для замыкания системы уравнений (1)—(5) можно использовать какое-либо полуэмпирическое соотношение, определяющее масштаб турбулентности l , или соответствующее уравнение переноса, определяющее либо l , либо ε . Простейшие граничные условия таковы:

$$\text{при } z = z_0 \quad u = v = 0, \quad db/dz = 0, \quad \Theta = \Theta_0 = \text{const}, \quad l = \kappa z_0;$$

$$\text{при } z = h \quad u = G, \quad v = 0, \quad b = 0, \quad \Theta = \Theta_h = \text{const}.$$

Если провести замену переменных $\tilde{z} = \frac{z - z_0}{h - z_0}$ и пронормировать искомые характеристики таким образом:

$$\tilde{\Theta} = \frac{\Theta - \Theta_h}{\Theta_0 - \Theta_h}, \quad \tilde{b} = b/G^2, \quad \tilde{u} = u/G, \quad \tilde{v} = v/G,$$

$$\tilde{\kappa} = \kappa/Gh, \quad \tilde{l} = l/h,$$

то исходные уравнения и граничные условия переписутся в виде:

$$\frac{d}{d\tilde{z}} \tilde{\kappa} \frac{d\tilde{u}}{d\tilde{z}} + \frac{\tilde{f}(h - z_0)}{G} \tilde{v} = 0, \quad (6)$$

$$\frac{d}{d\tilde{z}} \tilde{\kappa} \frac{d\tilde{v}}{d\tilde{z}} - \frac{\tilde{f}(h - z_0)}{G} (\tilde{u} - 1) = 0, \quad (7)$$

$$\frac{d}{d\tilde{z}} \alpha_r \tilde{\kappa} \frac{d\tilde{\Theta}}{d\tilde{z}} + \frac{d\tilde{R}}{d\tilde{z}} = 0, \quad (8)$$

$$\tilde{\kappa} \left[\left(\frac{d\tilde{u}}{d\tilde{z}} \right)^2 + \left(\frac{d\tilde{v}}{d\tilde{z}} \right)^2 - \frac{g}{\Theta} \frac{\alpha_r(\Theta_0 - \Theta_h)h}{G^2} \frac{d\tilde{\Theta}}{d\tilde{z}} \right] - \frac{cb^2}{\tilde{\kappa}} + \alpha_b \frac{d}{d\tilde{z}} \tilde{\kappa} \frac{d\tilde{b}}{d\tilde{z}} = 0 \quad (9)$$

$$\text{при } \tilde{z} = 0, \quad \tilde{u} = \tilde{v} = 0, \quad d\tilde{b}/d\tilde{z} = 0, \quad \tilde{\Theta} = 1, \quad \tilde{l} = \kappa z_0/h; \quad (10)$$

$$\text{при } \tilde{z} = 1, \quad \tilde{u} = 1, \quad \tilde{v} = 0, \quad \tilde{b} = 0, \quad \tilde{\Theta} = 0. \quad (11)$$

Если теперь положить $h \neq \text{const}$ или задавать $h = f(u_{*0}, H_0/\rho c_p)$, то можно решить систему уравнений (6)—(11), применив, например, численный метод решения, изложенный в [1]. При отыскании решения нужно учесть, разумеется, необходимость дополнительных итераций по нелинейным членам, содержащим h .

Естественно исследовать простейшие известные соотношения, определяющие h как функцию турбулентных потоков количества движения и тепла вблизи земной поверхности. В случае нейтральной стратификации почти не вызывает сомнений [10, 11] наличие зависимости вида $h = m u_* / \bar{f}$, где $m = \text{const}$. Величина константы m определена не вполне надежно. По-видимому, наиболее вероятное значение m заключено в пределах от 0,25 до 0,4.

Задавая $h = tu_* \sqrt{f}$, замкнем таким образом систему уравнений (6)–(11) и получим распределение характеристик внутри слоя ($z_0 \leq z \leq h$) в зависимости от числа $Ro = G/\sqrt{f}z_0$. Это дает возможность проверить, при каких условиях с помощью модели можно обеспечить выполнение экспериментальных зависимостей. Так, экспериментально установленная в работе [6] по данным натуральных (аэрологических) наблюдений связь между высотой ППС

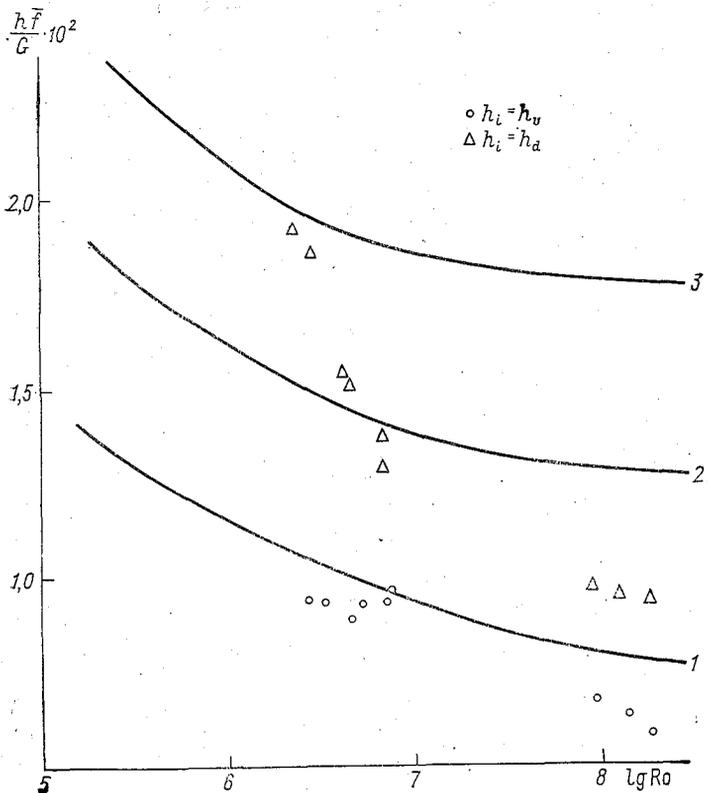


Рис. 1. Зависимость высоты квазистационарного ППС в условиях нейтральной стратификации от числа Ro .
 1) $m=0,1$; 2) $m=0,2$; 3) $m=0,3$.

в условиях нейтральной стратификации и параметром Ro изменяется в зависимости от способа определения h .

На рис. 1 приведены экспериментальные точки из [6], определяющие функцию $h\sqrt{f}/G = f(Ro)$. Эта же зависимость, рассчитанная по модели (6)–(11), представлена на рис. 1 кривыми, полученными для разных значений параметра m . Если h выбирается по экспериментальным данным как уровень, на котором ветер достигает геострофического по направлению ($h = h_d$), то, как

видно из рис. 1, величины h_d количественно близки к значениям h , рассчитанным по модели при m , заключенном в пределах $0,1 \leq m \leq 0,3$. Средняя величина m при этом несколько меньше оценки Дирдорфа [11]. Скорость изменения величины $h_i \bar{f}/G$ с ростом числа Ro , рассчитанная по модели, оказывается меньше, чем скорость изменения $h_d \bar{f}/G$, полученная в [6]. Наклон кривой $h_i \bar{f}/G = f(Ro)$ гораздо ближе к точкам экспериментальной зависимости $h_v \bar{f}/G = \bar{f}(Ro)$ (здесь h_v — уровень, на котором реальный ветер достигает геострофического по модулю). Однако численные значения рассматриваемых величин ($h_i \bar{f}/G$) оказываются близки к экспериментальным только при малых m .

Следует заметить, что значения u_{*0} и, следовательно, высота h , рассчитанные по модели, зависят от способа замыкания системы уравнений пограничного слоя. Приведенные выше оценки получены для случая, когда масштаб турбулентности задавался формулой из [5]:

$$l = -\alpha c^{1/4} \frac{b/k}{\frac{\partial}{\partial z} \left(\frac{b}{k} \right)}. \quad (12)$$

Помимо структуры квазистационарного ППС в условиях нейтральной стратификации наиболее изученной представляется физическая ситуация, определяющая режим ППС в условиях неустойчивой стратификации. Большинство экспериментальных исследований было посвящено изучению именно этого режима [3, 4, 13]. Однако попытки обобщения имеющегося экспериментального материала встречают затруднения, связанные как с неполнотой опубликованных исходных данных, так и с различными методами анализа, используемыми в разных работах.

Рисунок 2 дает некоторое представление об имеющихся в литературе экспериментальных данных по изменчивости h в условиях неустойчивой стратификации. Параметр стратификации на этом рисунке определен независимо от способа нахождения h как $\mu = h_i/L$. Сами же величины h оценены разными способами по профилям ветра [6] или температуры [3] в пограничном слое. Из рисунка видно, что наибольший разброс экспериментальных значений h отмечается в области значений μ , близких к равновесным. Очевидно, для получения надежных экспериментальных связей необходимо значительно увеличить число экспериментальных точек на графике. Заметим, что при оценке величин $h_i \bar{f}/u_*$ и L , использованных для построения рис. 2, не учтены различия в методике

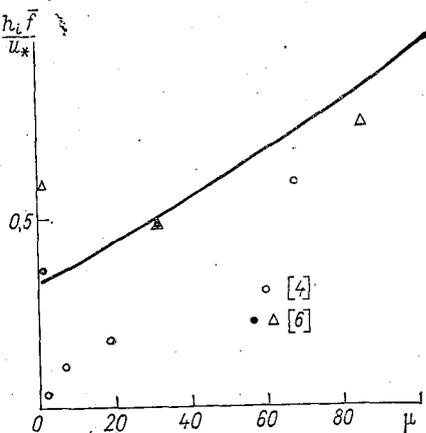


Рис. 2. Зависимость безразмерной высоты пограничного слоя от параметра устойчивости μ .

определения турбулентных потоков у земли и в методике определения h .

Из рассмотрения рис. 2 следует, что в условиях неустойчивой стратификации не сохраняется зависимость вида $h = tu_* / \bar{f}$. Как бы ни было мало количество экспериментальных точек на этом рисунке, все же и этого количества достаточно для того, чтобы заметить, что отношение $h_i \bar{f} / u_*$ растет с увеличением неустойчивости. Если бы выполнялось соотношение $h_i = tu_* / \bar{f}$, то для всех u было бы справедливо равенство $h_i \bar{f} / u_* = m = \text{const}$, и график указанной функции обратился бы в прямую. Вывод, следующий из анализа экспериментальных точек на рис. 2, не подтверждает результата, полученного в работе [10]. При анализе данных натурных экспериментов, выполненном в работе [13], наличие связи $h = tu_* / \bar{f}$ в условиях неустойчивости тоже не подтвердилось. При этом анализе достаточно хорошо оправдалась формула из работы [3], связывающая h с величиной приземного потока тепла,

$$h = a \left(\frac{\beta H_0}{\rho c_p \bar{f}^3} \right)^{1/2}, \quad (13)$$

где $H_0 / \rho c_p = -(\overline{w' \Theta'})$, $a = \text{const}$, $\beta = g / \Theta$, w — вертикальная скорость.

Выражение (13) было использовано в модели ППС для расчета характеристик пограничного слоя в условиях неустойчивой стратификации. Результат расчета представлен на рис. 2 кривой, которая близка к верхней границе поля экспериментальных точек. Очевидно, однако, что экспериментальные связи нуждаются в существенном уточнении. В этом плане весьма перспективным представляется использование материалов по фактическому распределению диссипации кинетической энергии турбулентности в пограничном слое атмосферы. Одновременное использование модели ППС, в которую включено уравнение переноса для ϵ , позволит расширить наши представления об особенностях формирования пограничного слоя в различных условиях.

СПИСОК ЛИТЕРАТУРЫ

1. Вагер Б. Г., Надежина Е. Д. Численные эксперименты по расчету горизонтально-неоднородного планетарного пограничного слоя. — Труды ГГО, 1974, вып. 326, с. 15—23.
2. Зубковский С. Л., Кухарец В. П., Цванг Л. Р. Вертикальные профили характеристик турбулентности в приземном и пограничном слоях атмосферы при неустойчивой стратификации. — Изв. АН СССР. Физика атмосферы и океана, 1979, т. 15, № 1, с. 44—52.
3. Океанология. Физика океана. Т. 1. Гидрофизика океана/Под ред. А. С. Монина. — М.: Наука, 1978. — 456 с.
4. Кухарец В. П., Цванг Л. Р. О параметризации турбулентных потоков тепла в неустойчиво стратифицированном пограничном слое атмосферы. — Изв. АН СССР. Физика атмосферы и океана, 1976, т. 12, № 1, с. 13—20.

5. Кухарец В. П., Цванг Л. Р. О скорости диссипации турбулентной энергии в неустойчиво стратифицированном атмосферном пограничном слое.—Изв. АН СССР. Физика атмосферы и океана. 1977, т. 13, № 6, с. 620—628.
6. Орленко Л. Р. Строение планетарного пограничного слоя атмосферы. Автореф. дисс. на соискание учен. степени д-ра физ.-мат. наук.—Л., 1975.
7. Симонов В. В. Энергетический метод оценки высоты стационарного нейтрально-стратифицированного пограничного слоя атмосферы.—Труды ГГО, 1977, вып. 382, с. 116—123.
8. Симонов В. В. Об определении высоты стационарного, стратифицированного слоя атмосферы.—Труды ГГО, 1977, вып. 398, с. 59—68.
9. Bradshaw P., Ferris D. H., Atwell N. P. Calculation on boundary layer development using the turbulent energy equation.—J. Fluid. Mech. 1967, vol. 28, p. 2, p. 593—618.
10. Clarke R. H., Hess G. D. On the appropriate scaling for velocity and temperature in the planetary boundary layer.—J. Atm. Sci., 1973, vol. 30, N 7, p. 1346—1355.
11. Deardorff J. W. Three-dimensional numerical study of the height and mean structure of a heated planetary boundary layer.—Bound. Lay. Met. 1974, vol. 7, p. 81—106.
12. Stull R. B. The energetics of entrainment across a density interface.—J. Atm. Sci, 1977, vol. 33, N 7, p. 1260—1267.
13. Structure of the atmospheric boundary layer derived from airborne measurements of the energy dissipation rate ϵ/M . Gamo, O. Yokoyama, S. Yamamoto, Y. Mitsuta.—J. Meteor. Soc., Japan, 1976, ser. 11, 54, N 4, p. 75—84.
14. Tennekes H. A model for the dynamics of the inversion above a convective boundary layer.—J. Atm. Sci, 1973, vol. 30, p. 558—567.
15. Zeman O., Tennekes H. Parametrization of the turbulent energy budget at the top of the day-time atmospheric boundary layer.—J. Atm. Sci, 1977, vol. 34, N 1, p. 111—123.

К АНАЛИЗУ КОРРЕЛЯЦИОННОЙ СТРУКТУРЫ СЕЗОННЫХ ВРЕМЕННЫХ РЯДОВ ТЕМПЕРАТУРЫ

Настоящая работа посвящена анализу корреляционной структуры временных рядов температуры воздуха с целью определения статистической значимости полученных оценок.

В работах [1, 2] показано, что временные ряды температуры (среднегодовые, сезонные) имеют очень небольшую коррелированную компоненту и основная часть сезонной изменчивости приходится на белый шум. Представляло интерес для временных рядов с наибольшей продолжительностью, во-первых, оценить статистическую значимость значений коэффициента корреляции средних многолетних значений температуры для запаздывания в один год и, во-вторых, попытаться определить, является ли статистически значимым различие корреляционной структуры отдельных сезонов. Исследования основаны на критерии, описание которого дано в книге Рао [2]. Возможность его применения обоснована тем фактом, что временные ряды имеют очень большой уровень белого шума. Поскольку имеется хоть небольшая коррелированная компонента, полученные результаты имеют приближенный характер.

В работе применялся критерий, связанный с преобразованием оценки коэффициента корреляции по формуле

$$\sqrt{n_i - 3}(z_i - E(z_i)), \quad (1)$$

где

$$z = \frac{1}{2} \ln \frac{1+r_i}{1-r_i}, \quad (2)$$

$$E(z)_i = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n_i-1)}, \quad (3)$$

r — исследуемый коэффициент корреляции, ρ — коэффициент корреляции генеральной совокупности, принимаемый в данном случае за нуль; n — объем выборки.

В табл. 1 приведены коэффициенты корреляции, полученные по данным сезонных рядов температуры.

В табл. 2 приведены величины, полученные по формуле (1), показывающие статистическую значимость рассчитанных коэффициентов корреляции. (Величинам, которым будут соответствовать значения с 95 %-ным уровнем, соответствуют значения табл. 2 больше двух.) Из табл. 2 следует, что в зимний период нет ни одного значимого коэффициента корреляции, в весенний и летний периоды имеется по 4 значимых коэффициента корреляции и в осенний период — 2 коэффициента.

Таблица 1

Статистические оценки корреляционных функций сезонных рядов температуры для запаздывания в один год

Станция	Период наблюдений, годы	Зима	Весна	Лето	Осень
Базель	1755—1957	-0,02	0,18	0,06	-0,04
Будапешт	1780—1960	-0,01	-0,01	0,22	0,04
Вена	1775—1960	0,01	0,17	0,16	0,07
Вильно	1781—1971	0,04	0,06	0,19	-0,03
Копенгаген	1800—1962	0,12	0,10	0,11	0,24
Ленинград	1805—1975	0,03	0,22	0,14	-0,01
Центральная Англия	1700—1972	-0,01	0,14	0,03	0,14

Итак, несмотря на то, что найденные оценки коэффициентов корреляции очень малы и основную роль в изменчивости анализируемых рядов играет белый шум, в ряде случаев эти оценки оказались статистически значимы, особенно в весенне-летний период. Полученные оценки подтверждают результаты, найденные

Таблица 2

Оценки корреляционных функций сезонных рядов температуры на статистическую значимость по формуле (1)

Станция	Зима	Весна	Лето	Осень
Базель	-0,3	2,6	0,9	-0,6
Будапешт	-0,1	-0,1	3,0	0,5
Вена	0,1	2,3	2,2	1,0
Вильно	0,5	0,7	2,7	-0,4
Копенгаген	1,5	1,3	1,4	3,1
Ленинград	0,3	2,9	1,8	-0,1
Центральная Англия	-0,1	2,4	0,5	2,3

Мадденом [3] с помощью рядов значительно меньшей продолжительности и меньшей точности, о том, что в летний период наиболее вероятны выявления «слабого сигнала», который содержится во временных рядах, так как в этот период изменчивость временных рядов температуры наименьшая. (Годовой ход стандартов норм температуры имеет минимум в летний период.)

В табл. 3 содержатся коэффициенты корреляции среднегодовых значений и величины, рассчитанные по формуле (1). Из семи коэффициентов три оказались статистически значимы с 95 %-ным уровнем. Заметим, что оценки статистически значимы для тех временных рядов среднегодовых наблюдений, для

Таблица 3

Коэффициенты корреляции (r) среднегодовых значений температуры и их оценки на статистическую значимость

Станция	r	$\sqrt{n-3}(z-E(z))$
Базель	0,15	2,1
Будапешт	0,01	0,1
Вена	0,12	1,7
Вильно	0,10	1,3
Копенгаген	0,02	0,3
Ленинград	0,19	2,6
Центральная Англия	0,15	2,5

которых они статистически значимы в весенний период. Следовательно, температурные особенности весеннего периода оказывают большее влияние на коэффициенты корреляции среднегодовых значений.

Таким образом, для сезонных рядов статистическая значимость коэффициентов корреляции имеет место в весенне-летний период и отсутствует в зимний. Более конкретные выводы можно будет сделать с помощью аналогичного анализа больших архивов наблюдений.

СПИСОК ЛИТЕРАТУРЫ

1. Поляк И. И. Численные методы анализа наблюдений. Л.: Гидрометеоиздат, 1975. — 212 с.
2. Рао С. Р. Линейные статистические методы и их применение. — М.: Наука, 1968.
3. Madden R. A. Estimates of the autocorrelations and spectra of seasonal mean temperatures over North America.— Mon. Wea. Rev., 1977, vol. 105, N 1.

СРАВНЕНИЕ СПЛАЙНОВЫХ И КОНЕЧНО-РАЗНОСТНЫХ РЕШЕНИЙ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ НА ПРИМЕРАХ ПРОСТЕЙШИХ ТЕЧЕНИЙ В ПОГРАНИЧНОМ СЛОЕ

Достигнутые в последние годы успехи в развитии математической теории сплайн-функций открыли широкие возможности для их практической реализации во многих разделах вычислительной математики [1, 4, 7]. Наиболее успешно сплайны применяются в теории приближений при решении задач интерполяции и экстраполяции. Довольно быстро проникают сплайны и в такой раздел вычислительной математики, как численное интегрирование дифференциальных уравнений.

Существуют два способа использования сплайнов при численном интегрировании дифференциальных уравнений. Первый состоит в том, что сначала находится с помощью конечно-разностных методов сеточное решение дифференциального уравнения, а затем строится интерполяционный или сглаживающий сплайн. Во втором случае решение сразу ищется в виде сплайна. Интерес к сплайновым решениям дифференциальных уравнений объясняется тем, что они соединяют одновременно идеи численных и аналитических методов. Действительно, при нахождении сплайновых решений вводится сетка и часто задача сводится к решению линейной алгебраической системы уравнений, как и в конечно-разностном методе, а с другой стороны, между узлами сетки получается полиномиальная аппроксимация искомой функции, что очень удобно, например, при нахождении функционалов от полученного приближенного решения.

В настоящей работе исследуется возможность использования кубических сплайнов при численном интегрировании дифференциальных уравнений, описывающих процессы в пограничном слое, и сплайновые решения этих уравнений сравниваются с конечно-разностными. В качестве основы для такого исследования был выбран алгоритм построения приближенного решения краевой задачи для дифференциального уравнения в виде кубического сплайна, предложенный в работе [3]. Одно из преимуществ этого

алгоритма по сравнению с другими методами нахождения сплайновых решений состоит в том, что он требует проведения всего лишь $O(N)$ арифметических действий, где N — количество узлов сетки.

Рассмотрим согласно [3] краевую задачу для линейного дифференциального уравнения второго порядка:

$$Ly \equiv \frac{d^2y}{dx^2} + p(x) \frac{dy}{dx} + q(x)y - f(x) = 0, \quad x \in [a, b], \quad (1)$$

$$x = a: \quad \alpha_a y + \beta_a \frac{dy}{dx} = \gamma_a, \quad x = b: \quad \alpha_b y + \beta_b \frac{dy}{dx} = \gamma_b. \quad (2)$$

Приближенное решение задачи (1), (2) будем искать в виде кубического сплайна $v(x)$. Для этого введем на интервале $[a, b]$ произвольную сетку $a = x_0 < x_1 < \dots < x_N < x_{N+1} = b$ и на каждом подынтервале $[x_i, x_{i+1}]$ зададим следующий вид сплайнового решения:

$$v_i(x) = v_{0i} + v_{1i}(x - x_i) + \frac{v_{2i}}{2}(x - x_i)^2 + \frac{v_{3i}}{6}(x - x_i)^3, \quad x \in [x_i, x_{i+1}], \quad (3)$$

где $v_{0i}, v_{1i}, v_{2i}, v_{3i}$ — пока неопределенные коэффициенты сплайна. Очевидно, что для их определения требуется $4N+4$ условия. Первые $3N$ уравнений получаются непосредственно из самого определения кубического сплайна как непрерывной функции на всем интервале, обладающей двумя первыми непрерывными производными [1]. Следовательно, во всех внутренних узлах сетки должны выполняться равенства:

$$v_{i-1}(x_i) = v_i(x_i), \quad v'_{i-1}(x_i) = v'_i(x_i), \quad v''_{i-1}(x_i) = v''_i(x_i), \quad i = 1, 2, \dots, N \quad (4)$$

Еще $N+2$ уравнения добавляются из условия удовлетворения функции $v(x)$ уравнению (1) во всех узлах сетки. Два последних уравнения, необходимых для замыкания системы, записываются из граничных условий (2). Полученная таким образом система алгебраических уравнений оказывается линейной. Как показано в [3], путем элементарных преобразований удастся выразить коэффициенты v_{1i}, v_{2i}, v_{3i} через v_{0i} , а для вычисления последних получается система уравнений с трехдиагональной матрицей, которая легко решается методом прогонки.

Подробный алгоритм нахождения всех коэффициентов сплайна приведен в [3]. В этой же работе получены оценки погрешности приближенного решения, а именно, если обозначить через $y(x)$ точное решение краевой задачи (1), (2) и ввести функцию $z(x) = y(x) - v(x)$, то справедливы оценки:

$$\max_i |z_i| \leq c_0 h^3, \quad \max_i |z'_i| \leq c_1 h^2, \quad \max_i |z''_i| \leq c_2 h^2, \quad \max_i |z'''_i| \leq c_3 h, \quad (5)$$

где h — максимальный шаг сетки, а c_0, c_1, c_2, c_3 — не зависящие от h константы.

Для исследования возможности использования описанного выше метода в задачах пограничного слоя рассмотрим в качестве примера ряд простейших задач, допускающих точные решения. В каждой из этих задач кроме сплайнового решения для сравнения приводятся аналитическое и конечно-разностное решения, причем при интегрировании методом сеток используется обычная конечно-разностная схема [6], имеющая второй порядок точности и сводящая краевую задачу (1), (2) к решению системы:

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} + p_i \frac{y_{i+1} - y_{i-1}}{2h} + g_i y_i = f_i, \quad i = 1, 2, \dots, N, \quad (6)$$

$$\alpha_a y_0 + \beta_a \frac{y_1 - y_0}{h} = \gamma_a, \quad \alpha_b y_{N+1} + \beta_b \frac{y_{N+1} - y_N}{h} = \gamma_b.$$

Подробная постановка первых трех задач и используемые в них обозначения приведены в гл. 5 монографии Шлихтинга [8], а последней задачи — в учебнике [2]. Мы же ограничимся лишь записью самих дифференциальных уравнений, граничных условий и точных решений этих задач.

Пример 1. Течение Куэтта.

$$\frac{dp}{dx} = \mu \frac{d^2 u}{dz^2}, \quad u|_{z=0} = 0, \quad u|_{z=H} = U,$$

$$u(z) = \frac{z}{H} U - \frac{H^2}{2\mu} \frac{dp}{dx} \frac{z}{H} \left(1 - \frac{z}{H}\right).$$

Пример 2. Течение Пуазейля.

$$\mu \left(\frac{d^2 u}{dz^2} + \frac{1}{z} \frac{du}{dz} \right) = \frac{dp}{dx}, \quad u|_{z=R} = 0, \quad u|_{z=-R} = 0,$$

$$u(z) = -\frac{1}{4\mu} \frac{dp}{dx} (R^2 - z^2).$$

Пример 3. Вращающиеся цилиндры.

$$\rho \frac{u^2}{r} = \frac{dp}{dr}, \quad \frac{d^2 u}{dr^2} + \frac{d}{dr} \left(\frac{u}{r} \right) = 0, \quad u|_{r=r_1} = r_1 \omega_1, \quad u|_{r=r_2} = r_2 \omega_2,$$

$$u(r) = \frac{1}{r_2^2 - r_1^2} \left[r(\omega_2 r_2^2 - \omega_1 r_1^2) - \frac{r_1^2 r_2^2}{r} (\omega_2 - \omega_1) \right].$$

Пример 4. Уравнение теплопроводности.

$$\frac{d}{dz} a_\tau k \frac{d\theta}{dz} = -\frac{P_0}{\rho c_p}, \quad T|_{z=H} = T_H, \quad T|_{z=0} = T_0,$$

$$T(z) = T_H + \gamma_a (H - z) + \frac{P_0}{\rho c_p k} \frac{(H - z)^2}{2H}.$$

Промежуток интегрирования всех четырех задач путем соответствующей нормировки сводился к интервалу $[0, 1]$, а в пер-

вых трех примерах аналогичным образом граничные условия приводились к виду: $y(0)=0$, $y(1)=1$. Число узлов сетки выбиралось равным 20, что при однородной сетке соответствует шагу $h=0,05$. Все решения получены как в однородной, так и в неоднородной сетке. Результаты расчетов представлены в табличной форме, что позволяет провести полное сравнение методов интегрирования. Для всех примеров в таблице для четырех значений независимой переменной ($z=0,2$, $z=0,4$, $z=0,6$, $z=0,8$)

Сравнение конечно-разностного и сплайнового решений с точным решением

Номер примера	y_k	v	y	$z_k= y-y_k $	$z_c= y-v $
1	-0,119996	-0,120000	-0,120000	$0,4 \cdot 10^{-5}$	0
	-0,079997	-0,080000	-0,080000	$0,3 \cdot 10^{-5}$	0
	0,120001	0,120000	0,120000	$0,1 \cdot 10^{-5}$	0
	0,479999	0,480000	0,480000	$0,1 \cdot 10^{-5}$	0
	0,039987	0,039994	0,040000	$0,13 \cdot 10^{-4}$	$0,6 \cdot 10^{-5}$
2	0,159983	0,159993	0,160000	$0,17 \cdot 10^{-4}$	$0,7 \cdot 10^{-5}$
	0,359983	0,359993	0,360000	$0,17 \cdot 10^{-4}$	$0,7 \cdot 10^{-5}$
	0,639988	0,639995	0,640000	$0,12 \cdot 10^{-4}$	$0,5 \cdot 10^{-5}$
	0,229596	0,229668	0,229648	$0,52 \cdot 10^{-4}$	$0,20 \cdot 10^{-4}$
3	0,438054	0,438142	0,438117	$0,63 \cdot 10^{-4}$	$0,25 \cdot 10^{-4}$
	0,633297	0,633373	0,633349	$0,52 \cdot 10^{-4}$	$0,24 \cdot 10^{-4}$
	0,819730	0,819774	0,819761	$0,31 \cdot 10^{-4}$	$0,13 \cdot 10^{-4}$
4	2,483798	2,483846	2,483838	$0,40 \cdot 10^{-4}$	$0,8 \cdot 10^{-5}$
	1,984607	1,984645	1,984659	$0,52 \cdot 10^{-4}$	$0,14 \cdot 10^{-4}$
	1,570913	1,570939	1,570959	$0,46 \cdot 10^{-4}$	$0,20 \cdot 10^{-4}$
	1,242713	1,242723	1,242740	$0,27 \cdot 10^{-4}$	$0,17 \cdot 10^{-4}$

приведены соответствующие значения конечно-разностного (y_k), сплайнового (v) и точного (y) решений, а также разности между приближенным и точным решениями (z_k и z_c).

Как видно из таблицы, сплайновые решения даже на простых линейных задачах оказываются точнее конечно-разностных. Такой же вывод получен и в работах других авторов. Так, например, Рубин и Хосла [5, 9] показали, что сплайн-процедура при решении уравнения диффузии обладает большей точностью по сравнению с конечно-разностными схемами, что связано с более высоким порядком точности сплайн-аппроксимации конвективных и диффузионных членов.

Если еще учесть, уже отмечавшееся преимущество сплайнового решения по сравнению с конечно-разностным, то можно наде-

яться, что при численном интегрировании нелинейных задач, описывающих процессы в пограничных слоях, применение сплайнов будет еще эффективнее.

СПИСОК ЛИТЕРАТУРЫ

1. Алберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее применения. — Мир, 1972. — 316 с.
2. Динамическая метеорология. — Л.: Гидрометеоиздат, 1976. — 607 с.
3. Ильин В. П. О сплайновых решениях обыкновенных дифференциальных уравнений. — Журнал вычислительной математики и математической физики, 1978, т. 18, № 3, с. 620—627.
4. Марчук Г. И. Методы вычислительной математики. — М.: Наука, 1976. — 456 с.
5. Рубин С. Г., Хосла П. К. Численные решения повышенной точности, использующие кубические сплайны. — Ракетная техника и космонавтика, 1976, т. 14, № 7, с. 25—34.
6. Самарский А. А. Теория разностных схем. — М.: Наука, 1977. — 656 с.
7. Стечкин С. Б., Субботин Ю. П. Сплайны в вычислительной математике. — М.: Наука, 1976. — 215 с.
8. Шлихтинг Г. Теория пограничного слоя. — М.: Наука, 1969. — 735 с.
9. Rubin S. G., Khosla P. K. Higher order numerical methods derived from three-point polynomial interpolation. General App. Sci. Labs., GASL TR 228, 1976.

А. И. Полищук, Т. Н. Юринова

АРХИВАЦИЯ ДАННЫХ В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ ОБРАБОТКИ ИНФОРМАЦИИ О ЗАГРЯЗНЕНИИ АТМОСФЕРЫ (АСОИЗА)

1. При решении многих научных и практических задач, связанных с охраной воздушного бассейна городов от загрязнения, возникает необходимость обращения к материалам наблюдений за длительные промежутки времени, в том числе и за прошлые годы. Особая потребность в большом объеме данных появляется для задач статистического характера. При этом желательно располагать сведениями, занесенными на технический носитель, что позволяет воспользоваться известными алгоритмами статистического анализа и программами их реализации на ЭВМ (см., например, [5, 8]).

Другой областью использования материала массовых наблюдений можно считать информационное обслуживание различных народнохозяйственных задач. В настоящее время подобные вопросы решаются путем создания автоматизированных информационно-справочных систем (АИСС), с применением средств вычислительной техники. Для функционирования АИСС требуется наличие информационной базы системы на технических носителях.

В связи с перечисленными выше задачами очевидна необходимость размещения материалов наблюдений на таком техническом носителе, который отвечает требованиям долговременного хранения данных и допускает многократное обращение к массиву данных. В системе гидрометеорологического обслуживания в качестве таких носителей рассматриваются магнитные ленты МЛ [4].

2. Информационное содержание АИСС загрязнения воздуха составляют материалы наблюдений на общегосударственной сети контроля загрязнения атмосферы (СКЗА). В настоящее время СКЗА организована более чем в 250 городах и включает около 1000 постов наблюдений. Поток первичной информации от СКЗА состоит из сведений о содержании примесей в воздухе, дан-

ных сопутствующих метеорологических измерений и результатов аэрологических наблюдений. Специфика режима работы СКЗА приводит к тому, что объем и состав информации за месяц от каждого поста измерения концентраций примесей и от каждого вида наблюдений не являются постоянными и колеблются в довольно широких пределах (табл. 1). Наибольший объем информации поступает от одного поста в городе, для которого вместе с данными о концентрациях примесей сообщаются сведения об аэрологических наблюдениях и о наблюдениях на метеостанции. (Далее этот пункт условно называется «метеостанция»). Для других постов СКЗА сведения о концентрациях примесей сопровождаются меньшим числом метеопараметров.

Таблица 1

Структура информационных потоков

Информация	Объем за месяц
Дискретные наблюдения:	
на стационарных постах	1,5 . . . 15 тыс. байтов по посту
маршрутные	800 . . . 4,2 тыс. байтов по посту
подфакельные	800 . . . 8 тыс. байтов под одним факелом
Непрерывные наблюдения:	
автоматические газоанализаторы	4,2 тыс. байтов по одному посту

Дальнейшее развитие СКЗА предполагается в направлении как увеличения числа постов измерений, так и в расширении перечня измеряемых параметров. Широкое распространение получают автоматические и полуавтоматические средства измерения содержания примесей в атмосфере. Таким образом, объем информации, поступающей от СКЗА, будет возрастать. Естественно, что такой объем данных целесообразно анализировать и обобщать с помощью ЭВМ.

3. В течение ряда последних лет в Главной геофизической обсерватории им. А. И. Воейкова разрабатывалась на базе ЭВМ М-222 автоматизированная система обработки информации о загрязнении воздуха (АСОИЗА), которая предусматривает подготовку перфолент с данными наблюдений СКЗА в городах, где эти наблюдения проводятся, передачу их в центр для контроля, обработки, обобщения и архивации информации на МЛ.

Вопросы построения АСОИЗА, принципы, заложенные при конкретной реализации схем машинной обработки, освещены

в [7]. Как и предполагалось ранее, комплекс автоматизированной обработки по мере его освоения и внедрения расширился и усовершенствовался. При этом выявилась необходимость разделить процесс машинной обработки на четыре самостоятельных этапа — подготовительный, обработки, архивации, обобщения. На каждом из них циклически обрабатывается определенный объем информации, которая записывается на магнитную ленту (МЛ). Такое разделение оказалось целесообразным с точки зрения распределения трудовых ресурсов и времени ЭВМ. В настоящее время система состоит из шести основных программ и двадцати вспомогательных (сервисных), обеспечивающих взаимосвязанное действие основных программ и облегчающих эксплуатацию системы.

Известно, что информация, поступающая на вход системы, содержит ошибки различного характера, для выявления которых в систему введен синтаксический и семантический контроль. Устранение обнаруженных ошибок выполняется в рамках системы с помощью корректирующих перфоленг, вводимых при необходимости. Повторение обработки позволяет получить на выходе информацию, содержащую надежные сведения. Основными выходными документами системы являются таблицы стандартных форм, а также магнитные ленты архивного типа.

Технология обработки допускает ввод информации отдельных постов наблюдений в любой последовательности независимо от принадлежности к конкретному городу или региону. Упорядочивание массивов и организация их в определенном порядке выполняется программно только при архивации данных. Это создает удобство при эксплуатации системы обработки массового материала наблюдений. Поскольку все этапы обработки завершаются записью информации на МЛ, то на каждой из них организованы «информационные зоны» (ИЗ), содержащие сведения о наличии информации отдельных постов СКЗА на МЛ, состоянии обработки, корректировке ошибок, на них также фиксируется переход к следующему этапу. Так, по завершении второго этапа — обработки — проконтролированная первичная информация, а также отдельные результаты расчетов заносятся на МЛ промежуточного архива (МЛарп). После анализа качества машинной обработки, который в настоящей технологической схеме выполняется вручную, на МЛарп заносится признак «разрешение архивации» массива данных отдельного поста наблюдений. Этот же признак вырабатывается после повторения обработки в случае выявленных ошибок.

4. Рассмотрим подробнее архивацию данных как этап машинной обработки в АСОИЗА.

Та информация, которая не содержит ошибок и не вызывает сомнений, записывается на две отдельные архивные магнитные ленты (МЛар). При этом на одну из них заносятся проконтролированные данные наблюдений СКЗА (МЛар «срочные»), на другую — результаты статистической обработки (МЛар «средние»).

В основу архивации на МЛ данных о загрязнении атмосферы положены разработки, выполненные Л. П. Афиногеновым и под его руководством в ГО [1—3], а также во ВНИИГМИ—МЦД [4]. В связи с тем что информация, поступающая от СКЗА, имеет переменный объем, не удалось полностью принять рекомендации ВНИИГМИ—МЦД, ориентированные на массивы данных постоянной длины. Кроме того, в АСОИЗА предусмотрена запись на МЛар сведений, регулярно поступающих с СКЗА. Следовательно, заранее определить объем и место информации по каждому городу (пункту наблюдений), как это принято при ар-

Таблица 2

Структура архивов на МЛ

Номер зоны	Характер информации	Количество разрядов в слове	Номер зоны	Характер информации	Количество разрядов в слове	
1	Резерв	45	1	Резерв	45	
2	ГЗЛ „Срочные“	45	2	ГЗЛ „Средние“	45	
3		45	3		45	
4		45	4		45	
5		45	5		45	
6		45	6		45	
7	ГЗФ	45	7	ГЗФ	45	
8	Месяц	45	8	Год	45	
9		45	9		45	
10		Метеостанция	32		10	ГЗГ—город
11	Пост 1	33 П	11	Месяц I	32	
	Пост 2		11	Месяц II		
		12	Месяц III		
		12	Месяц IV		
		13	Месяц V		
		13	Месяц VI		
	Пост PP-2		32	14	
	Пост PP-1		15	
	16		Месяц XI
	16		Месяц XII
	17		ГЗГ—город
.....	18	Месяц I	32		
.....	18	Месяц II			
.....		
1777	1777	

хивации иных гидрометеорологических данных, становится невозможным. Поэтому в систему, рекомендуемую ВНИИГМИ—МЦД, пришлось внести изменения, учитывающие специфику данных СКЗА. При записи на МЛар информация программно комплексуется в пространственно-временной архив, который пополняется информацией в течение всего года.

Структуры МЛар «срочные» и «средние» представлены в табл. 2, из которой видно, что они отличаются по типу, объему и способу комплектации данных. Структуры этих МЛар отличаются и от структур, предложенных ВНИИГМИ—МЦД. Здесь запись ведется на заранее защищенную ленту, т. е. уже на МЛар. Кроме того, в головную зону ленты (ГЗЛ) заносится дополнительная информация, необходимая для формирования архива и последующего его использования, а в головную зону файла

Таблица 3

Структура ГЗЛ МЛар „срочные“

Адрес массива в зоне	Название массива	Пояснение
1	МЕТ [1:16]	Метка ленты
21	Ti [1:1]	Таблица разметки ленты
24	Tit [1:1, 1:30]	Каталог файлов
85	KML [1:1, 1:6]	Каталог МЛ
105	KM [1:30]	Массив переводных множителей для метеоэлементов
141	KK [1:45]	Массив переводных множителей для концентраций
203	KСЛ [1:1400]	Словарь станций

(ГЗФ) заносится каталог, отражающий пространственно-временную структуру МЛар. На МЛар «средние» организованы дополнительные служебные зоны — головные зоны каждого города (ГЗГ), это хотя и увеличивает объем архива, но представляет значительные удобства при поиске конкретной архивной информации (по отдельным городам, постам, месяцам) в АИСС. Головные зоны позволяют идентифицировать конкретную информацию в архиве, дают количественные, временные и пространственные характеристики информации. В служебных зонах на МЛар (ГЗЛ, ГЗФ, ГЗГ) используются 45-разрядные слова, а в зонах записи собственно информации — 32-разрядные слова. При переводе архива с ЭВМ М-222 на ЭВМ ЕС информация с 32-разрядными словами переписется практически без изменения.

5. Структура записи МЛар «срочные» такова, что позволяет на одну МЛар «срочные» записать информацию по наблюдениям за два месяца на всех постах СКЗА Советского Союза, т. е. за год таких МЛ должно быть шесть. Для записи данных за один

месяц отводится один файл со своей ГЗЛ, который допускает запись информации по 1000 постам СКЗА. В ГЗЛ по сравнению с разработками ВНИИГМИ—МЦД дополнительно внесены массивы переводных множителей и словаря станций (табл. 3).

До записи информации в ГЗЛ заносится массив МЕТ [1:16]—метка ленты, определяющий содержание головной зоны ленты, которая в свою очередь содержит массивы, идентифицирующие структуру разметки МЛ и структуру файлов. Адрес каталога разметки—ячейка 21—указывает место на МЛ, где записана таблица разметки МЛ Ti [1:1], содержащая полную информацию о разметке ленты.

Таблица 4

Структура ГЗФ МЛар „срочные“

Адрес массива в зоне	Название массива	Пояснение
1	Title [1:30]	Массивы, используемые при упаковке информации
31	fx [1:37]	
68	qx [1:37]	
105	rx [1:37]	
148	kt [1:7]	
156	KAт [1:800]	Бинарный каталог наличия на МЛ информации по каждому городу
		Каталог записанных на МЛар городов

Структура и объем данных, записанных на МЛар «срочные», определяется в ГЗФ, как видно из табл. 4, массивом Title [1:30]. Этот же массив характеризует и содержание самой ГЗФ. Заметим здесь, что в ГЗФ имеется дополнительный массив—каталог записанных на МЛар городов, определяющий наличие, количество и местонахождение на ленте информации по каждому записанному городу.

Бинарная таблица (массив kt) используется для фиксации наличия информации по каждому городу и пополняется, если поступили на МЛар сведения хотя бы одного (любого) поста СКЗА в этом городе. Массив КАт—каталог записанных городов—пополняется по мере поступления информации.

При записи информации первого поступившего на МЛар поста СКЗА данного города рассчитывается количество зон, необходимое для записи со всех постов этого города. Эти зоны резервируются. Затем рассчитывается номер первой свободной зоны для записи информации следующего города. В первую зону, отведенную для данного города, заносится информация по метеостанции, поскольку объем этой информации в 2 раза больше, чем для других постов. В последующие зоны по мере поступления информации записываются данные наблюдений по двум постам в зоне.

Перед записью информация отдельного поста СКЗА упако-

а)

32		30		20				10			
JJ		GG		УУУ				rrr ₁			
φ				rrr ₂				rrr ₃			
ff				rrr ₄				rrr ₅			
eee				rrr ₆				rrr ₇			
		dd		vv				± TTT			
τ	NN	nn	K ₁				T ₁ T ₁		T ₂ T ₂		
dd ₁		vv ₁		dd ₂		vv ₂		dd ₃			
		vv ₃		dd ₄		vv ₄		vv ₅			
		H ₁ H ₁ H ₁		H ₂ H ₂ H ₂		H ₃ H ₃ H ₃					
				± T _{0,5} T _{0,5}		± T _{1,0} T _{1,0}		± T _{2,0} T _{2,0}			

б)

Статистические характеристики

δδ		N				Q̄			
		S				v			
Q _{макс}		G				G1			
		M		M1		M2			

Характеристики суммарного действия

02+10		02+05		02+13	
02+20		02+05+19		08+09	
02+04		02+10		02+05	
02+13		02+20		02+05+19	
		08+09		02+04	

Рис. 1. Макеты упаковки информации при записи в архив на МЛ.
 а) МЛар «срочные»; б) МЛар «средние» (для характеристик суммарного действия указаны шифры примесей в соответствии с [6]).

ывается по несколько слов в ячейку согласно макету упаковки, представленному на рис. 1. Обозначения на этом рисунке соответствуют принятым в [6]. Информация в зоне записи данных на МЛар «срочные» содержит служебный массив для отдельного поста СКЗА, который дублирует ту часть массива KAt, которая относится к данному посту и определяет качественный и количественный состав информации. Далее помещается упакованная в соответствии с массивами упаковки, описанными в ГЗФ, информация по данному посту за все сроки наблюдений, количество которых указывается в служебном массиве. Для записи информации «метеостанции» отводится целая зона, а для информации одного поста — половина зоны.

6. Результаты статистической обработки данных наблюдений на отдельных постах за весь год по всей территории Советского Союза заносятся на одну МЛар «средние». Для записи архива отведен один файл (см. табл. 2). Разметка МЛар «средние» такая же, как и МЛар «срочные», в ГЗЛ размещается аналогичная информация, отсутствует только массив КМ. Изменяется содержание массива Tit, который в данном случае состоит из одного массива Title [1 : 30] (по количеству файлов).

В ГЗФ записываются массив Title, массивы, описывающие упаковку данных, и формируется бинарная таблица наличия информации по отдельным городам за каждый месяц. В бинарной таблице для каждого месяца фиксируется наличие или отсутствие на МЛар информации от каждого города за каждый месяц (поступили ли данные хотя бы по одному посту СКЗА).

В каталоге городов KAtS для одного города заносятся номер города и номер головной зоны города (ГЗГ), так как для каждого города организована своя ГЗГ, которая, как это видно из табл. 5, характеризует наличие информации за каждый месяц, порядок записи, координаты записанных постов, количество записанных постов. Место записи информации по каждому городу строго фиксируется на МЛар «средние» с помощью каталога городов (KAtS). Для одного города отведено 7 зон вместе с ГЗГ, в одну зону можно записать информацию по 15 постам за два месяца.

Для каждого поста СКЗА записывается служебная информация — постоянные характеристики поста, а также характеристики количественного и качественного содержания информации. После служебной информации размещаются непосредственно результаты статистической обработки за конкретный месяц, упакованные по макетам, показанным на рис. 1.

7. Запись информации для долговременного хранения ее на МЛар выполняется после проведения критического анализа качества первичных данных и качества машинной обработки. МЛарп, имеющая в информационной зоне признаки «разрешение архивации», является исходной информацией для данного этапа — записи информации в архив. Блок-схема программы «архивация» представлена на рис. 2.

Структура ГЗГ на МЛар „средние“

Адрес массива в зоне	Название массива	Номер слова в зоне	Содержание	Пояснение	
1	ГОР [1 : 7]	1	φφλλλλ	Номер города	
		2	hhh	Высота станции	
		3	КЗП	Количество записанных постов	
		4	пзГ	Номер ГЗГ	
		5	пз1.	Номер зоны записи информации за данный месяц	
		6	ММ	Месяц	
		7	—		
21	ЗП* [1 : 15]	21	NN ₁	Номера постов определяют последовательность их записи в зонах, отведенных для записи данного города	
		22	NN ₂		
		:	:		
		20+j	NN _j		
		:	:		
41	мп* [1 : 12]	35	NN ₁₅	мп[i]=i если записывалась информация хотя бы одного поста i-го месяца	
		41	ММ ₁		
		42	ММ ₂		
		:	:		
		40+i	ММ _i		
61	П _j [1 : 15]	52	ММ ₁₂	Номер поста (j=1) координаты поста	
		61	NN ₁		
		62	±xx		
		63	±xx		
		64	ММ ₁		Номера записанных месяцев данного поста
		65	ММ ₂		
		:	:		
60+20	П _(j-1) [1 : 15]	60+i	ММ _i	Количество массивов П [1 : 15] определяется КЗП (ГОР [3])	
		:	:		
		75	ММ ₁₂		

Примечание. Звездочкой отмечены величины, меняющиеся в процессе записи информации.

Перед каждой первой записью информации на данную МЛар программно проверяется соответствие метки обеих лент (МЕТ [1]), что гарантирует запись информации на нужную ленту. Аналогично выполняется анализ на соответствие месяца на МЛар «срочные» и на МЛарп, что обеспечивает запись нужной информации на соответствующую ленту.

ИЗ МЛарп предварительно обрабатывается таким образом, чтобы сформировать каталог городов, информация которых находится на данной МЛарп. Это упрощает и ускоряет запись информации отдельного города на МЛар, особенно удобен полученный каталог городов при записи на МЛар «средние». Затем вся информация от каждого поста СКЗА одного города считывается с МЛарп и упаковывается по несколько слов в ячейку согласно макетам упаковки.

Информация по данным наблюдений на СКЗА сразу же записывается с защитным кодированием на МЛар «срочные». Место на МЛар «срочные», отводимое для записи информации по каждому городу, предварительно не фиксируется. Последовательность записи городов (и постов СКЗА в них) для разных месяцев не совпадает и определяется последовательностью поступления информации на начало этапа обработки.

Информация о результатах статистической обработки накапливается на МБ до тех пор, пока не будет записана на магнитный барабан (МБ) информация от всех постов данного города, имеющаяся на МЛарп. После этого вся информация переписывается с защитным кодированием с МБ на МЛар «средние».

По мере записи на АЦПУ печатается информация: номер города, номер поста СКЗА в этом городе, номер зоны МЛар «срочные», в которую записаны данные наблюдений. Для информации записываемой на МЛар «средние» на печать выводятся сведения о записи всех постов одного города за каждый месяц, с указанием номера зоны, в которую производилась запись. Программой вносятся коррективы в соответствующие каталоги, бинарные таблицы, ГЗФ, а также в ИЗ МЛарп, отражающие процесс записи информации в архив.

8. Таким образом, созданная АСОИЗА помимо функций сбора, первичного контроля и обработки информации о загрязнении атмосферы выполняет также и функции начального режимного обслуживания материалами, выходящими непосредственно после обработки информации. Накопление проконтролированных данных на долговременном техническом носителе (МЛ) с одновременной организацией и комплектацией массивов данных по определенной структуре уже сейчас предоставляет широкие возможности использования информации для режимно-справочных обобщений. Кроме того, получаемые на выходе АСОИЗА архивы удобны для применения их в рамках АИСС гидрометеорологической информации, работы по созданию которых в настоящее время начинают развиваться. Поскольку в качестве технической базы АИСС Госкомгидромета рассматриваются ЭВМ третьего

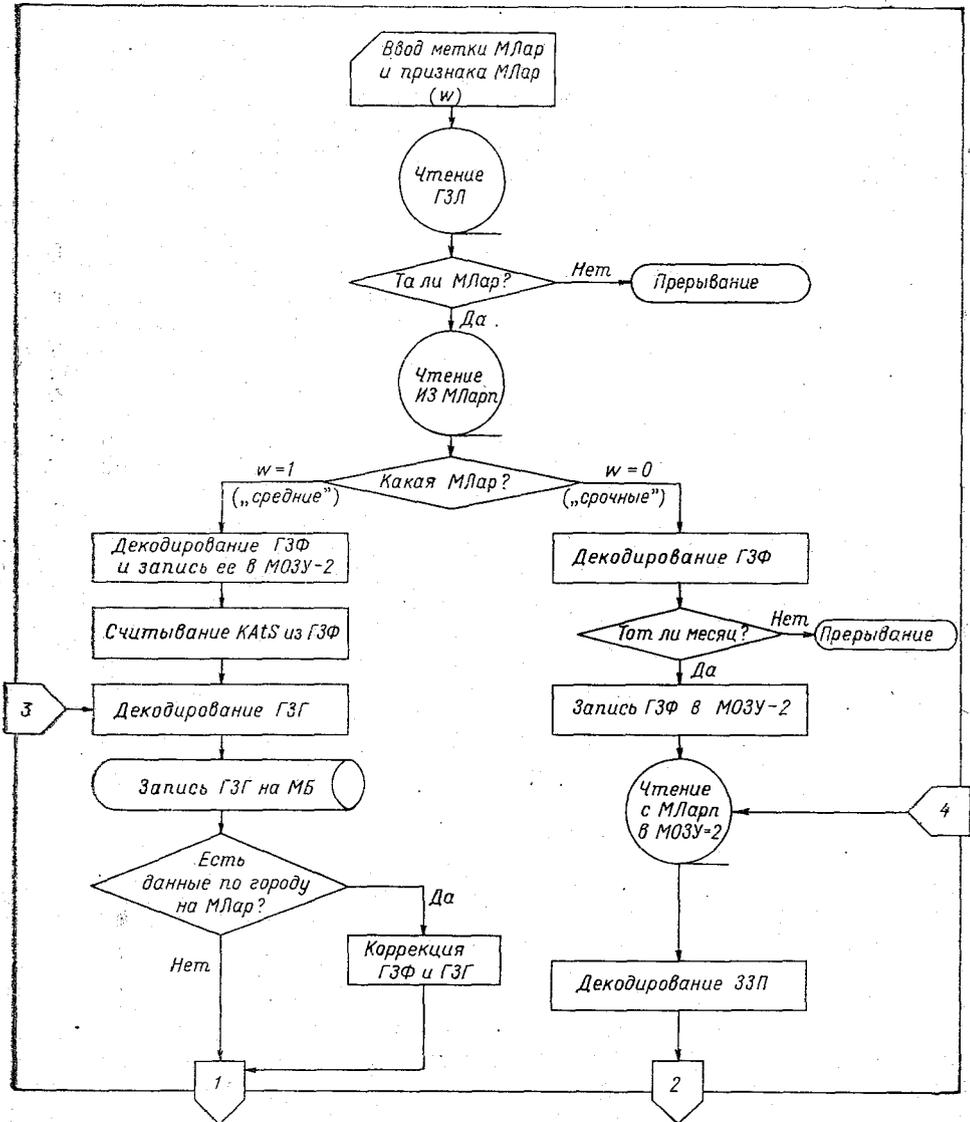
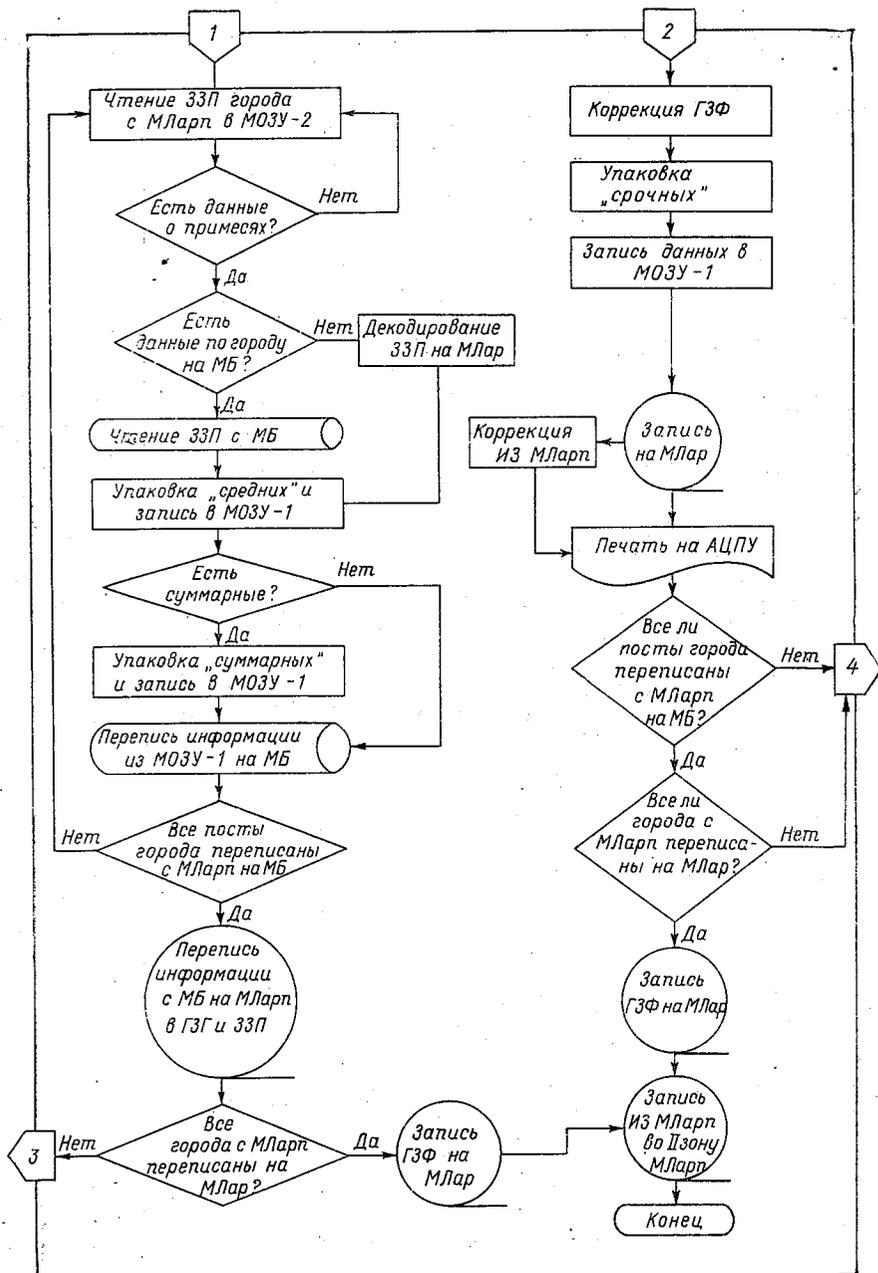


Рис. 2. Блок-схема



программы «архивация».

поколения, очень важно, чтобы информационная база, получаемая от систем обработки, функционирующих на ЭВМ других типов, была удобна для переноса на носители ЭВМ ЕС. Именно такие цели мы имели в виду при разработке структуры архивов на МЛ.

СПИСОК ЛИТЕРАТУРЫ

1. Абашев Г. А. и др. Алгоритмы и программы кодирования и декодирования четырехмерной кодовой матрицы. — Труды ГГО, 1975, вып. 347, с. 68—77.
2. Афиногенов Л. П. Длительное хранение информации при использовании помехоустойчивого кодирования и периодической регенерации. — Труды ГГО, 1975, вып. 347, с. 55—62.
3. Афиногенов Л. П. Особенности хранения информации при помехоустойчивом кодировании и дублировании. — Труды ГГО, 1977, вып. 377, с. 92—104.
4. Веселов В. М. Об организации работы с метеорологическими архивами, хранящимися на магнитных лентах ЭВМ М-222. — Труды ВНИИГМИ — МЦД, 1976, вып. 13, с. 196—210.
5. Винников К. Я., Гриб Н. К., Поляк И. И. Методика расчета корреляционных функций и спектров временных метеорологических рядов. — Труды ГГО, 1973, вып. 308, с. 27—46.
6. Временные методические указания по машинной обработке и контролю данных гидрометеорологических наблюдений. Вып. 10, часть 1, разд. 1. — Обнинск: Ротапринт. ВНИИГМИ — МЦД, 1974. — 24 с.
7. Полищук А. И., Толкачева Н. Н., Юринова Т. Н. Автоматизированная система сбора, обработки и обобщения режимной информации о загрязнении воздуха в городах. — Труды ГГО, 1979, вып. 436, с. 124—131.
8. Поляк И. И. Алгоритмы и программы статистического анализа метеорологических наблюдений. — Труды ГГО, 1975, вып. 364, с. 40—44.

Л. П. Афиногенов, Г. А. Абашев

КОЛИЧЕСТВО ИНФОРМАЦИИ, ПЕРЕДАВАЕМОЙ В СИСТЕМЕ ДЛИТЕЛЬНОГО ХРАНЕНИЯ, ИСПОЛЬЗУЮЩЕЙ АЛФАВИТ С ФИКСИРОВАННЫМИ ОГРАНИЧЕНИЯМИ

Систему длительного хранения информации можно рассматривать как канал связи и применять к ней методы теории информации. В ряде случаев множество (алфавит) символов, используемых при записи (на входе канала) и при чтении (на выходе канала), превышает множество символов, поступающих от источника, но зато подчинено системе фиксированных ограничений, обусловленных физическими свойствами носителя и устройств записи-чтения.

Простейшим и широко распространенным примером такого случая является канал дискретной двухуровневой записи/считывания магнитной ленты (МЛ). Здесь используются два состояния предельного намагничивания магнитного слоя ($+B_r$ и $-B_r$) и чередующиеся переходы из одного состояния в другое. При методе записи NRZI [8] двоичному «0» соответствует неизменение состояния магнитного слоя, а двоичной «1» — изменение состояния (любой переход). Таким образом, на входе рассматриваемого канала используются четыре символа [1]: $F_1 \sim +B_r$; $F_0 \sim -B_r$; $\Delta_0 \sim -B_r \rightarrow +B_r$; $\Delta_1 \sim +B_r \rightarrow -B_r$, а ограничения состоят в том, что после символов F_1 и Δ_0 могут следовать только символы F_1 и Δ_1 , после F_0 и Δ_1 — F_0 и Δ_0 . Алфавит на выходе канала, т. е. алфавит сигналов головки считывания, состоит из трех символов: E_0 — отсутствие напряжения, возникает при чтении с МЛ символов F_0 и F_1 ; E_+ — положительный импульс, соответствующий символу Δ_0 ; E_- — отрицательный импульс, возникающий при чтении символа Δ_1 . Здесь также имеются ограничения, связанные с последовательностью символов на МЛ и состоящие в том, что символы E_+ и E_- чередуются, причем они могут быть разделены произвольным числом символов 0. В процессе считывания МЛ возможны искажения сигналов, которые соответствуют помехам в обычном канале связи.

Систему такого типа можно представить в виде преобразователя, имеющего k внутренних состояний $C_1 \dots C_k$, и обычного канала с помехами (рис. 1). На вход преобразователя от источника информации поступают символы x_i ($i=1 \dots n$), выбираемые независимо с вероятностями $p_1 \dots p_n$. В ответ на каждый входной символ преобразователь генерирует символ z_g ($g=1 \dots m$), который зависит от предыдущего состояния преобразователя C_j и входного символа x_i . При этом преобразователь переходит в новое состояние $C_{j'}$, которое также зависит от предыдущего состояния и входного символа. Далее символы z_g поступают в канал, с выхода которого снимаются символы y_t ($t=1 \dots l$). Канал в этой модели можно рассматривать как преобразователь, осуществляющий однозначное преобразование алфавита $z_1 \dots z_m$ в алфавит $y_1 \dots y_l$ при отсутствии искажений и вероятностное преобразование [2] при наличии искажений.

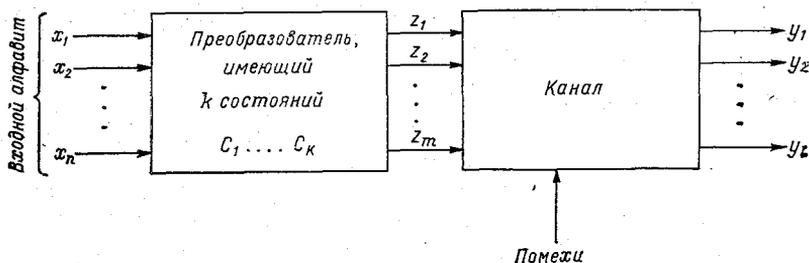


Рис. 1. Представление канала записи/считывания в виде преобразователя и канала с помехами.

Преобразователь, имеющий k состояний и работающий от источника с независимым выбором символов, можно представить как источник информации марковского типа. Такой источник, а также канал с помехами и без них рассматривались в классической работе Шеннона [11] и затем другими авторами [6, 9, 10 и др.]. В этих работах предполагается, что между источниками и каналом имеется кодирующее устройство, которое обеспечивает их согласование так, что символы на входе канала могут выбираться независимо. Методы расчета информационных характеристик дискретного канала с шумами также основаны на предположении о возможности независимого выбора входных символов. Однако основной особенностью рассматриваемой системы является то, что символы на входе канала (после преобразователя) не могут выбираться произвольно. В данной работе рассматривается метод расчета количества информации, передаваемой системой рис. 1 с ограничениями на входе канала.

Основные информационные характеристики (энтропия, количество информации, пропускная способность) будут относиться к одному входному символу (а не к единице времени), поэтому

длительность символов может не учитываться. Это отражает ту реальную особенность рассматриваемой системы, что каждому входному символу $x_1 \dots x_n$ соответствует один символ на входе и один на выходе канала.

Сам дискретный канал может быть описан обычным образом с помощью стохастической матрицы канала [2]

$$M_k = \|\delta_{gt}\|; \quad g = 1 \dots m; \quad t = 1 \dots l, \quad (1)$$

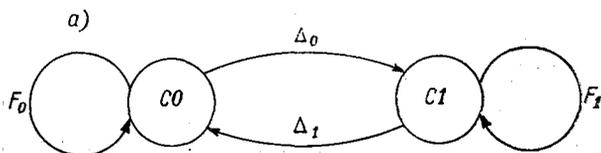
содержащей условные вероятности δ_{gt} появления на выходе символа y_t , если на вход поступил символ z_g .

Например, для МЛ это будет матрица размером 4×3 (4 входных и 3 выходных символа) вида

$$M_{к_{МЛ}} = \begin{vmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \\ \delta_{41} & \delta_{42} & \delta_{43} \end{vmatrix}. \quad (2)$$

Здесь первая цифра индекса (1, 2, 3, 4) соответствует входным символам $F_0, F_1, \Delta_0, \Delta_1$, а вторая (1, 2, 3) — выходным символам E_0, E_+, E_- .

Для описания преобразователя необходимо задать закон, определяющий выходной символ z_g и новое состояние C_{j_1} в зависимости от входного символа x_i и предыдущего состояния C_{j_0} . Это можно сделать с помощью графов и таблиц переходов [11]



б)

Входной символ	Состояние	
	C0	C1
0	$\frac{F_0}{C0}$	$\frac{F_1}{C1}$
1	$\frac{\Delta_0}{C1}$	$\frac{\Delta_1}{C0}$

Рис. 2. Описание преобразователя для МЛ.

а — граф состояний, б — таблица переходов.

или двумерных матриц. На рис. 2 в качестве примера показаны граф и таблица переходов преобразователя для МЛ. Здесь в каждой клетке таблицы сверху указан выдаваемый символ, а снизу — новое состояние в зависимости от входного символа (0 или 1) и состояния преобразователя (C0 или C1).

Для дальнейшего анализа воспользуемся описанием с помощью двумерных матриц. Каждому состоянию преобразователя C_j соответствует матрица преобразования $M_{пj}$ размером $n \times m$ и матрица состояний M_{cj} размером $n \times k$. Первая из них определяет выходные символы, а вторая — новые состояния. Число строк в каждой матрице равно числу выходных символов (n), число столбцов в $M_{пj}$ равно числу входных символов (m), а в M_{cj} — числу состояний (k). В каждой (i -й) строке матриц содержится информация о выходных символах z_g (для $M_{пj}$) и новых состояниях C_{j_1} (для M_{cj}), возникающих, когда при данном состоянии преобразователя C_{j_0} на вход поступает символ x_i . В случае однозначного преобразователя каждая строка содержит одну 1 (в колонке, соответствующей генерируемому выходному символу для $M_{пj}$ или новому состоянию для M_{cj}) и нули в остальных позициях. Для иллюстрации приведем матрицы преобразователя при считывании МЛ:

$$M_{п0} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{vmatrix}, \quad M_{c0} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix},$$

$$M_{п1} = \begin{vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}, \quad M_{c1} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}.$$

Здесь строки всех матриц соответствуют входным символам $x_1=0$ (верхняя строка) и $x_2=1$ (нижняя строка); четыре колонки матриц $M_{п0}$ и $M_{п1}$ соответствуют четырем выходным символам $F_0, F_1, \Delta_0, \Delta_1$; две колонки в матрицах M_{c0} и M_{c1} соответствуют новым состояниям C0 (левая колонка) и C1 (правая колонка); матрицы $M_{п0}$ и M_{c0} соответствуют предыдущему состоянию C0, матрицы $M_{п1}$ и M_{c1} — предыдущему состоянию C1.

Матричный метод описания позволяет также рассматривать более общий случай, когда возможны ошибки преобразователя. При этом в строках матриц $M_{пj}$ и M_{cj} указываются условные вероятности выдачи выходных символов и переходов в новое состояние при данном входном символе x_i и данном состоянии C_{j_0} . В дальнейшем мы будем использовать именно это более общее представление для матриц $M_{пj}$, M_{cj} , поскольку оно не усложняет анализа.

Обозначим $\alpha_{gij} = \alpha(z_g/x_i, C_j)$ — условные вероятности выдачи преобразователем символа z_g ($g=1\dots m$) при поступлении на вход символа x_i ($i=1\dots n$), если преобразователь находится в состоянии C_j ($j=1\dots k$); $\beta_{j_1j_0} = \beta(C_{j_1}/x_i, C_{j_0})$ — условные вероятности перехода преобразователя в состояние C_{j_1} из состояния C_{j_0} ($j_1,$

$j_0=1...k$), если на вход поступило x_i . Эти вероятности удовлетворяют условиям

$$0 \leq \alpha_{gij} \leq 1, \quad 0 \leq \beta_{j_i i j_0} \leq 1, \quad (3)$$

$$\sum_g \alpha_{gij} = \sum_{j_i} \beta_{j_i i j_0} = 1. \quad (4)$$

Таким образом, матрицы $\mathbf{M}_{\Pi j} = \|\alpha_{gij}\|$ и $\mathbf{M}_{C_j} = \|\beta_{j_i i j_0}\|$ являются стохастическими.

Для того чтобы система (см. рис. 1) была полностью описана, необходимо задать распределение вероятностей входных символов p_i ($i=1...n$); мы будем считать его известным.

Как и при любом преобразовании, количество информации, передаваемое с входа на выход системы, определяется формулой

$$I(Y, X) = H(Y) - H(Y/X), \quad (5)$$

где $H(Y)$ — полная энтропия выхода; $H(Y/X)$ — условная энтропия выхода при известном входном символе, осредненная по всем символам с учетом их вероятностей. При определении $H(Y)$ и $H(Y/X)$ для рассматриваемой системы должны учитываться ее особенности. Это означает, что должны быть определены частные характеристики $H(Y/C_j)$, $H(Y/X, C_j)$, $I(Y/X, C_j)$ отдельно для каждого состояния C_j преобразователя, а затем могут быть найдены значения $H(Y)$, $H(Y/X)$, $I(Y, X)$ путем осреднения частных характеристик по всем состояниям C_j с учетом их вероятностей. Таким образом, методика вычисления $I(Y, X)$ состоит в следующем.

1. Для каждого состояния преобразователя C_j составляется общая матрица преобразователя и канала $\mathbf{M}_{\Pi K j}$, содержащая основные вероятности появления символов y_t ($t=1...l$) на выходе канала при данных символах на входе преобразователя x_i и данном состоянии канала. Эта матрица равна произведению

$$\mathbf{M}_{\Pi K j} = \mathbf{M}_{\Pi j} \mathbf{M}_K. \quad (6)$$

2. По матрице $\mathbf{M}_{\Pi K j}$ и входному распределению p_i находятся $H(Y/C_j)$, $H(Y/X, C_j)$, $I(Y/C_j)$ для системы «преобразователь в состоянии C_j — канал». Приведем формулы для расчета этих величин. При известном входном распределении p_i ($i=1...n$) и матрице системы $\mathbf{M}_{\Pi K j} = \|p_{it}\|$ ($t=1...l$; p_{it} — элементы матрицы $\mathbf{M}_{\Pi K j}$):

$$q_t = \sum_{i=1}^n p_i p_{it}, \quad (7)$$

где q_t — вероятности символов y_t на выходе;

$$H(Y) = - \sum_{t=1}^l q_t \lg q_t, \quad (8)$$

$$H(Y/X) = - \sum_{i=1}^n p_i \sum_{t=1}^l p_{it} \lg p_{it}, \quad (9)$$

$$\begin{aligned}
 I(Y, X) &= H(Y) - H(Y|X) = \\
 &= \sum_{i=1}^n p_i \sum_{t=1}^l p_{it} \lg p_{it} - \sum_{t=1}^l q_t \lg q_t
 \end{aligned} \tag{10}$$

3. Определяются вероятности γ_j каждого из состояний преобразователя C_j . Для этого необходимо предварительно найти квадратную ($k \times k$) матрицу условных вероятностей $\eta_{j_i j_0}$ перехода в состояние C_{j_i} , если преобразователь находится в состоянии C_{j_0} : $M_{\text{пер}} = \|\eta_{j_i j_0}\|$ ($j_i, j_0 = 1 \dots k$). Каждый элемент этой матрицы, т. е. вероятность перехода из состояния C_{j_0} в состояние C_{j_i} , равен сумме элементов соответствующего состоянию C_{j_i} столбца матрицы состояний $M_{C_{j_0}}$, взятых с весами, равными вероятностям входных символов x_i . Отсюда каждая строка $M_{\text{пер}}$ может быть найдена как произведение

$$\|\eta_{j_i j_0}\| = \|p_i\| \cdot \|\beta_{j_i j_0}\|, \tag{11}$$

где $\|\eta_{j_i j_0}\|$ — матрица-строка, являющаяся j_0 -й строкой $M_{\text{пер}}$ и представляющая набор вероятностей перехода из состояния C_{j_0} в состояния $C_1 \dots C_k$; $\|p_i\|$ — матрица-строка вероятностей входных символов; $\|\beta_{j_i j_0}\|$ — матрица состояний $M_{C_{j_0}}$ для данного состояния преобразователя C_{j_0} . Определив все k строк $\|\eta_{j_i j_0}\|$ по формуле (11) и записав их в виде матрицы, получим квадратную матрицу $M_{\text{пер}}$. Для получения вероятностей γ_j состояний C_j необходимо найти предельную матрицу

$$M_{\text{пред}} = \lim_{a \rightarrow \infty} M_{\text{пер}}^a. \tag{12}$$

В большинстве случаев, когда выполняются условия связности марковской цепи [7], каковой по существу является преобразователь системы рис. 1, этот предел существует и представляет собой стохастическую матрицу ранга 1 (т. е. матрицу с одинаковыми строками) [4], причем строка этой матрицы состоит из искомым вероятностей γ_j . Для ее определения составим матричное уравнение

$$\|\gamma_j\| \cdot M_{\text{пред}} = \|\gamma_j\| \cdot \|\eta_{j_i j_0}\| = \|\gamma_j\|. \tag{13}$$

Это уравнение отражает тот факт, что при умножении предельного значения $\|\gamma_j\|$ на $M_{\text{пред}}$ должно получаться то же предельное значение. Уравнение (13) сводится к системе из k однородных линейных уравнений относительно k неизвестных с определителем, равным 0, и, следовательно, имеет множество решений. Присоединив к (13) естественное условие

$$\sum_j \gamma_j = 1 \tag{14}$$

(вытекающее из того, что k -состояний преобразователя составляют полную группу событий), получим неоднородную систему

($k+1$ уравнение, k неизвестных), которая в данном случае, как можно показать [5], имеет единственное решение.

4. После определения вероятностей состояний можно найти значения

$$H(Y) = \sum_{j=1}^k \gamma_j H(Y/C_j), \quad (15)$$

$$H(Y/X) = \sum_{j=1}^k \gamma_j H(Y/X, C_j), \quad (16)$$

$$\begin{aligned} I(Y, X) &= H(Y) - H(Y/X) = \sum_{j=1}^k \gamma_j I(Y/X, C_j) = \\ &= \sum_{j=1}^k \gamma_j H(Y/C_j) - \sum_{j=1}^k \gamma_j H(Y/X, C_j). \end{aligned} \quad (17)$$

Полученное таким образом количество передаваемой информации $I(Y/X)$ является функцией входных вероятностей $p_1 \dots p_n$. Для определения пропускной способности системы $C = \sup I(Y/X)$ можно воспользоваться известными аналитическими приемами вычисления экстремума функции нескольких переменных. Методы расчета пропускной способности обычных дискретных каналов с шумами рассмотрены в работах [3, 12].

Для иллюстрации описанной методики приведем пример расчета $I(Y, X)$ для канала записи/считывания МЛ. Пусть p_0 и p_1 — вероятности записываемых двоичных символов 0 и 1; C_0 и C_1 — состояния преобразователя (т. е. состояния сигналограммы на ленте);

$$\begin{aligned} M_{п0} &= \begin{vmatrix} 0,99 & 0 & 0,01 & 0 \\ 0,02 & 0 & 0,98 & 0 \end{vmatrix}, & M_{п1} &= \begin{vmatrix} 0 & 0,99 & 0 & 0,01 \\ 0 & 0,01 & 0 & 0,99 \end{vmatrix}, \\ M_{с0} &= \begin{vmatrix} 0,99 & 0,01 \\ 0,02 & 0,98 \end{vmatrix}, & M_{с1} &= \begin{vmatrix} 0,01 & 0,99 \\ 0,99 & 0,01 \end{vmatrix} \end{aligned}$$

— матрицы преобразований и состояний для C_0 и C_1 . Здесь рассматривается общий случай и допускается ошибочная работа преобразователя;

$$M_k = \begin{vmatrix} 0,96 & 0,01 & 0,03 \\ 0,96 & 0,03 & 0,01 \\ 0,02 & 0,97 & 0,01 \\ 0,01 & 0,01 & 0,98 \end{vmatrix}$$

— матрица канала.

1. Находим общие матрицы преобразователя и канала для обоих состояний МЛ по формуле (6):

$$M_{пк0} = M_{п0} M_k = \begin{vmatrix} 0,9506 & 0,0196 & 0,0298 \\ 0,0388 & 0,9508 & 0,0104 \end{vmatrix},$$

$$M_{\text{пкГ}} = M_{\text{пГ}} M_{\text{к}} = \begin{vmatrix} 0,9505 & 0,0298 & 0,0197 \\ 0,0195 & 0,0102 & 0,9703 \end{vmatrix}.$$

2. Для входных вероятностей p_0 и p_1 (числовые значения которых введем позже) по формулам (7)–(10) находим: для состояния C0:

$$q_{E_0} = 0,9506p_0 + 0,0388p_1,$$

$$q_{E_+} = 0,0196p_0 + 0,9508p_1,$$

$$q_{E_-} = 0,0298p_0 + 0,0104p_1,$$

$$H(Y/C0) = -q_{E_0} \lg q_{E_0} - q_{E_+} \lg q_{E_+} - q_{E_-} \lg q_{E_-},$$

$$H(Y/X, C0) = -p_0(0,9506 \lg 0,9506 + 0,0196 \lg 0,0196 + 0,0298 \lg 0,0298) - p_1(0,0388 \lg 0,0388 + 0,9508 \lg 0,9508 + 0,104 \lg 0,0104) = 0,331\ 712\ 508p_0 + 0,319\ 599\ 368p_1,$$

$$I(Y/X, C0) = H(Y/C0) - H(Y/X, C0);$$

для состояния C1:

$$q_{E_0} = 0,9505p_0 + 0,0195p_1,$$

$$q_{E_+} = 0,0298p_0 + 0,0102p_1,$$

$$q_{E_-} = 0,0197p_0 + 0,0703p_1,$$

$$H(Y/C1) = q_{E_0} \lg q_{E_0} - q_{E_+} \lg q_{E_+} - q_{E_-} \lg q_{E_-},$$

$$H(Y/X, C1) = -p_0(0,9505 \lg 0,9505 + 0,0298 \lg 0,0298 + 0,0197 \lg 0,0197) - p_1(0,0195 \lg 0,0195 + 0,0102 \lg 0,0102 + 0,9703 \lg 0,9703) = 0,332\ 272\ 129p_0 + 0,220\ 448\ 73p_1,$$

$$I(Y/X, C1) = H(Y/C1) - H(Y/X, C1).$$

3. С помощью (11) по матрицам M_{c0} и M_{c1} находим матрицу $M_{\text{пер}}$:

$$M_{\text{пер}} = \begin{vmatrix} \eta_{c0c0} & \eta_{c0c1} \\ \eta_{c1c0} & \eta_{c1c1} \end{vmatrix},$$

где

$$\eta_{c0c0} = \|p_0 p_1\| \cdot \begin{vmatrix} 0,99 \\ 0,02 \end{vmatrix} = 0,99p_0 + 0,02p_1,$$

$$\eta_{c0c1} = \|p_0 p_1\| \cdot \begin{vmatrix} 0,01 \\ 0,98 \end{vmatrix} = 0,01p_0 + 0,98p_1,$$

$$\gamma_{c_1 c_0} = \|p_0 p_1\| \cdot \left\| \begin{array}{c} 0,01 \\ 0,99 \end{array} \right\| = 0,01 p_0 + 0,99 p_1,$$

$$\gamma_{c_1 c_1} = \|p_0 p_1\| \cdot \left\| \begin{array}{c} 0,99 \\ 0,01 \end{array} \right\| = 0,99 p_0 + 0,01 p_1.$$

4. Составляем матричное уравнение (13)

$$\|\gamma_{c_0} \gamma_{c_1}\| \cdot M_{\text{пер}} = \|\gamma_{c_0} \gamma_{c_1}\|.$$

Это уравнение и равенство $\gamma_{c_0} + \gamma_{c_1} = 1$ приводят к системе уравнений:

$$\left. \begin{array}{l} \gamma_{c_0}(0,99 p_0 + 0,02 p_1 - 1) + \gamma_{c_1}(0,01 p_0 + 0,99 p_1) = 0, \\ \gamma_{c_0}(0,01 p_0 + 0,98 p_1) + \gamma_{c_1}(0,99 p_0 + 0,01 p_1(-1)) = 0, \\ \gamma_{c_0} + \gamma_{c_1} = 1. \end{array} \right\}$$

Решение системы дает вероятности состояний:

$$\gamma_{c_0} = \frac{0,01 p_0 + 0,99 p_1}{1 - 0,98 p_0 + 0,97 p_1}, \quad \gamma_{c_1} = \frac{1 - 0,99 p_0 - 0,02 p_1}{1 - 0,98 p_0 + 0,97 p_1}.$$

5. С учетом найденных значений γ_{c_0} , γ_{c_1} , $H(Y/C_0)$, $H(Y/C_1)$, $H(Y/X, C_0)$, $H(Y/X, C_1)$ по формуле (17) находим количество информации в зависимости от p_0 и p_1 . В частности, при $p_0 = p_1 = 0,5$, как это обычно бывает при записи на МЛ неизвестной заранее двоичной информации, $I(Y/X) = 0,902 292 8$.

Полученное для $I(Y/X)$ выражение содержит вероятности входных символов. Подставив в него $p_1 = 1 - p_0$ и дифференцируя по p_0 , можно найти экстремум функции $I(p_0)$, который будет представлять собой пропускную способность данного канала.

СПИСОК ЛИТЕРАТУРЫ

1. Абашев Г. А. Анализ методов дискретной записи на магнитный носитель. — Труды ВНИИГМИ—МЦД, 1978, вып. 31, с. 54—63.
2. Афиногенов Л. П. Информационные свойства вероятностных преобразований. — Труды ГГО, 1975, вып. 347, с. 5—12.
3. Афиногенов Л. П. Расчет информационной пропускной способности стохастических матриц. — Труды ГГО, 1975, вып. 364, с. 77—84.
4. Афиногенов Л. П. Информационные свойства цепей вероятностных преобразований. — Труды ГГО, 1975, вып. 347, с. 13—29.
5. Афиногенов Л. П. Условия разрешимости матричных уравнений. — Труды ВНИИГМИ—МЦД, 1975, вып. 12, с. 98—114.
6. Голдман С. Теория информации. — М.: Изд-во иностр. лит., 1957. — 446 с.
7. Колмогоров А. Н. Цепи Маркова со счетным множеством возможных состояний. — Бюлл. МГУ, 1937, т. 1, вып. 3.
8. Накопители на магнитной ленте для электронных вычислительных машин. Типы. Технические требования. ГОСТ 14127—69. — М.: Изд-во стандартов, 1969.

9. Фано Р. Передача информации. Статистическая теория связи.—М.: Мир, 1965.—438 с.
10. Финк Л. М. Теория передачи дискретных сообщений.—М.: Сов. радио, 1970.—728 с.
11. Шеннон К. Математическая теория связи.—В кн.: К. Шеннон. Работы по теории информации и кибернетике.—М.: Изд-во Иностран. лит., 1963, с. 243—332.
12. Mi g o g a S. On the capacity of a discrete channel I.—J. Phys. Soc. Japan, 1953, vol. 8, N 4, p. 484—494.

ОРГАНИЗАЦИЯ ДИАЛОГА ПРИ РЕШЕНИИ ЗАДАЧ ОКЕАНОЛОГИИ

Введение

В настоящее время использование ЭВМ в исследованиях Мирового океана стало обычным явлением. Более того, появились средства диалога человека с машиной [3]. Общение исследователя с ЭВМ в различных океанологических экспериментах имеет в большинстве случаев общие черты. Это позволяет алгоритмизировать процесс диалога.

При проведении исследования океанолог должен, например, быстро проверить построение ряда возможных вариантов моделей поля, отфильтровать исходные данные, оценить данные по структурным функциям или промежуточным результатам, внести коррекцию в исходные данные, окончательные результаты. С современных позиций такой процесс реализуется в рамках интерактивной (диалоговой) графической системы (ГС) на базе мини-ЭВМ с графическим дисплеем, снабженным световым пером, соответствующим программным обеспечением, и при активном участии человека.

Процесс диалога

Работу человека в диалоговой системе можно представить в виде итерационных процедур ввода, решения, вывода.

Исходные данные в процессе экспериментов или расчетов могут быть записаны на перфоленту, перфокарты, магнитную ленту, магнитный диск или вводятся в графическом виде через графический дисплей посредством светового пера либо в цифровом виде — посредством клавиатуры.

Особенности каждого эксперимента учитываются в тематических программах расчета, выполняющих обработку исходных данных. Основными тематическими программами, например, для статистической обработки пространственно-временных рядов являются [1]:

— определение среднего арифметического, среднего квадратического, дисперсии, асимметрии, эксцесса;

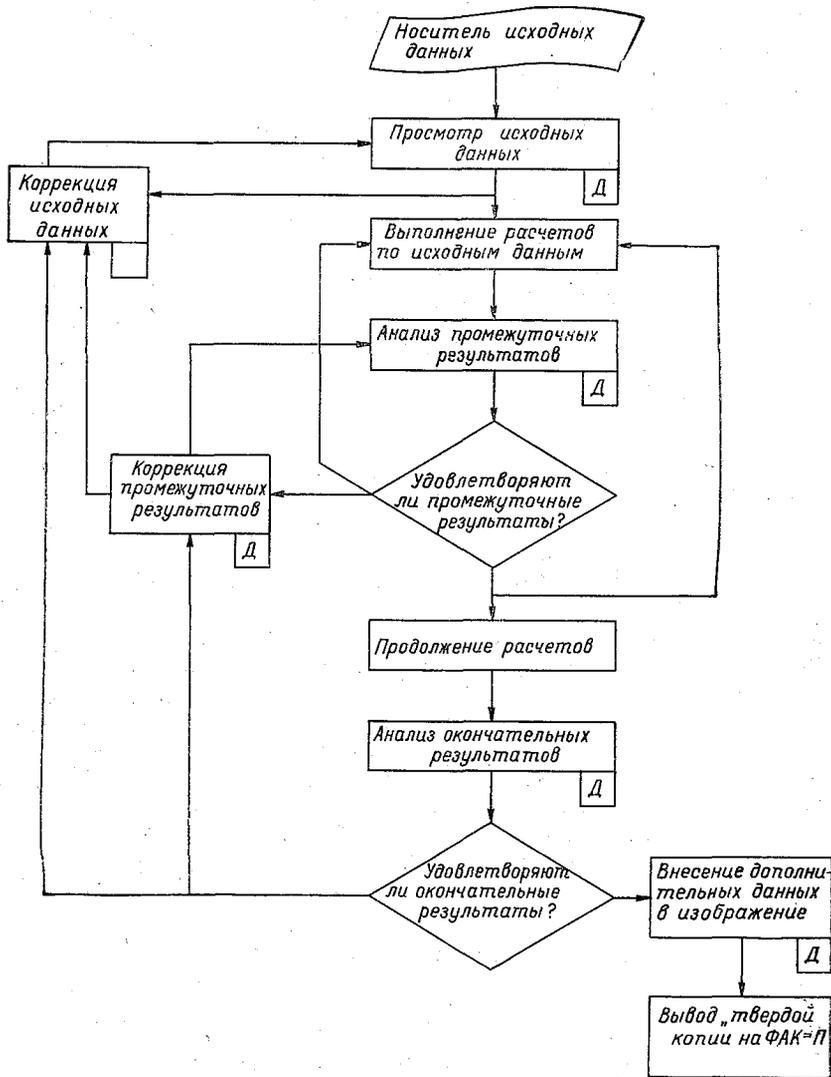


Рис. 1. Информационная схема процесса обработки данных.

- структурных функций;
- корреляционных функций;
- спектральных функций;
- фильтрация рядов.

Эти программы, программы формирования массивов вывода на экран, редактирующие и документирующие программы, программы управления составляют полный перечень операций в си-

стеме. Для нормального функционирования выполнение каждого действия подвергается семантическому и синтаксическому контролю.

Процесс исследования выполняется следующим образом. Вводя исходные данные, океанолог просматривает их на экране, корректируя вручную, затем при необходимости фильтрует ряд (ряды). После этого проводит расчеты по тематическим программам с выводом результатов на экран. При удовлетворении результатом его графическое изображение выводится в виде «твердой» копии на ФАК-П «Ладога», графопостроитель или АЦПУ.

Особенности интерактивной океанологической системы

Обработка длинных рядов океанологических данных представляет довольно трудоемкий процесс. Этот процесс включает целый ряд операций, которые в силу своей специфики должны выполняться специалистом океанологом. Задача обработки заключается в вычислении определенных характеристик и графическом представлении их на картах, схемах, диаграммах, графиках.

Информационная схема процесса обработки данных океанологического эксперимента, представленная на рис. 1, может выполняться вручную или на ЭВМ. Процесс обработки является циклическим, и чем больше циклов можно выполнить в определенный отрезок времени с последующим визуальным сравнением полученных результатов с предполагаемыми из опыта, знаний и интуиции экспериментатора, тем выше качество получаемых материалов.

Основным назначением диалоговой части океанологической графической системы является:

- организация ввода и вывода данных в графической и цифровой форме;
- обеспечение диалога океанолога с ЭВМ.

Гибкость и удобство связи океанолога с ЭВМ определяется возможностями и параметрами используемых технических средств и программным обеспечением (рис. 2). Программное обеспечение ГС на ЭВМ III поколения формируется на базе пакетов прикладных программ, которые состоят из обособленных элементов, образующих комплекс.

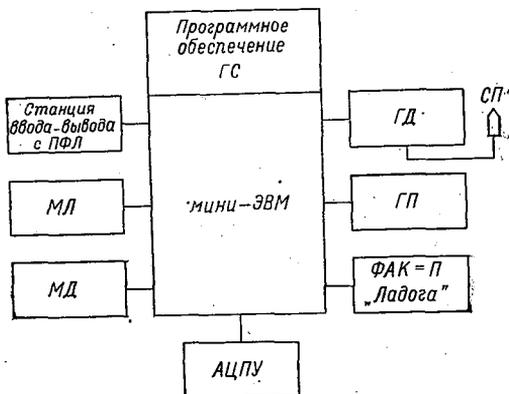


Рис. 2. Структурная схема графической системы.

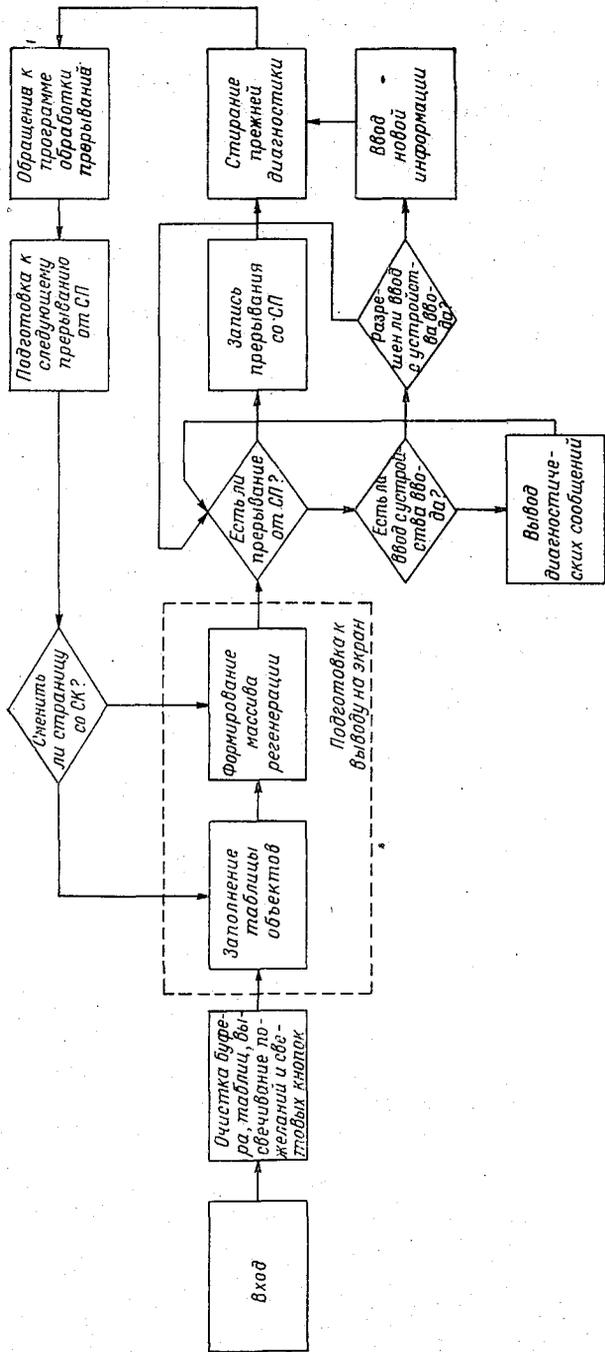


Рис. 3. Структурная схема диалоговой управляющей программы.

Графическая система представляет собой конечный автомат Мили [2]. В произвольный момент времени она находится в одном из возможных для нее состояний, которое определяется входным воздействием и текущим состоянием. Реакция системы на действие пользователя будет выражаться в изменении изображенных на экране диагностических сообщений либо значений рассчитанных характеристик, графического представления данных или результата, либо в смене страниц световых кнопок [3].

Структурная схема диалоговой управляющей программы представлена на рис. 3. Диалог по этой схеме ведется на языке, привычном для человека: языке изображений, схем, графиков и т. п. В разработке языка ГС принимают участие специалисты океанологи. При этом используется метод представления языка систе-



Рис. 4. Цепной однонаправленный список.

мы с помощью диаграмм состояний, которые являются неформальным его описанием. Это требует специальной организации представления информации. Элементы данных размещаются в памяти определенным образом и составляют некоторую структуру данных. Последняя строится из:

- возможности быстрой идентификации точек, линий;
- возможности включения и исключения элементов данных произвольным образом с незначительными затратам машинного времени;
- обеспечения эффективного поиска необходимых данных.

В современных графических системах используется списочная структура данных, которая эффективна при частых изменениях изображений. Учитывая специфику океанологических задач, имеет смысл применить в ГС цепной однонаправленный список (рис. 4). На рис. 4 A_1 — адрес последующего элемента списка, \mathcal{E}_1 — описание элемента изображения. Признаком окончания списка является нулевой адрес элемента. Использование относительной индексной адресации адреса связи позволит обойтись, в частности для мини-ЭВМ ЕС-1010, для A_1 и \mathcal{E}_1 одним словом (16 разрядов). В этом случае для получения действительных адресов необходимо относительные адреса связи суммировать с постоянной для всего списка величиной, представляющей собой адрес начала списка. Кроме того, станет возможной обработка списка с помощью языка Фортран путем соответствующего индексирования.

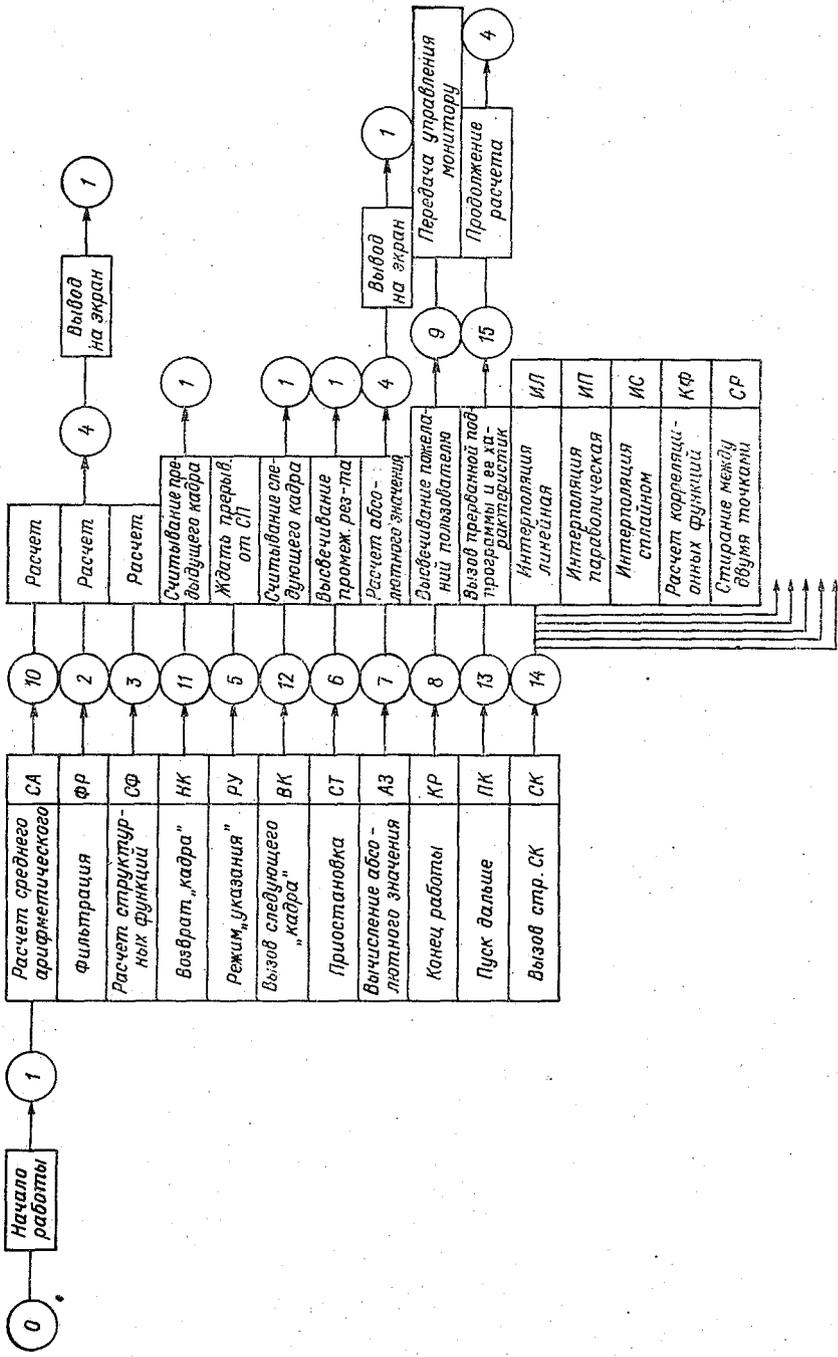


Рис. 5. Диаграмма состояний (фрагмент).

Управление структурой осуществляет программа, работающая согласно диаграмме состояний. Диаграмма состояний, фрагмент из которой представлен на рис. 5, описывает логику работы системы.

Заключение

Графическая система для статистической обработки рядов океанологических данных обладает целым рядом достоинств:

— система организована в виде блочной структуры, что позволяет наращивать и модернизировать ее (посредством изменения тематических программ без изменения состава программного обеспечения системы);

— в описании языка взаимодействия принимают участие будущие пользователи, что позволяет максимально учесть специфику конкретного исследования;

— предусмотрена возможность адаптации системы к уровню подготовленности пользователя;

— система позволяет осуществлять пошаговый контроль за действиями пользователя (вывод обширной диагностики на экран дисплея);

— система позволяет редактировать графическую информацию в процессе ввода исходных данных и анализа результатов;

— система выдает результаты в форме, которая пригодна для использования в отчетах, статьях;

— есть возможности создания многопультных систем.

Учитывая достоинства интерактивных графических систем, их применение при решении задач океанологии целесообразно. Правильный подход к реализации диалога океанолога с ЭВМ намного повысит эффективность ГС и оперативность обработки информации, расширит круг пользователей ЭВМ.

СПИСОК ЛИТЕРАТУРЫ

1. Бышев В. И. Программа статистического анализа наблюдений, выполненных на Атлантическом гидрофизическом полигоне. — В кн.: Атлантический гидрофизический полигон-70. — М.: Наука, 1974.
2. Глушков В. М. Синтез цифровых автоматов. — М.: Физматгиз, 1962.
3. Печискер Л. Г. Диалоговое взаимодействие в задачах океанологии. — См. наст. сб.

Л. Г. Печискер

ДИАЛОГОВОЕ ВЗАИМОДЕЙСТВИЕ В ЗАДАЧАХ ОКЕАНОЛОГИИ

Введение

Судовые системы автоматизации научных исследований (ССАНИ) в последние годы стали одним из мощных средств изучения Мирового океана. Существующие в нашей стране системы базируются на ЭВМ II поколения («Минск-2», «Минск-32»), предназначенных в основном для проведения научных расчетов и не имеющих широкого набора устройств ввода-вывода данных.

Современный этап развития вычислительной техники связан со становлением ЭВМ III поколения. Они характеризуются повышенным быстродействием и надежностью, увеличенным объемом запоминающих устройств, а также улучшенным взаимодействием между ЭВМ и человеком. Последнее выразилось, в частности, в появлении экранных пультов (дисплеев) — принципиально нового средства общения человека с ЭВМ, позволяющего оперативно выводить на экран текст, плоские и объемные изображения. Дисплеи снабжены функциональной и буквенно-цифровой клавиатурой, световым пером, что позволяет человеку вводить в ЭВМ информацию в алфавитно-цифровой и графической формах.

Дисплеи в зависимости от вида представимой информации делятся на алфавитно-цифровые и графические. Алфавитно-цифровые дисплеи предназначены для работы с текстовой информацией и основное применение находят в информационных системах. Графические дисплеи позволяют выводить на экран графики, диаграммы, пространственные фигуры. Такие дисплеи более сложны по структуре, чем алфавитно-цифровые. Наиболее широкое применение получили графические дисплеи в автоматизированных системах проектирования, а также в системах автоматизации научных исследований (САНИ) [2, 10, 14].

Результаты обработки океанологических данных в большинстве случаев представляются в наглядной форме, т. е. в виде

графиков, карт, разрезов. Получение графических изображений на алфавитно-цифровых печатающих устройствах (АЦПУ) не приемлемо из-за ограниченной точности и низкой скорости построения. Известны успешные попытки вывода результатов обработки на графопостроитель ДГУ-2 [1].

Применение графических дисплеев в ССНИ позволит оперативно просматривать на экране результаты обработки, а также организовать диалог «экспериментатор — ЭВМ».

Данная работа посвящена рассмотрению некоторых возможностей использования графических дисплеев в океанологических исследованиях.

Классификация и краткий обзор графических дисплеев

Графический дисплей (рис. 1) представляет устройство, в состав которого входят следующие функциональные узлы.

Центральный процессор ЭВМ через блок сопряжения связан с памятью. Назначение блока сопряжения состоит в приеме и выдаче специальных сигналов обмена, приеме данных, их контроле и т. д. В памяти хранится отображаемая информация (массив регенерации). В зависимости от типа данные из памяти поступают на соответствующие генераторы аналоговых сигналов. В качестве таких генераторов используются генератор точек (ГТ),

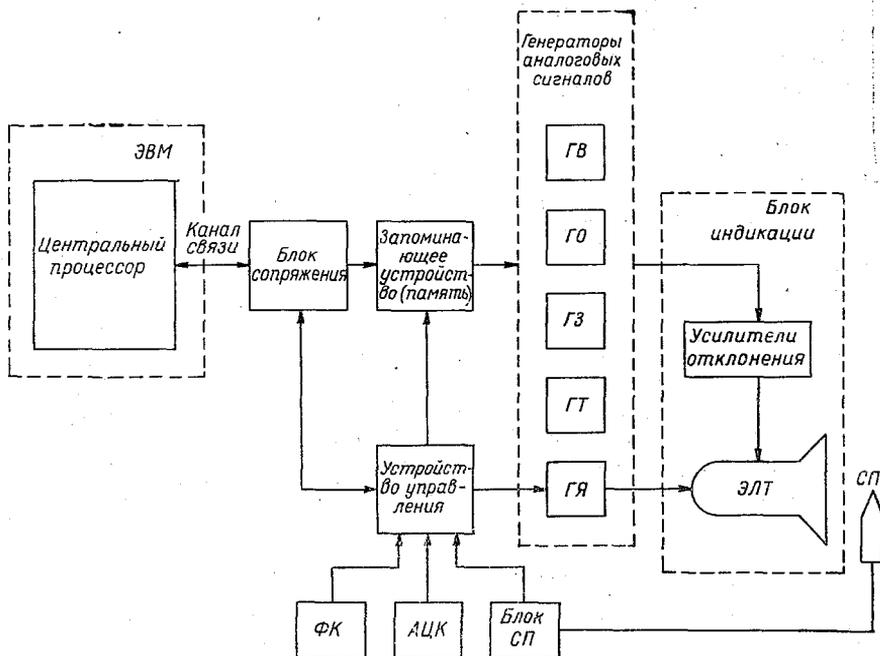


Рис. 1. Структурная схема графического дисплея.

генератор векторов (ГВ), генератор окружностей [ГО], генератор знаков (ГЗ). Усиленные сигналы с выходов этих генераторов с помощью отклоняющей системы управляют движением луча по экрану электронно-лучевой трубки (ЭЛТ). Усилители отклонения и ЭЛТ составляют блок индикации [БИ] графического дисплея.

Управление процессом отображения информации, организация обмена с памятью, декодирование и выдача специальных сигналов осуществляются устройством управления. С помощью функциональной (ФК) и алфавитно-цифровой (АЦК) клавиатуры выполняется ввод командной и буквенно-цифровой информации соответственно. Для оперативной связи с ЭВМ используется световое перо (СП). Оно позволяет указывать на любой элемент

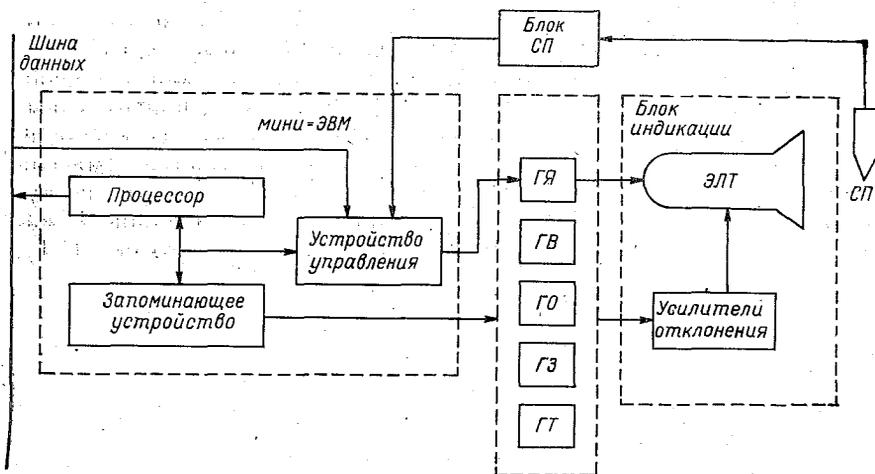


Рис. 2. Схема дисплея на мини-ЭВМ.

изображения (режим «указания»), служит для ввода графических элементов (режим «рисования») в ЭВМ. Сигнал со светового пера усиливается и формируется по уровню в блоке светового пера (блок СП). Этот блок организует «слежение» за движением светового пера. Цифроаналоговый преобразователь (ГЯ) управляет яркостью изображения на экране.

По современным понятиям графический дисплей входит в состав графической системы вместе с ЭВМ и человеком, взаимодействующим с ЭВМ.

В качестве удачного примера можно привести графическую систему, созданную в ИПМ АН СССР [5]. Система позволяет наблюдать на экране дисплея SDS-910 характер протекания во времени любых процессов, моделирование которых происходит на БЭСМ-6. Возможно построение графиков в процессе вычислений, наблюдение за движением тел в пространстве. Система

ориентирована на производство фильмов, снимаемых с экрана дисплея, с целью регистрации результатов научного исследования.

Стремление расширить функциональные возможности дисплея, освободить центральный процессор ЭВМ от процедур, связанных с модификацией изображения, привело к включению в состав дисплея мини-ЭВМ (рис. 2). Кроме того, это связано с увеличением выпуска недорогих быстродействующих мини-ЭВМ, а также с возможностью использования памяти мини-ЭВМ для хранения и регенерации изображения.

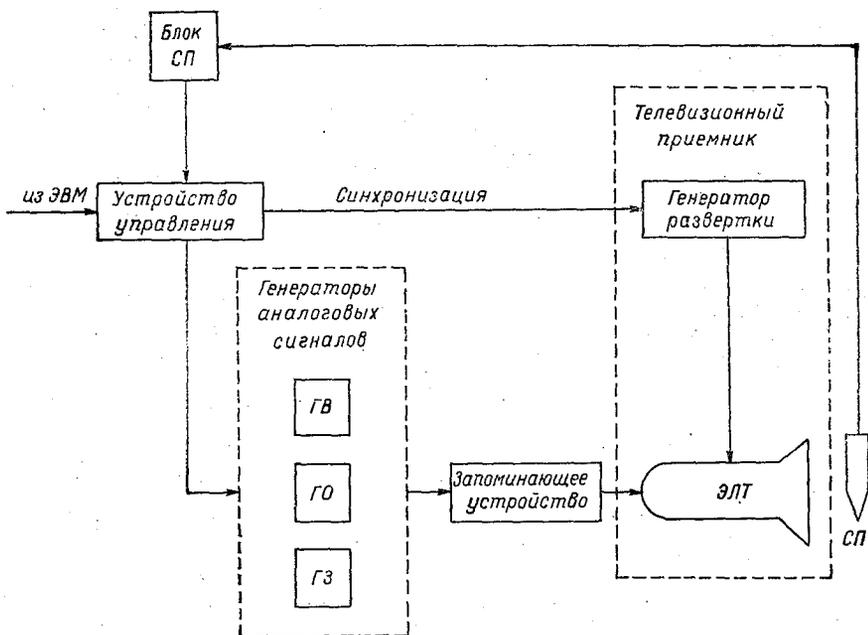


Рис. 3. Схема растрового дисплея.

В ИАЭ СО АН СССР разработан графический дисплей «Дельта» на базе мини-ЭВМ «Электроника-100» [8]. Дисплей «Дельта» снабжен световым пером, функциональной и алфавитно-цифровой клавиатурой. На мини-ЭВМ возложено проведение большинства операций по редактированию и модификации изображения. Графическая система «Дельта Минск-32» предназначена для оперативного взаимодействия исследователя с системой автоматизации экспериментов, цифрового моделирования, а также решения задач автоматизации проектирования.

Графический дисплей представляет собой довольно сложное и дорогостоящее устройство. Особенно сложны в техническом отношении генераторы аналоговых сигналов. Использование этих

генераторов связано с наличием в отображаемой информации однотипных элементов (точек, векторов, окружностей). Последнее наиболее характерно для задач автоматизации проектирования. В системах автоматизации научных исследований применение развитых графических дисплеев не всегда оправдано из-за их высокой стоимости. Кроме того, нерегулярный случайный характер исследуемых процессов и соответственно изображений в океанологических исследованиях позволяет применить в ССНИ существенно более простые графические дисплеи с поточечным выводом изображения. Существует два метода формирования изображения по точкам: координатный и растровый. Сравнивая эти два метода, следует отметить, что первый применим для вывода сравнительно небольшого количества точек на экран ЭЛТ, в то

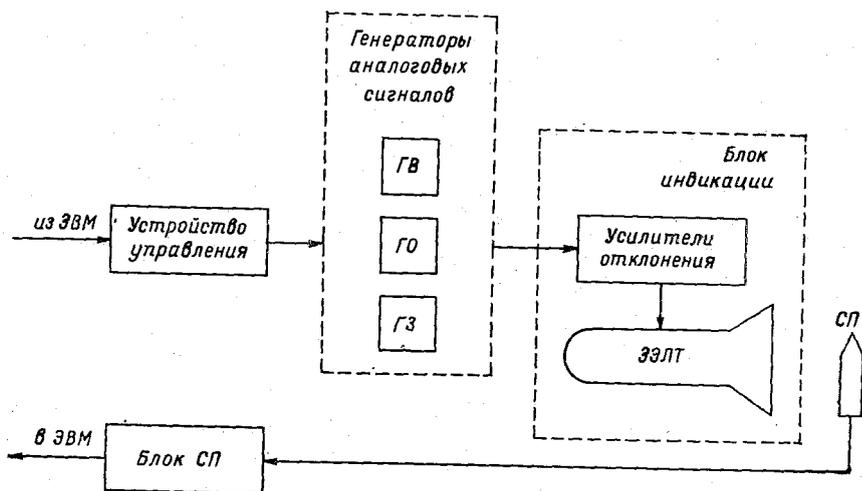


Рис. 4. Схема дисплея на ЭЛТ.

время как второй (рис. 3) позволяет выводить весьма сложные изображения с большим количеством точек. Однако программное обеспечение растрового метода сложнее.

Растровый метод на практике реализуется на базе телевизионной техники. При этом время построения изображения не зависит от сложности последнего, имеется возможность совмещения изображений (в частности, координатной сетки с данными), получения цветных и полутоновых изображений.

Телевизионная техника и каналы телевизионной связи широко распространены. Это обстоятельство может быть использовано в ряде случаев. Проводя исследования в океане, иногда бывает важно получить консультацию специалистов, передав им изображение результатов на береговой вычислительный центр (ВВЦ), или получить на борт НИС с ВВЦ изображение необходимых архивных данных.

Поточечный вывод изображений был использован в устройстве отображения спектрометрической информации в графической системе ПОФИ-2, созданной в лаборатории вычислительной техники и автоматизации ОИЯИ (г. Дубна). Режим диалога с ЭВМ реализован при помощи светового пера [6].

В случаях, когда необходима высокая скорость построения изображения, применяют дисплеи на запоминающих ЭЛТ (ЗЭЛТ). Они (рис. 4) позволяют сохранить построенные изображения длительное время. Хорошее качество изображения и невысокая стоимость вызвали широкое использование ЗЭЛТ в графических дисплеях. Фирма «Tektronix» (США) выпускает устройства 4014, 4015 на ЗЭЛТ, применяемые в геофизических и сейсмических исследованиях [7]. Получили распространение коррелятор 3721А и спектральный индикатор 3720А фирмы «Hewlett — Packard» (США), которые выводят результаты вычислений на осциллоскопы с запоминанием [13]. Японская фирма «Takeda Riken» выпускает систему, предназначенную для измерения и обработки динамических процессов. На осциллографе с запоминанием выполнена система отображения графической информации [13].

Характеристики ряда графических дисплеев, используемых в САНИ, приведены в таблице.

Задачи океанологии и графические системы

Многие исследуемые в океанологии процессы характеризуются широким спектром и большим объемом экспериментальных данных [15]. В последних часто имеют место случайные и систематические ошибки. Для примера рассмотрим задачу обработки данных, полученных при длительных (в течение рейса) исследованиях гидрофизических полей. Эта работа занимает тысячи человеко-часов. Кроме того, не существует уверенности в правильности результатов, так как расчеты ведутся «вслепую». Представим решение этой задачи в рамках графической системы. Необработанные записи данных считываются в систему. Устройство отображения информации в системе показывает кривые или единичные отметки данных. Исследователь редактирует случайные ошибки в необработанных данных путем отбора различных точек и частей кривой световым пером и удаления их из дальнейшего рассмотрения. Когда данные отредактированы, они обращаются в бумажную копию или подвергаются дальнейшей обработке. Такой подход к решению проблемы очистки исходных данных дает прямую экономию во времени, требуемом для обработки данных, и уверенность в более осмысленных данных для дальнейшего анализа.

Оценка сейсмических данных также требует субъективного вмешательства. Сейсмические данные содержат значительный объем помех. Опытный геофизик может «профильтровать» и оценить данные, глядя на воспроизведение, идущее на экран дисп-

Характеристика графических дисплеев, применяемых в научных исследованиях

Фирма, организация									
Control data Corporation	Information Displays	Xerox Data Systems	Vidi	Tektronix		ЕО ЭВМ	ОИЯИ	ИЛЭ	
Страна	США	США	Франция	США	США	СССР	СССР	СССР	
Модель	240	7580		4014	4015	ЕС-7064	ОСК-2	„Дельта“	
Размер экрана, см	диаг. 53,3	25,4×25,4	диаг. 30	38×28	38×28	25×25	17×17	28×28	
Количество цветов	1	1	1	1	1	1	2	1	
Входной код	ASC II, BCD	EBSDIC	ASC II, ISO, CCITTN5	ASC II	ASC II APL	ДКОИ	Код базовой мини-ЭВМ		
Количество адресуемых позиций	1к×1к	1к×1к	1к×1к	4к×4к	4к×4к	1к×1к	1к×1к	1к×1к	
Частота обновления изображения (кадров в 1 с)	50	60	15 . . 50	3ЭЛТ	3ЭЛТ	50	50; 25	50; 25	
Средства редакции и ввода	ФК, АЦК, МК, ШТ, СП, ТВ	ФК, АЦК, СП	ФК, АЦК, СП	ФК, СП, АЦК	ФК, СП, АЦК	ФК, АЦК, СП	ФК, АЦК, СП	ФК, АЦК, СП	

Примечание. МК — магнитный карандаш, ШТ — штурвал, ТВ — планшет (таблетка)

дея прямо с цифровых лент. В. И. Бышев и Ю. А. Иванов в работе [3] указывают на два основных недостатка данных измерений над течениями. Это разрывность временных рядов и значительный уровень шума по сравнению с полезным сигналом. Разрывность связана с неизбежными отказами в работе приборов, дрейфом буев, утратой части буйковых станций с рабочими записями. Встают проблемы объективной интерполяции и фильтрации полей течений. В настоящее время пробелы заполняют, исходя из принципов линейной интерполяции. Вывод имеющихся данных на экран позволит сглаживать данные более высокими порядками полиномов, что будет меньше искажать результаты, а применение итеративного способа решения с просмотром результатов приблизит результат к наилучшему.

Итеративный способ с визуализацией результатов может быть использован и для решения задачи оценки временного масштаба долгопериодных доминирующих колебаний течений [4]. При определении этой оценки проводится низкочастотная фильтрация исходных рядов зональной и меридиональной компонент скорости течения. Параметр фильтрации задается несколько раз для того, чтобы оцениваемый период времени попал в наблюдаемый интервал.

Многие гидрофизические задачи связаны со статистической обработкой временных и пространственно-временных рядов. При этом весьма удобным представляется вывод, кроме исходных, также и некоторых промежуточных и окончательных результатов обработки на световой экран. По промежуточным можно судить о необходимости продолжения расчетов, вызове к действию другой программы обработки или об изменении параметров расчета. Оперативное представление результатов расчета на экране позволяет быстро сделать выводы о характере исследуемых процессов и активного вмешательства в ход проведения расчетов [12].

Графическое изображение информации дает возможность охватить взглядом всю картину распределения физических величин, сократить время полной обработки. Например, при невозможности определения экстремума функции, ввиду большого количества ложных экстремумов, исходная точка поиска решения определяется человеком визуально. Если же задача не может быть решена на ЭВМ (ввиду ее сложности, из-за большого объема, большого количества решений и т. д.), то, по сведениям, которые предоставит исследователю графическое изображение части решения, он, используя свой опыт, знания, интуицию, выберет решение, которое его удовлетворит.

Так, при изучении водных масс Мирового океана «методом ядра» [11] все данные гидрометеорологических станций, полученные в интересующем нас районе, собираются и качественно контролируются. Основные водные массы, образующиеся в результате термохалинных изменений, имеют в различных слоях экстремумы по одной или нескольким характеристикам. Необходимо найти экстремумы, общие для большинства станций, нане-

сти их на карту, через них провести изолинии и определить местоположение основной оси распространения водных масс. Необходимо также получить вертикальный и горизонтальный разрезы параллельно основной оси. После чего выделяются несколько срединных слоев и для каждого из них должна быть изготовлена карта. Просмотр исходных данных со станций, определение экстремумов в ряде случаев, проведение изолиний, вращение пространственного изображения поля в исследуемом районе, определение места, где требуется произвести разрез, получение копий и карт — все это предоставит использование графического дисплея при решении этой задачи.

Графические дисплеи могут быть использованы и для «обывания» каких-то ситуаций в планируемых экспериментах

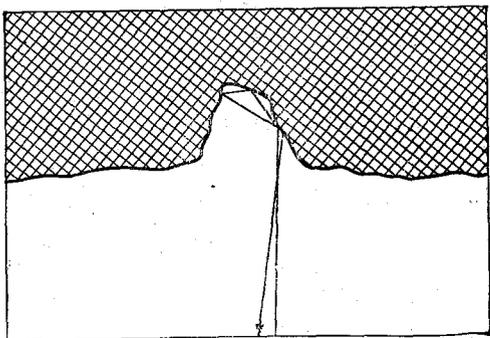


Рис. 5. Участок рельефа дна.

Приведем пример. Одно из направлений химии моря — определение и извлечение полезных веществ из морской воды. Оно связано с многочисленными экспериментами, постановка которых основана на планировании эксперимента. Графическое представление функции отклика на экране дисплея в виде «поверхности отклика» в факторном пространстве поможет быстрее определить экстремум функции отклика, т. е. необходимую стратегию эксперимента.

Важным вопросом является автоматизация обработки и анализа геолого-геофизических наблюдений при изучении рельефа дна. Получаемые данные, обычно при эхолотировании, несут в себе ряд принципиальных ошибок, например, при таком рельефе (рис. 5), т. е. резком изменении глубины. Просмотр получаемых с эхолота данных на экране дисплея и их редактирование в реальном масштабе времени, вывод достоверных данных на карту значительно повысил бы качество результатов и уменьшил бы время оформления отчетов.

Применение графических дисплеев в ССАНИ позволит решать такие задачи, решение которых раньше было нереально.

ак, при анализе вынужденных возмущений синоптической изменчивости течений выяснилось, что они являются шумом [9]. Одним из путей подавления этих возмущений является рассмотрение вместо отдельных реализаций интегральных характеристик, затем их осреднение. При интегрировании возникает проблема задания соответствующих весов информации, подлежащей интегрированию. Итеративно изменяя веса и сравнивая на экране дисплея исходную реализацию с расчетной, легко и удобно можно разрешить эту проблему. Существует и другой путь. Используя возможность получения на экране дисплея изображений различных масштабах, исходную реализацию в виде шума можно подвергнуть визуальному редактированию. Неоценимую помощь окажет применение дисплея при изучении пространственной структуры гидрофизических полей. В этих задачах необходимо трехмерное представление поля, вращение его вокруг осей, получение разрезов и т. д.

Иногда возникают ситуации, когда отдельные массивы срочных данных необходимо учитывать в расчетах. Такое возможно при получении спутниковых или радиозондовых данных. Результаты решения этой задачи помогут более точному прогнозированию. В программном обеспечении графической системы имеется архив, в котором хранится изображение любой части земного шара. Вызывая на экран нужную часть в приемлемом масштабе, вводим вновь полученные данные с помощью светового пера и задаем выполнение расчета метеобстановки снова. Результат расчета выводится на карту при помощи графопостроителя.

В конечном итоге применение графических дисплеев поможет перейти к более прогрессивной, чем существующая, форме эксперимента — управляемому эксперименту, при котором обработка данных ведется в реальном масштабе времени и режим наблюдений корректируется с учетом результатов обработки. В ходе обработки и анализа океанологических наблюдений будет осуществляться обратная связь с подсистемой сбора информации, т. е. решаться задача о наилучшем размещении в пространстве и во времени датчиков и приборов, измеряющих элементы полей, о выборе достаточной дискретности отсчетов, чувствительности датчиков, разрядности кодов и других параметров аппаратуры.

Заключение

Применение графических дисплеев в САНИ, а также в ССАНИ чрезвычайно перспективно. Графические устройства планируется использовать почти во всех фазах океанологических экспериментов. Многие из применений требуют оперативного отображения информации, с помощью которой океанолог сможет управлять и модифицировать процесс наблюдений. Оперативные графические дисплеи должны использоваться для контроля и управления измерительным комплексом, системами накопления и системами

обработки и корректировки информации. Помимо значительного увеличения точности определения океанологических полей, появляется возможность комплексной обработки информации.

Применение графических дисплеев в ССАНИ сократит время и повысит качество обработки значительных объемов экспериментальной информации. Это повысит эффективность, а также культуру труда океанолога-экспериментатора. Расширится доступ пользователей к ЭВМ, так как не потребует от них знаний по вычислительной технике и программированию.

Использование главного качества графических систем — оперативного взаимодействия человека с ЭВМ — сделает возможным проведение управляемых экспериментов.

Автор выражает благодарность В. П. Непрочнову, В. В. Здоренину, Б. Н. Гринько, В. И. Бышеву, оказавшим существенную помощь в подборе задач для данной статьи.

СПИСОК ЛИТЕРАТУРЫ

1. Арван Б. А. и др. Оперативная обработка геофизической информации на судовой ЭВМ. — В кн.: Вычислительные и статистические методы интерпретации сейсмических данных (вычислительная сейсмология). Вып. 6. — М.: Наука, 1973.
2. Беляев В. И. Обработка и теоретический анализ океанографических наблюдений. — Киев: Наукова думка, 1973.
3. Бышев В. И., Иванов Ю. А. Некоторые результаты исследования реакции поля течений на возмущения атмосферы. — В кн.: Атлантический гидрофизический полигон-70. — М.: Наука, 1974.
4. Бышев В. И., Чекотилло К. А. Особенности крупномасштабного движения вод в океане. — В кн.: Атлантический гидрофизический полигон-70. — М.: Наука, 1974.
5. Боровин Г. К. и др. Система вывода информации на экран дисплея ИПМ АН СССР. — М.: Препринт № 48, 1973.
6. Злоказов В. Б., Нефедьева Л. С. Математическое обеспечение осциллографа со световым пером в системе ПОФИ-2. — Дубна: ОИЯИ, 1971.
7. Каталог/Tektronix—1974.
8. Ковалев А. М. и др. Универсальный графический дисплей с мини-ЭВМ. — Управляющие системы и машины, 1974, № 5.
9. Корт В. Г., Бышев В. И., Тарасенко В. М. Синоптическая изменчивость течений на Атлантическом полигоне. — В кн.: Атлантический гидрофизический полигон-70. — М.: Наука, 1974.
10. Нестерихин Ю. Е. и др. Организация систем автоматизации научных исследований (проблемы, методы, перспективы). — Автометрия, 1974, № 4.
11. Океанографическая энциклопедия. — Л.: Гидрометеоиздат, 1974.
12. Печискер Л. Г. Небольшая графическая система для обработки и анализа рядов океанологических данных. — Океанология, 1976, т. 16, вып. 5, с. 929—934.
13. Попов В. А. Аппаратура иностранных фирм для статистического анализа (обзор). — В кн.: Прикладной анализ случайных сигналов. Новосибирск, 1973.
14. Ситников Л. С. Состояние и задачи автоматизации эксперимента в океанологии. — Автометрия, 1974, № 4.
15. Тимченко И. Е. Измерение и расчет случайных полей океана с использованием автоматизированных систем сбора и обработки наблюдений. — В кн.: Материалы IV Всесоюзной школы «Автоматизация научных исследований морей и океанов». Севастополь, МГИ АН УССР, 1973.

Р. Л. Каган, Е. Е. Сибир, Е. И. Хлебникова

О МОДЕЛИРОВАНИИ СВЯЗНЫХ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ С РАВНОМЕРНЫМ РАСПРЕДЕЛЕНИЕМ

1. Для решения задач, связанных с обработкой временных рядов, оказывается полезным использование методов статистического моделирования. Эти методы хорошо разработаны применительно к гауссовским случайным процессам [1, 2]. В случае негауссовских процессов моделирование связанных рядов осложняется. В этом случае обычно используется метод обратного преобразования, состоящий в моделировании связанных гауссовских величин с последующим преобразованием их в величины, распределенные в соответствии с заданным законом. Естественно, что при этом приходится учитывать трансформацию характеристик многомерного распределения, в частности характеристик корреляции между членами ряда. Алгоритмы, основанные на таком принципе, являются очень трудоемкими. Гораздо более экономичными могли бы быть алгоритмы, основанные на использовании предварительно построенных связанных последовательностей величин, обладающих равномерным распределением в заданном интервале. Использование таких алгоритмов могло бы быть особенно перспективным при моделировании процессов, для которых легко находится функция, обратная к функции распределения. К числу таких распределений относятся, например, распределения Релея — Райса и Вейбулла, широко используемые при статистическом описании скорости ветра.

В работе Д. Я. Ратковича [3] для моделирования связанных равномерно распределенных последовательностей с заданной корреляцией был использован метод, предложенный И. О. Сармановым [4], в основу которого положено использование представляемой моделируемой величины в виде разложения в ряд по двойным полиномам Лежандра. Такое разложение, как показано в [5], в пределе при стремлении числа членов ряда к бесконеч-

ному обеспечивает получение связанных величин с заданной корреляцией. При использовании ограниченного числа членов ряда моделирование, естественно, осуществляется приближенно. В частности, оказывается, что для обеспечения точного воспроизведения статистических характеристик при коэффициентах корреляции r между смежными членами ряда больше 0,3 необходимо удержание более трех членов ряда, а при $r > 0,5$ — более пяти членов ряда. Таким образом, точное моделирование рядов высокой связности этим методом также требует достаточно трудоемких расчетов.

Ниже будет изложен способ получения связанных равномерно распределенных величин, который при сравнительно небольшой трудоемкости его обеспечивает точное воспроизведение коэффициента корреляции между величинами в диапазоне $0 \leq r \leq 1$. Будут представлены некоторые результаты реализации этого метода и использования полученных равномерно распределенных последовательностей для моделирования величин, распределение которых описывается законом Вейбулла.

2. Пусть x_i — равномерно распределенные в интервале $(0,1)$ случайные бессвязные величины. Образует на их основе величины z_i по следующему алгоритму:

$$z_0 = x_0,$$

$$y_i = \frac{p}{1+p} z_{i-1} + \frac{1}{1+p} x_i,$$

$$z_i = \begin{cases} \frac{(1+p)^2}{2p} y_i^2, & 0 \leq y_i \leq A, \\ y_i / (1-A) - B, & A \leq y_i \leq 1-A, \\ 1 - \frac{(1+p)^2}{2p} (1-y_i)^2, & 1-A \leq y_i \leq 1, \end{cases} \quad (1)$$

где p — некоторая постоянная положительная величина,

$$A = \min\left(\frac{1}{1+p}, \frac{p}{1+p}\right),$$

$$B = \frac{A}{2(1-A)}.$$

Нетрудно показать, что полученные таким образом величины обладают равномерным распределением на $(0, 1)$.

Очевидно также, что величины z_i являются связными, причем корреляция их r между смежными членами последовательности существенно зависит от значения параметра p .

Можно показать, что двумерная плотность распределения смежных величин z_i, z_{i+1} имеет вид:

$$p(z_i, z_{i+1}) = \begin{cases} \frac{1}{\sqrt{2qz_{i+1}}}, & \text{если } \left\{ z_i \leq q, \frac{z_i^2}{2q} \leq z_{i+1} \leq B \right\}, \\ \frac{1}{\sqrt{2q(1-z_{i+1})}}, & \text{если } \left\{ z_i \geq 1-q, 1-B \leq z_{i+1} \leq 1 - \frac{(1-z_i)^2}{2q} \right\}, \\ \frac{1}{A(1+q)}, & \text{если } \{ 1-q \leq z_i \leq q, B \leq z_{i+1} \leq 1-B \}, \\ & \text{или } \{ z_i \leq q, z_i \leq 1-q, B \leq z_{i+1} \leq z_i + B \}, \\ & \text{или } \{ z_i \geq q, z_i \geq 1-q, z_i - B \leq z_{i+1} \leq 1-B \}, \\ & \text{или } \{ q \leq z_i \leq 1-q, z_i - B \leq z_{i+1} \leq z_i + B \}, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (2)$$

Здесь и далее $q=1/p$.

Используя эту функцию, получаем выражения для некоторых смешанных центральных моментов до 4-го порядка включительно.

При $q < 1$

$$\begin{aligned} \mu_{11}(1) &= r(1) = 1 - \frac{q^2}{2} + \frac{q^3}{5}, \\ \mu_{21}(1) &= 0, \\ \mu_{31}(1) &= \frac{9}{5} - \frac{3}{2}q^2 + \frac{9}{5}q^3 - \frac{6}{5}q^4 + \frac{12}{35}q^5, \\ \mu_{22}(1) &= \frac{9}{5} - 2q^2 + \frac{12}{5}q^3 - \frac{8}{7}q^4 + \frac{1}{5}q^5. \end{aligned} \quad (3)$$

При $q > 1$

$$\begin{aligned} \mu_{11}(1) &= r(1) = 1/q - \frac{3}{10q^2}, \\ \mu_{21}(1) &= 0, \\ \mu_{31}(1) &= \frac{9}{5q} - \frac{39}{70q^2}, \\ \mu_{22}(1) &= 1 + \frac{2}{5q^2} - \frac{1}{7q^3}. \end{aligned} \quad (4)$$

Получаемая описанным образом последовательность z_i является марковской, так что многомерные плотности распределения ее могут быть получены из двумерной. В частности, для $q > 1$ коэффициент корреляции величин z_i и z_{i+2} может быть рассчитан по формуле

$$\mu_{11}(2) = \frac{1}{q^2} - \frac{19}{30q^3} + \frac{3}{20q^4} - \frac{1}{28q^5} + \frac{1}{144q^6}. \quad (5)$$

Указанный алгоритм был реализован на ЭВМ БЭСМ-6. Поскольку имеющиеся в математическом обеспечении этой машины подпрограммы получения равномерно распределенных псевдослучайных чисел, написанные на языках АЛГОЛ и ФОРТРАН, не обеспечивают надлежащей точности моделирования, мы были вынуждены специально составить соответствующую подпрограмму на автокоде МАДЛЕН, реализующую мультипликативный датчик с учетом рекомендаций, содержащихся в [2].

Ниже приводятся некоторые результаты расчета одноточечных и смешанных моментов для разных сдвигов между членами последовательности в зависимости от задания параметра q .

Таблица 1

Выборочные моменты равномерно распределенной последовательности, полученной по алгоритму (1) (длина реализации $m=1200$)

Номер выборки	Число реализаций	q	\bar{z}	σ_z^2	A_z	E_z
1	50	0,25	0,501	0,080	-0,013	-1,145
2	50	0,50	0,502	0,082	-0,016	-1,195
3	50	0,75	0,500	0,082	-0,006	-1,193
4	50	1,00	0,500	0,082	-0,003	-1,189
5	20	1,27	0,496	0,083	0,013	-1,197
6	20	1,63	0,497	0,083	0,014	-1,202
7	20	2,15	0,497	0,083	0,016	-1,205
8	20	3,00	0,497	0,083	0,017	-1,206
9	20	4,68	0,498	0,083	0,018	-1,205
10	20	9,69	0,498	0,083	0,019	-1,203
Теоретическое	—	—	0,500	0,082	0,000	-1,200

В табл. 1 приводятся значения моментов (среднего \bar{z} , дисперсии σ_z^2 , коэффициента асимметрии A_z , коэффициента эксцесса E_z), полученные по ряду выборок, в сопоставлении с теоретическими значениями для равномерного распределения. Из таблицы видно, что отклонения выборочных значений моментов от теоретических находятся в пределах, допустимых для использованного объема выборок.

К аналогичному выводу приводит рассмотрение табл. 2, в которой представлены выборочные значения $\tilde{\mu}_{ij}(1)$ двухточечных моментов для смежных членов последовательности и теоретические значения их $\mu_{ij}(1)$, определяемые по формулам (3) и (4). Таким образом, указанный алгоритм действительно обеспечивает реализацию поставленной задачи моделирования равномерно распределенных связанных последовательностей.

Выборочные (μ_{ij}) и теоретические ($\tilde{\mu}_{ij}$) значения моментов двумерного распределения для смежных членов равномерной последовательности (длина реализации $m=1200$; при $q \leq 1$ число реализаций $N=50$; при $q > 1$ $N=20$)

μ_{ij}	q									
	0,25	0,50	0,75	1,00	1,27	1,63	2,15	3,00	4,68	9,69
$\tilde{\mu}_{11}$	0,97	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10
μ_{11}	0,97	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10
$\tilde{\mu}_{31}$	1,78	1,59	1,42	1,25	1,07	0,89	0,71	0,53	0,35	0,17
μ_{31}	1,73	1,59	1,42	1,24	1,07	0,89	0,72	0,54	0,36	0,18
$\tilde{\mu}_{22}$	1,76	1,54	1,38	1,26	1,18	1,12	1,07	1,04	1,01	1,00
μ_{22}	1,71	1,53	1,37	1,26	1,18	1,12	1,07	1,04	1,02	1,00

Практический интерес представляет поведение моментов не только для смежных членов последовательности, но и более удаленных, т. е. для сдвигов τ больше 1. Из формулы (5) видно, что отклонение $\tilde{\mu}_{11}$ (2) от значения $(\tilde{\mu}_{11}(1))^2$ незначительно. Расчет моментов $\tilde{\mu}_{11}(\tau)$ для $\tau > 2$ показывает, что убывание корреляции между членами последовательности при увеличении сдвига между ними близко к экспоненциальному. Это хорошо видно из рис. 1, на котором представлена зависимость $\tilde{r}(\tau) = \tilde{\mu}_{11}(\tau)$ в полупологарифмической системе координат. Выборочные значения коэффициентов корреляции очень хорошо описываются зависимостью вида

$$\tilde{r}(\tau) = (\tilde{r}(1))^\tau, \quad (6)$$

т. е. существует такая функция f , что $\tilde{r}(k\tau) = f(\tilde{r}(\tau))$ для любых k и τ . В связи с этим представляет интерес выяснить, имеет ли место однозначная связь между моментами $\tilde{r}(\tau)$ и $\mu_{ij}(\tau)$. Из рис. 2, на котором представлены зависимости от $\tilde{r}(\tau)$ моментов $\tilde{\mu}_{31}(\tau)$ (рис. 2 а) и $\tilde{\mu}_{22}(\tau)$ (рис. 2 б), полученные в результате обработки реализаций длиной в 1200 членов (число реализаций варьировалось от 20 до 50), видно, что такая однозначная связь

имеет место лишь для моментов μ_{31} . Для моментов μ_{22} характер связи с $\tilde{r}(\tau)$ при разных q заметно различается. Тем более можно ожидать различий в моментах распределения более высоких порядков. Следовательно, при использовании полученных по данному алгоритму равномерно распределенных последовательностей для моделирования последовательностей, распределенных по другому закону, можно ожидать, что уже ход корреляционных функций для таких величин будет существенно зависеть от зада-

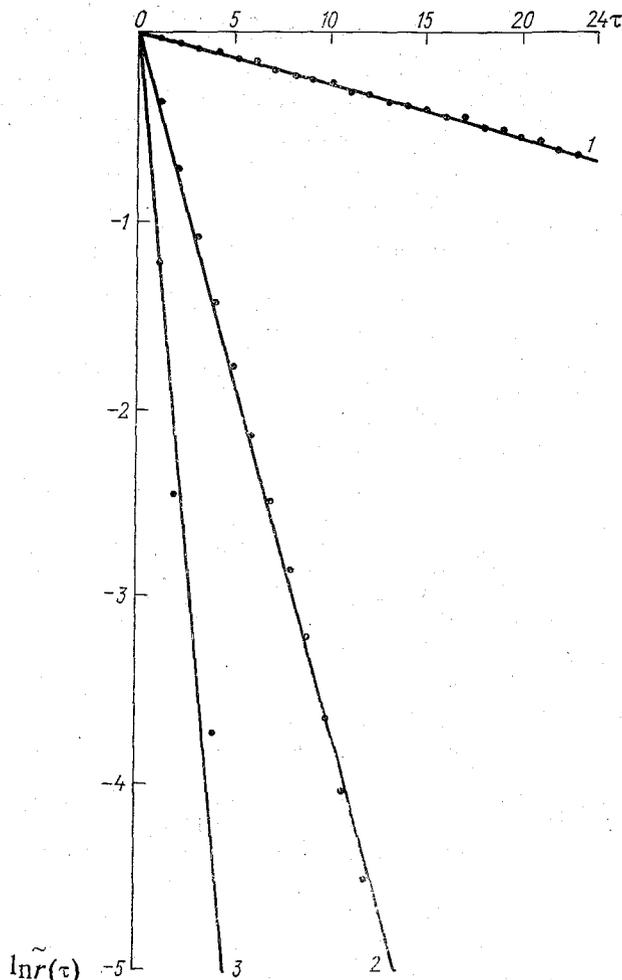


Рис. 1. Корреляционные функции равномерно распределенных последовательностей, полученных по алгоритму (1).

1) $q=0,25$; 2) $q=1,0$; 3) $q=3,0$.

ния параметра q или, что то же самое, от задания коэффициента корреляции $r(1)$ между смежными членами равномерной последовательности.

3. Описанный выше алгоритм был использован для моделирования последовательностей w_i , функция распределения которых описывается законом Вейбулла

$$F(V) = P\{v_i \leq V\} = 1 - \exp(-(V/\beta)^\gamma). \quad (7)$$

Из (7) следует, что при подстановке в формулу

$$v_i = \beta(-\ln(1 - z_i))^{1/\gamma} \quad (8)$$

равномерно распределенных значений z_i мы получим значения v_i , распределенные в соответствии с законом Вейбулла.

Моделирование по указанной формуле осуществлялось для нескольких значений параметра γ , существенно определяющих характер изменчивости величины v . Поскольку параметр β определяет лишь масштаб величины v и не сказывается на характе-

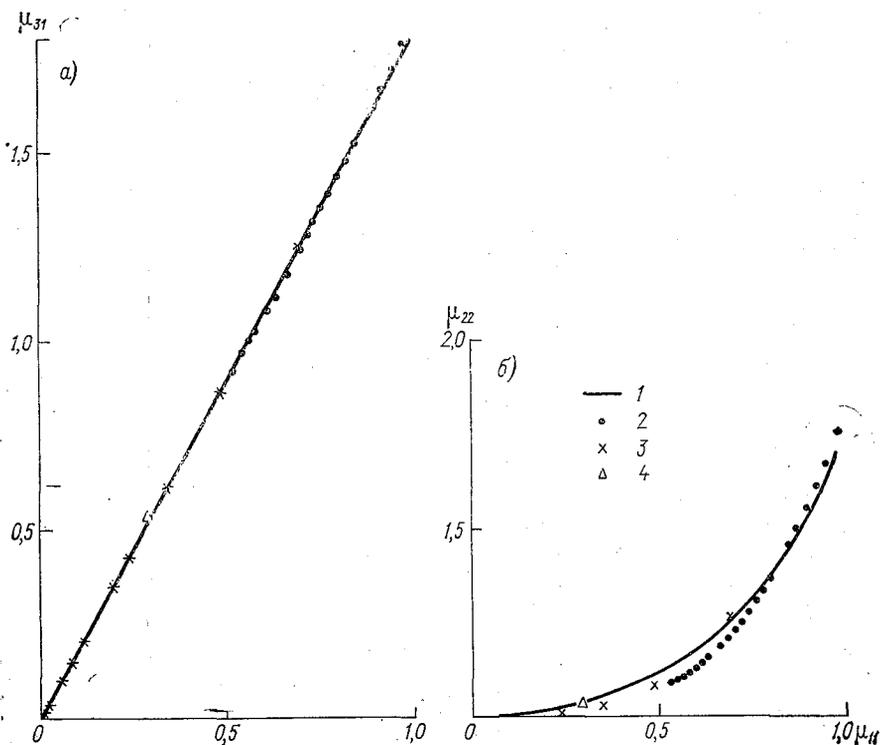


Рис. 2. Связь между моментами μ_{31} и μ_{11} (а) и моментами μ_{22} и μ_{11} (б) для последовательности, полученной по алгоритму (1).

1) Теоретические значения для смежных членов последовательности; выборочные значения для разных сдвигов: 2) $q=0,25$; 3) $q=1,0$; 4) $q=3,0$.

ре ее изменчивости, этот параметр не варьировался и был принят равным 1.

В табл. 3 для различных γ представлены значения моментов одномерного распределения (среднего \bar{v} , дисперсии σ^2 , коэффициента асимметрии A и эксцесса E), рассчитанные по данным 50 реализаций (длиной в 1200 членов) последовательностей v_i , моделирование которых осуществлялось по формуле (8) с использованием для получения равномерно распределенных связанных чисел z_i алгоритма (1) (при $p=1$). Сопоставление этих

Выборочные и теоретические моменты вейбулловской последовательности

γ	Способ получения	\bar{v}	σ^2	A	E
0,6	Теоретический	1,50	7,00	4,59	37,48
	Выборочный	1,49	6,76	4,67	37,10
1,0	Теоретический	1,00	1,00	2,00	6,00
	Выборочный	0,99	0,97	2,02	6,24
1,4	Теоретический	0,91	0,44	1,20	1,84
	Выборочный	0,91	0,42	1,20	1,92

значений моментов с теоретическими значениями моментов распределения Вейбулла, также представленными в табл. 3, свидетельствует об их согласии.

О характере изменения корреляции с увеличением сдвига между членами вейбулловской последовательности, смоделированной по предложенному алгоритму, можно судить по рис. 3, на котором для различных значений параметров q и γ представлены выборочные корреляционные функции соответствующих последовательностей.

В заключение отметим, что использование для моделирования связанных равномерно распределенных чисел алгоритма (1) требует примерно в 40 раз меньше машинного времени, чем применение метода, основанного на преобразовании связанных гаус-

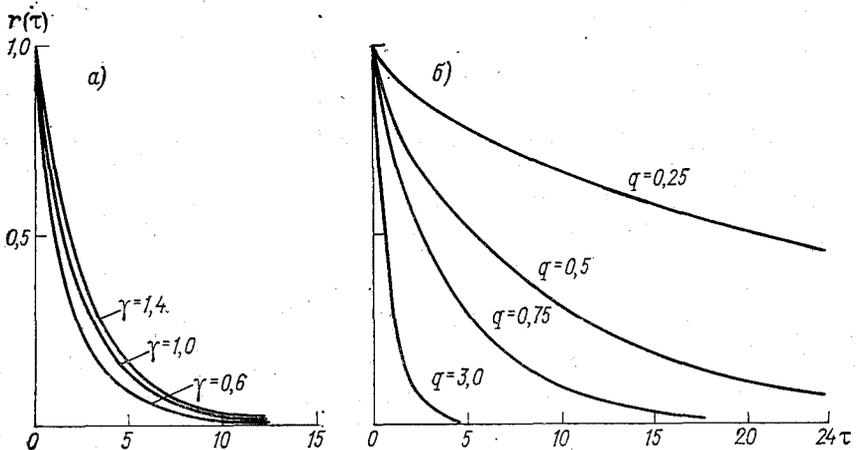


Рис. 3. Выборочные корреляционные функции вейбулловских последовательностей.

а) $q=1,0$; б) $\gamma=1,4$.

совских последовательностей. Поэтому использование алгоритма (1) для моделирования негауссовских последовательностей представляется весьма целесообразным. Особенно полезно его использование в тех случаях, когда надежные сведения о связности моделируемого процесса имеются лишь для смежных членов последовательности, что нередко имеет место для метеорологических рядов.

СПИСОК ЛИТЕРАТУРЫ

1. Выков В. В. Цифровое моделирование в статистической радиотехнике. — М.: Сов. радио, 1971. — 326 с.
2. Ермаков С. М., Михайлов Г. А. Курс статистического моделирования. — М.: Наука, 1976. — 319 с.
3. Раткович Д. Я. Закономерности чередования маловодных и многоводных лет как основа расчетов регулирования речного стока. — Труды ГГИ, 1969, вып. 162, с. 62—173.
4. Сарманов И. О. Построение корреляции между равномерно распределенными случайными величинами. — Труды ГГИ, 1968, вып. 160, с. 81—89.
5. Сарманов О. В. Исследование стационарных марковских процессов методом разложения по собственным функциям. — Труды Математического ин-та им. В. А. Стеклова АН СССР, 1961, т. 60, с. 238—261.

Р. Л. Каган, Е. Е. Сибир

О СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИКАХ НЕКОТОРЫХ МОДЕЛЕЙ СВЯЗНЫХ ВЕЙБУЛЛОВСКИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

1. При статистической обработке данных наблюдений за скоростью ветра широко используется распределение Вейбулла (см., например, [1, 7]). В связи с этим представляет интерес разработка статистических моделей, которые позволяли бы адекватно описывать это распределение и, в частности, имитировать ряды, соответствующие данному распределению, на ЭВМ.

Статистическое моделирование бессвязных вейбулловских рядов не представляет существенных трудностей и может осуществляться стандартными методами [2, 3]. Однако реальные метеорологические ряды в большинстве случаев характеризуются заметной связностью, которую желательно учитывать и при моделировании. Для этого необходимо знать статистические характеристики этой связи. К сожалению, данными о многомерных функциях распределения временных рядов для скорости ветра мы не располагаем, в настоящее время имеется мало сведений даже о временной корреляционной функции этого элемента, не говоря уже о моментных функциях более высокого порядка. Поэтому при построении многомерных вейбулловских моделей для описания этих рядов имеется значительный произвол. Может быть предложен ряд различных моделей, каждая из которых обеспечивает правильное воспроизведение маргинальных одномерных распределений и заданной временной корреляционной функции. Вместе с тем моментные функции более высоких порядков у них, вообще говоря, различны. Соответственно могут различаться и многие важные для практики статистические характеристики.

В работах А. С. Марченко [8] предложено получать связанные вейбулловские ряды путем степенного преобразования предварительно моделированных связанных гауссовских последовательностей. Другой вариант использования таких последовательностей методом обратного преобразования был предложен в [6]. Там же приведены некоторые результаты, характеризующие различия этих алгоритмов.

С точки зрения экономичности расчетов при больших объемах моделируемых выборок представляется заманчивым использование предложенного в [5] алгоритма, в котором вейбулловские последовательности моделируются путем преобразования равномерно распределенных рядов, минуя промежуточный этап преобразования таких рядов в гауссовские. Как показано в [5], моделирование вейбулловских последовательностей с помощью этого алгоритма требует в десятки раз меньше машинного времени.

Разумеется, без сравнения с фактическими данными нельзя гарантировать, что хотя бы один из трех перечисленных алгоритмов удовлетворительно им соответствует. Такое сравнение может производиться разными способами. В работах [9, 10] сравниваются не основные статистические характеристики моделируемых и фактических данных, а сравнительно сложно с ними связанные величины, такие, как число и продолжительность выбросов. Совпадение их, (всегда, конечно, приближенное), однако, не обязательно означает, что будут совпадать и другие статистические характеристики, потребность в расчете которых может возникнуть. Представляется поэтому полезным сопоставлять также основные статистические характеристики, соответствующие рассматриваемым моделям, в качестве которых естественно рассматривать их моментные функции.

Расчет моментных функций может быть выполнен путем численного интегрирования соответствующих функций распределения. Поскольку, однако, для практических целей большой точности задания их не требуется, представляется более целесообразным использование метода статистического моделирования. При этом мы, с одной стороны, имеем возможность проверить практическую пригодность моделирующего алгоритма. С другой стороны, все интересующие нас статистические характеристики моделируемых рядов могут быть получены путем непосредственного расчета на ЭВМ. Такие расчеты были нами выполнены применительно к трем упомянутым выше алгоритмам. Ниже приводятся некоторые из полученных при этом результатов.

2. Дадим схематическое описание использованных алгоритмов, которые мы впредь для краткости будем называть алгоритмами А, Б и В соответственно.

А. *Преобразование равномерно распределенных связанных последовательностей.*

В качестве исходных используются равномерно распределенные в интервале (0,1) случайные величины, на основе которых получают связанные величины z_i по формулам [5]:

$$z_0 = x_0,$$

$$z_i = \begin{cases} \frac{(1+q)^2}{2q} y_i^2 & \text{при } 0 \leq y_i \leq A, \\ \frac{y_i - A/2}{1-A} & \text{при } A \leq y_i \leq 1-A, \\ 1 - \frac{(1+q)^2}{2q} (1-y_i)^2 & \text{при } 1-A \leq y_i \leq 1, \end{cases} \quad (1)$$

где

$$y_i = \frac{1}{1+q} (z_{i-1} + qx_i), \quad (2)$$

q — некоторая постоянная положительная величина,

$$A = \begin{cases} \frac{q}{1+q} & \text{при } q \leq 1, \\ \frac{1}{1+q} & \text{при } q \geq 1. \end{cases} \quad (3)$$

Преобразование чисел z_i в вейбулловские производится по формуле

$$v_i = \beta(-\ln(1 - z_i))^{1/\gamma}. \quad (4)$$

Б. Обратное преобразование гауссовских связанных последовательностей.

В этом случае также используется преобразование (4) по отношению к равномерно распределенным числам z_i , которые получаются из гауссовских чисел y_i по формуле

$$z_i = \frac{1}{2\pi} \int_{-\infty}^{y_i} \exp\left(-\frac{\xi^2}{2}\right) d\xi, \quad (5)$$

где y_i — нормально распределенные, нормированные (0,1) связанные числа. Они получаются из псевдослучайных гауссовских, нормированных (0,1) последовательностей чисел t_i по обычному алгоритму моделирования гауссовских марковских рядов

$$y_0 = t_0, \quad y_i = py_{i-1} + \sqrt{1-p^2} t_i, \quad (6)$$

где p численно равно коэффициенту корреляции между смежными членами последовательности t .

Напомним, что при выработке на ЭВМ гауссовских псевдослучайных чисел, как правило, осуществляется сглаживание нескольких предварительно выработанных равномерно распределенных чисел. Наряду с необходимостью расчета интеграла вероятностей (5) это и определяет значительно большую трудоемкость алгоритма Б.

В. Степенное преобразование гауссовских связанных последовательностей.

В этом преобразовании, предложенном в [8], в качестве исходных берутся два связанных гауссовских ряда, каждый из которых получается по формуле типа (6), но при использовании различных псевдослучайных чисел t_i , что обеспечивает некоррелированность этих рядов:

$$\begin{aligned} y_{10} &= \sqrt{\lambda} t_0, & y_{20} &= \sqrt{\lambda} t_1, \\ y_{1i} &= py_{1i-1} + \sqrt{\lambda(1-p^2)} t_{2i}, \\ y_{2i} &= py_{2i-1} + \sqrt{\lambda(1-p^2)} t_{2i-1}, \end{aligned} \quad (7)$$

где $\lambda = \beta^\gamma$, а β и γ представляют собой параметры распределения Вейбулла. Вейбулловская последовательность получается по формуле

$$v_i = \left(\frac{y_{1i}^2 + y_{2i}^2}{2} \right)^{1/\gamma} \quad (8)$$

Каждый из этих трех алгоритмов обеспечивает получение последовательностей, одномерное распределение которых описывается законом Вейбулла

$$F(V) = P(v_i < V) = 1 - \exp\left(-\left(\frac{V}{\beta}\right)^\gamma\right) \quad (V > 0) \quad (9)$$

с плотностью распределения

$$f(V) = \begin{cases} \frac{\gamma}{\beta} \left(\frac{V}{\beta}\right)^{\gamma-1} \exp\left(-\left(\frac{V}{\beta}\right)^\gamma\right) & \text{при } V \geq 0, \\ 0 & \text{при } V \leq 0, \end{cases} \quad (10)$$

Начальные моменты одномерного распределения описываются простыми формулами

$$m_k = \overline{v^k} = \int_0^\infty \xi^k f(\xi) d\xi = \beta^k \Gamma\left(1 + \frac{k}{\gamma}\right), \quad (11)$$

где Γ — гамма-функция.

Легко получаются и другие статистические характеристики одномерного распределения Вейбулла. Для ряда характеристик значения их в наиболее употребительном диапазоне изменений параметра γ приводятся в приложении 1. В этой таблице приняты следующие обозначения: \bar{v} — среднее значение величины v , σ — ее среднее квадратическое отклонение, $c_v = \sigma/\bar{v}$ — коэффициент изменчивости, A и E — коэффициенты асимметрии и эксцесса, $V(F)$ — квантиль распределения Вейбулла обеспеченности F , т. е. значения величины v , вероятность непревышения которой составляет F (соответственно $1-F$ — вероятность ее превышения).

Моменты двумерных распределений удобно вычислять непосредственно по моделированным рядам. Для этой цели указанные алгоритмы были реализованы на ЭВМ БЭСМ-6, на которой моделировались выборки вейбулловских последовательностей, состоящих из ряда (m от 20 до 50) реализаций, длина каждой из которых n составляла 1224 члена. В действительности при расчетах одномерного распределения использовалось лишь 1200 первых членов, остальные применялись при расчете характеристик связи с целью исключения их смещения за счет неодинакового объема выборок (эти характеристики оценивались лишь для сдвигов, не превышающих 24 интервала между членами выборки).

Рассчитывались нормированные центральные моменты связи до 4-го порядка включительно:

$$\begin{aligned} \mu_{11}(\tau) &= r(\tau) = \frac{(\overline{v^I - \bar{v}})(\overline{v^{II} - \bar{v}})}{\sigma^2}, \\ \mu_{21}(\tau) &= \frac{(\overline{v^I - \bar{v}})^2(\overline{v^{II} - \bar{v}}) + (\overline{v^I - \bar{v}})(\overline{v^{II} - \bar{v}})^2}{2\sigma^3}, \\ \mu_{31}(\tau) &= \frac{(\overline{v^I - \bar{v}})^3(\overline{v^{II} - \bar{v}}) + (\overline{v^I - \bar{v}})(\overline{v^{II} - \bar{v}})^3}{2\sigma^4}, \\ \mu_{22}(\tau) &= \frac{(\overline{v^I - \bar{v}})^2(\overline{v^{II} - \bar{v}})^2}{\sigma^4}. \end{aligned} \quad (12)$$

Здесь черта сверху означает арифметическое осреднение по всему объему выборки. Индексы I и II означают, что в каждой реализации рассматриваются члены от 1-го до 1200-го и от $\tau+1$ до $1200+\tau$ соответственно. Средние значения \bar{v} и дисперсия σ^2 определялись по основному объему выборки (в каждой реализации использовались члены от 1-го до 1200-го включительно).

С целью сокращения расчетов непосредственно вычислялись начальные моменты

$$m_{kl}(\tau) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (v_j^i)^k (v_{j+\tau}^i)^l, \quad (13)$$

где n — объем реализации, m — число их, верхний индекс при v означает номер реализации, нижний — номер члена реализации.

Переход от начальных моментов к центральным производился по стандартным формулам, приводящимся, например, в [11].

Расчеты производились для нескольких значений параметра γ , лежащих в пределах от 0,6 до 2,0, поскольку именно в этом диапазоне изменяется значение γ для данных наблюдений о скорости ветра [7]. Что касается параметра β , то его изменение не имеет существенного значения для основных характеристик распределения, меняя лишь линейный масштаб оценок. Поэтому во всех случаях моделировались ряды, соответствующие значениям $\beta=1$.

3. При оценке различных характеристик методом статистического моделирования необходимо иметь в виду, что эти оценки являются выборочными, так что точность их зависит от объема использованной выборки. Некоторое представление о точности расчета одномерных моментов при использованных нами объемах выборки дает табл. 3 из работы [5], из которой следует, что погрешности их расчета лежат в пределах нескольких процентов от оцениваемых величин. Применительно к ряду других характеристик одномерного распределения систематические оценки погрешностей даны в [4].

Рассмотрение моментных функций, полученных путем статистической обработки моделированных выборок, показывает наличие в них как случайных ошибок, проявляющихся в разбросе отдельных значений на фоне общего хода функций, так и ошибок систематического характера, проявляющихся в общем смещении функции от выборки к выборке. Как случайные, так и систематические ошибки оказываются наиболее существенными для малых значений γ , соответствующих большой изменчивости величины v . Для исключения больших ошибок случайного характера данные расчетов дополнительно сглаживались графически, однако систематические ошибки при этом, естественно, не исключались. Приводящиеся ниже данные о корреляционных функциях μ_{11} могут поэтому содержать ошибки до 0,02, μ_{21} — до 0,05, μ_{22} и μ_{31} — до 0,2.

Таблица 1

Зависимость от γ параметров формул (14), (17) и (18) для моделей А и Б

Параметр	γ			
	0,6	1,0	1,4	1,8
R	0,48	0,61	0,66	0,68
α_1	0,60	0,89	1,03	1,24
α_2	0,60	0,64	0,66	0,68
δ	1,22	1,10	1,06	1,04
ϵ_B	1,37	1,05	1,04	1,02

Особый интерес представляют оценки моментных функций для вейбулловских последовательностей, соответствующих модели А, поскольку именно эта модель используется в статье [4] для определения точности статистической экстраполяции связанных вейбулловских рядов. Моментные функции для значений параметров q , используемых в [4], приводятся нами в приложении 2. При их расчете для $q > 1$ использовалось по 20, а при $q \leq 1$ — по 50 реализаций длиной 1200 членов каждая.

Анализ этих таблиц показывает, что как корреляционная функция $\mu_{11}(\tau)$, так и функции более высоких порядков существенно зависят не только от параметра q , определяющего связность исходной равномерной последовательности (увеличение q приводит к уменьшению всех функций μ), но и от вейбулловского параметра γ (увеличение его приводит к увеличению μ_{11} и уменьшению остальных функций). При этом зависимость от γ особенно заметна при $\gamma < 1$, при $\gamma > 1$ она сравнительно слабо выражена.

Наиболее существенно поведение величины $r_1 = \mu_{11}(1)$, т. е. коэффициента корреляции для смежных членов последовательно-

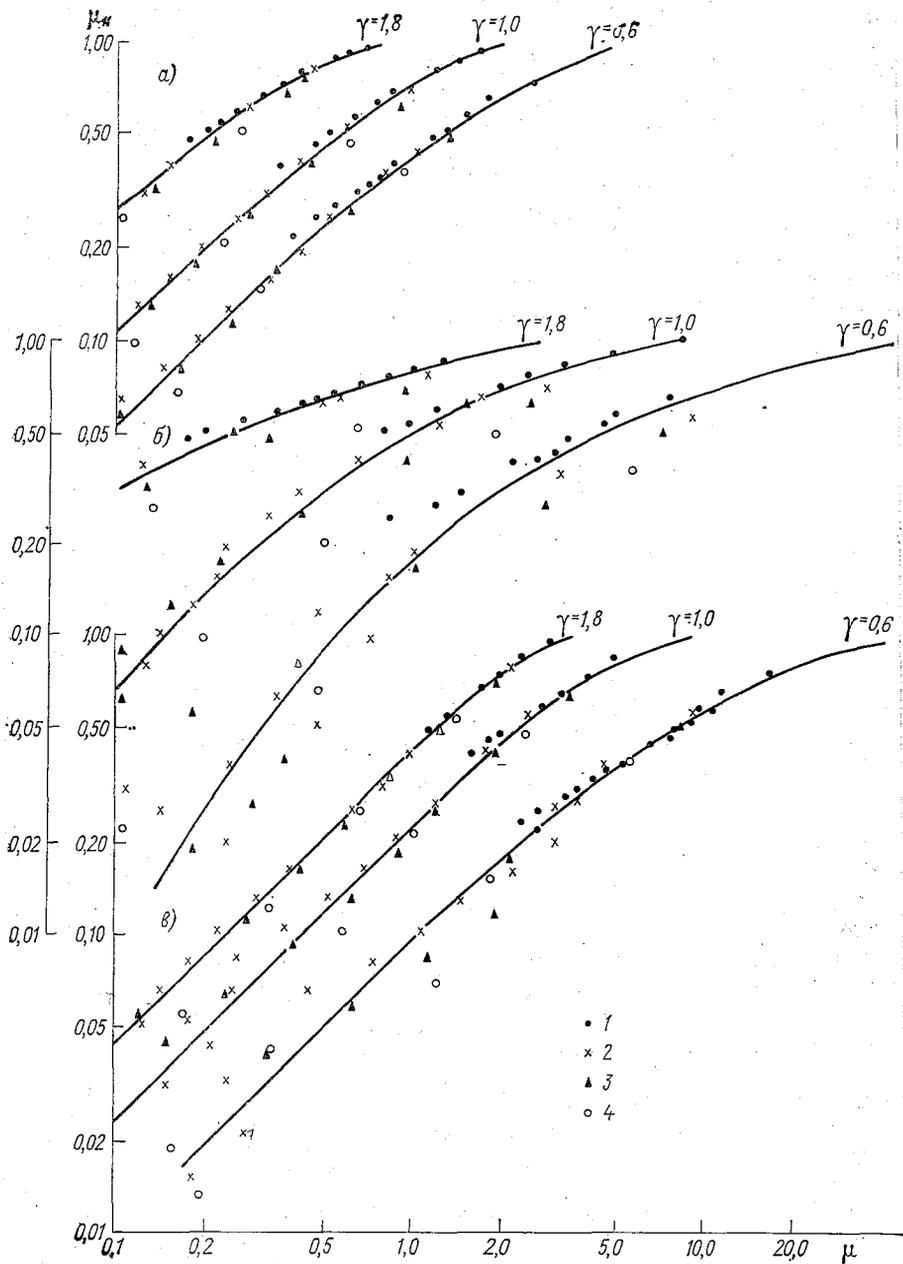


Рис. 1. Зависимость моментных функций μ_{21} (а), $\mu_{22} - 1$ (б), μ_{31} (в) от корреляционной функции μ_{11} для последовательностей, соответствующих модели А.

1) $q=0.25$; 2) $q=0.75$; 3) $q=1$; 4) $q=1.63$.

сти. На практике эта величина нередко оказывается единственной характеристикой связности данных фактических наблюдений.

Зависимость этой величины от параметра q хорошо описывается формулами

$$r_1 = \begin{cases} 1 - (1 - R)q^{\alpha_1} & \text{при } q \leq 1, \\ R/q^{\alpha_2} & \text{при } q > 1. \end{cases} \quad (14)$$

Здесь R представляет собой коэффициент корреляции при $q=1$.

Зависимость параметров R , α_1 и α_2 от γ представлена в табл. 1.

Из рис. 1, на котором представлена зависимость значений разных моментных функций от корреляционной функции μ_{11} , полученных для разных q , видно, что для фиксированных значений γ они практически укладываются на одну кривую. Разброс точек относительно этих кривых, по-видимому, связан с упоминавшимися выше выборочными погрешностями.

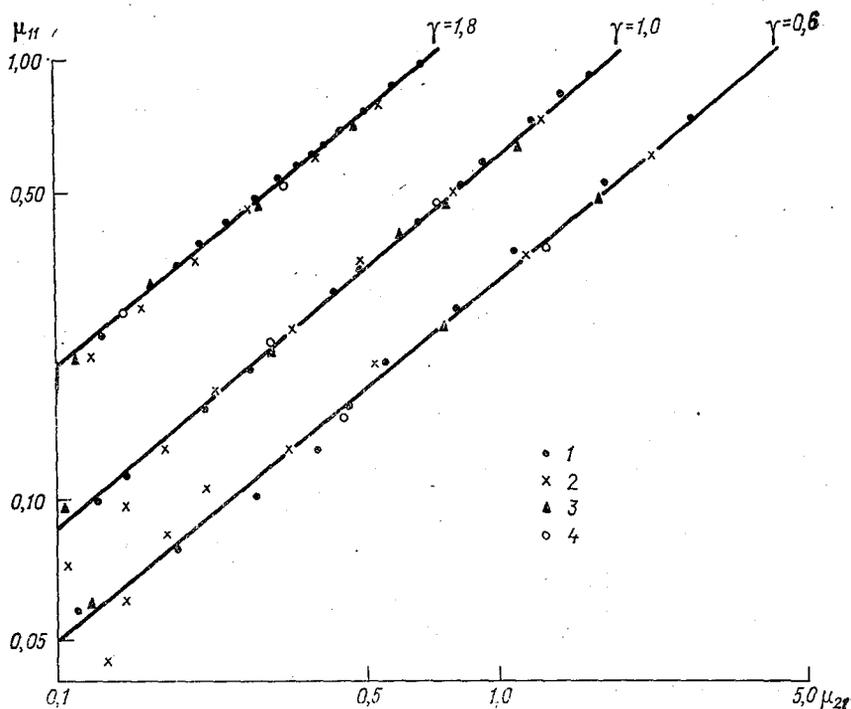


Рис. 2. Зависимость функции μ_{21} от корреляционной функции μ_{11} для последовательностей модели Б. Значения параметра p выбраны таким образом, чтобы коэффициенты корреляции r_1 совпадали с коэффициентами корреляции для модели А, соответствующими значениям q на рис. 1.

Уел. обозначения см. рис. 1.

Однозначная связь между моментами связи различных порядков имеет место и для вейбулловских последовательностей, моделированных в соответствии с другими алгоритмами. Так, из рис. 2, на котором в логарифмических координатах представлена зависимость между $\mu_{21}(\tau)$ и $\mu_{11}(\tau)$ для последовательностей, соответствующих модели Б, видно, что эта зависимость хорошо описывается степенным законом, причем показатель степени для различных значений γ практически один и тот же. В результате для этой модели имеем

$$\mu_{21}(\tau) = A(\mu_{11}(\tau))^{1,25}. \quad (15)$$

Степенная зависимость получается и для других моментов связи:

$$\mu_{22}(\tau) = 1 + (E + 2) [\mu_{31}(\tau)]^{1,9}, \quad (16)$$

$$\mu_{31}(\tau) = (E + 3) [\mu_{11}(\tau)]^{\delta}. \quad (17)$$

Однако показатель степени δ в (17) оказывается зависящим от γ . Значения его также представлены в табл. 1.

Связь коэффициента корреляции для смежных членов последовательностей, полученных по модели Б, с соответствующим коэффициентом корреляции p исходной гауссовской последовательности приближенно можно аппроксимировать степенной зависимостью

$$r_1 = p^{\varepsilon}. \quad (18)$$

Из табл. 1, в которой приведены значения ε для модели Б (ε_B , видно, что при $\gamma \geq 1$ коэффициенты корреляции r_1 и p приближенно равны. Лишь при $\gamma < 1$ различие между ними существенно.

В отличие от последовательностей, соответствующих модели А, модель Б генерирует последовательности, имеющие практически экспоненциально убывающую корреляционную функцию, для которой

$$\mu_{11}(\tau) = r_1^{\tau} = p^{\varepsilon \tau}. \quad (19)$$

Формулы (15)–(19) могут быть использованы для приближенной оценки моментов вейбулловской модели Б.

Для модели В А. С. Марченко [8] указана точная формула, связывающая величины p и r_1 . Она имеет вид

$$r_1 = \frac{\Gamma^2\left(1 + \frac{1}{\gamma}\right)}{\Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma^2\left(1 + \frac{1}{\gamma}\right)} \left[F\left(-\frac{1}{\gamma}; -\frac{1}{\gamma}; 1; p^2\right) - 1 \right], \quad (20)$$

где F — гипергеометрическая функция.

Практически использовать эту формулу неудобно. Как можно видеть из табл. 2, аппроксимация зависимости $r_1(p)$ степен-

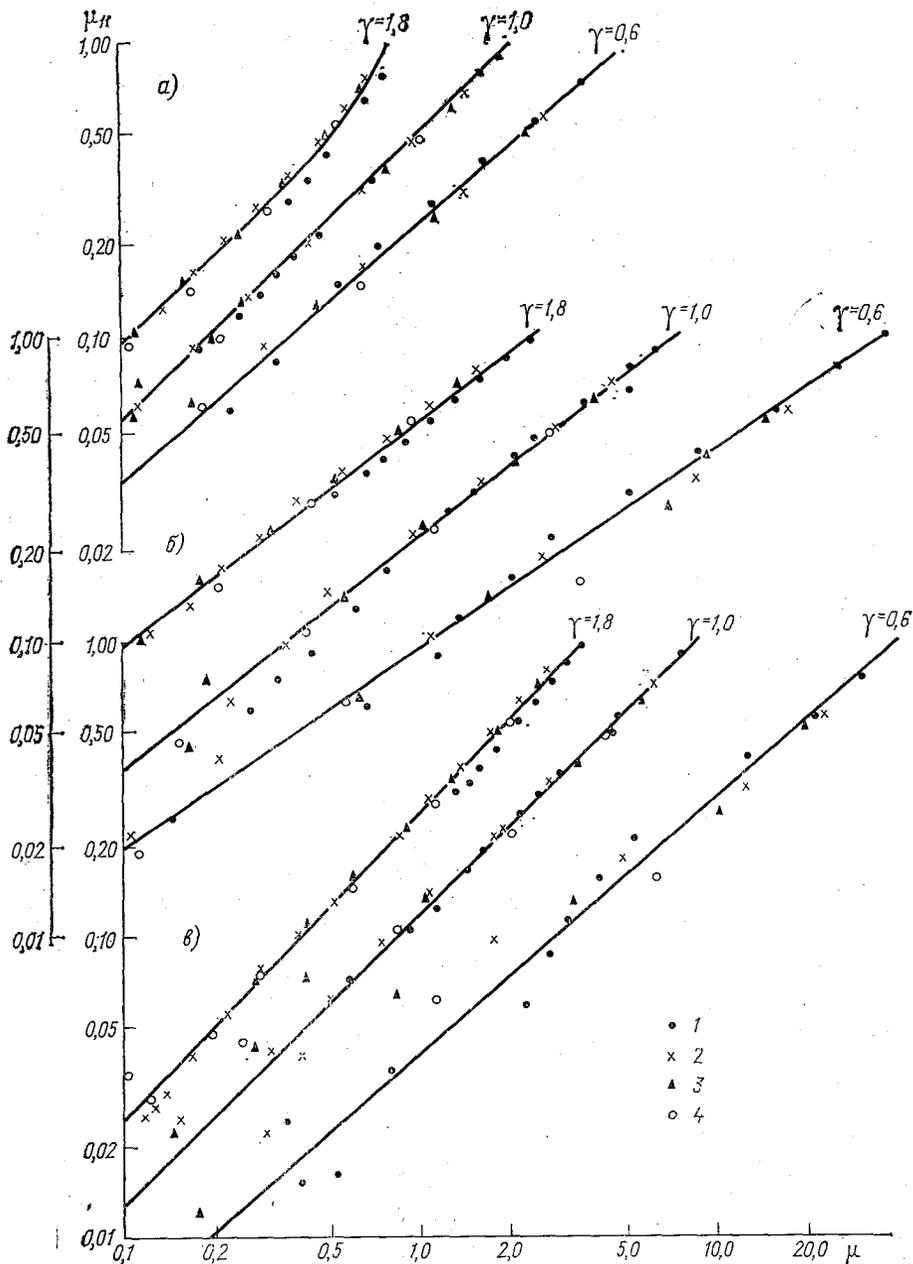


Рис. 3. Зависимость моментных функций μ_{21} (а), $\mu_{22} - 1$ (б) и μ_{31} (в) от корреляционной функции μ_{11} для последовательностей модели В. Значения параметра p выбраны таким образом, чтобы коэффициенты корреляции γ_1 совпадали с коэффициентами корреляции для модели А, соответствующими значениям q на рис. 1.

Усл. обозначения см. рис. 1.

ным законом (18), значения параметров которого ϵ_B приводятся в этой же таблице, приводит к ошибке, не превышающей 0,02.

Учитывая это обстоятельство и практически экспоненциальное убывание $\mu_{11}(\tau)$ со сдвигом во времени, формула (19) также может использоваться для описания временной корреляции последовательностей, соответствующих модели В, при надлежащем выборе параметра ϵ .

Зависимость моментов μ_{21} , μ_{31} и μ_{22} от корреляционной функции μ_{11} представлена на рис. 3. В сочетании с формулой (18) эта зависимость позволяет оценивать значения всех этих моментов при конкретном задании τ и γ .

Таблица 2

Точные (r_1) и приближенные (r_ϵ) значения коэффициентов корреляции смежных членов вейбулловской последовательности, соответствующей модели В

p	γ							
	0,6		1,0		1,4		1,8	
	r_1	r_ϵ	r_1	r_ϵ	r_1	r_ϵ	r_1	r_ϵ
0,9	0,794	0,797	0,810	0,810	0,805	0,807	0,795	0,800
0,8	0,616	0,619	0,640	0,640	0,633	0,634	0,620	0,623
0,7	0,465	0,464	0,490	0,490	0,483	0,483	0,470	0,469
0,6	0,337	0,317	0,360	0,360	0,353	0,353	0,343	0,339
0,5	0,231	0,225	0,250	0,250	0,245	0,243	0,237	0,230
0,4	0,146	0,139	0,160	0,160	0,156	0,154	0,151	0,143
0,3	0,082	0,075	0,090	0,090	0,088	0,086	0,084	0,078
0,2	0,036	0,031	0,040	0,040	0,039	0,038	0,037	0,033
0,1	0,009	0,007	0,010	0,010	0,010	0,009	0,009	0,004
ϵ_B	2,15		2,00		2,04		2,12	

Сравнение рис. 3 с рис. 1 для последовательностей модели А и с рис. 2 для последовательностей модели В показывает, что эти последовательности при одной и той же временной корреляционной функции имеют существенно различные моментные функции более высокого порядка. Для выяснения возможности использования рассмотренных моделей для статистического описания данных наблюдений за скоростью ветра необходимо получение этих функций на эмпирическом материале и сравнение их с расчетными.

ПРИЛОЖЕНИЕ I

Характеристики распределения Вейбулла для различных значений параметра γ

Характеристика	γ								
	0,6	0,8	1,0	1,2	1,4	1,6	1,8	2,0	2,2
\bar{v}/β	1,51	1,13	1,00	0,94	0,91	0,90	0,90	0,89	0,89
σ/β	2,64	1,43	1,00	0,79	0,66	0,57	0,51	0,46	0,42
c_v	1,76	1,26	1,00	0,84	0,72	0,64	0,58	0,52	0,48
A	4,59	2,82	2,00	1,52	1,20	0,96	0,80	0,63	0,51
E	37,48	12,74	6,00	3,24	1,84	1,04	0,56	0,24	0,04
\bar{v}^3/β^3	120,00	16,59	6,00	3,32	2,29	1,79	1,51	1,33	1,21
$v(0,5)/\beta$	0,54	0,63	0,69	0,74	0,77	0,80	0,82	0,83	0,85
$v(0,9)/\beta$	4,02	2,84	2,30	2,00	1,81	1,68	1,59	1,52	1,46
$v(0,95)/\beta$	6,23	3,94	3,00	2,50	2,19	1,98	1,84	1,73	1,65
$v(0,98)/\beta$	9,71	5,50	3,91	3,12	2,65	2,35	2,13	1,98	1,86
$v(0,99)/\beta$	12,75	6,75	4,60	3,57	3,00	2,60	2,34	2,15	2,02
$v(0,995)/\beta$	16,10	8,04	5,30	4,01	3,29	2,84	2,52	2,30	2,13
$v(0,999)/\beta$	25,06	11,20	6,91	5,01	3,98	3,35	2,93	2,63	2,41
$v(0,9995)/\beta$	29,38	12,62	7,60	5,42	4,26	3,55	3,09	2,76	2,51
$v(0,9999)/\beta$	40,47	16,05	9,21	6,36	4,88	4,01	3,43	3,04	2,74

Моментные функции (в сотых долях единицы) для вейбулловской последовательности, соответствующей модели А

q	μ	τ									
		1	2	3	4	6	8	12	16	24	
γ=0,6											
0,25	μ ₁₁	75	65	58	54	48	44	36	31	23	
	μ ₂₁	254	179	153	137	116	101	78	61	39	
	μ ₃₁	1716	1300	1050	935	795	685	512	390	220	
	μ ₂₂	1585	860	620	535	450	390	305	250	180	
0,60	μ ₁₁	60	43	33	27	20	14	8	4	1	
	μ ₂₁	168	96	68	56	38	26	14	7	0	
	μ ₃₁	1015	560	423	355	240	170	90	42	0	
	μ ₂₂	1050	450	305	245	195	172	140	115	100	
1,00	μ ₁₁	48	28	17	11	6	3	1	0	0	
	μ ₂₁	133	62	35	25	11	4	0	0	0	
	μ ₃₁	820	455	240	170	70	15	0	0	0	
	μ ₂₂	908	335	230	175	130	115	105	100	100	
1,63	μ ₁₁	37	15	7	2	1	0	0	0	0	
	μ ₂₁	92	31	17	8	2	0	0	0	0	
	μ ₃₁	525	240	145	80	4	0	0	0	0	
	μ ₂₂	652	250	150	101	100	100	100	100	100	
γ=1,0											
0,25	μ ₁₁	89	83	78	74	69	64	56	49	38	
	μ ₂₁	149	121	107	99	86	77	62	51	35	
	μ ₃₁	625	500	445	415	370	330	270	220	155	
	μ ₂₂	580	410	350	320	275	250	205	175	145	
0,60	μ ₁₁	76	61	50	42	32	23	13	7	2	
	μ ₂₁	106	71	53	43	30	22	11	6	0	
	μ ₃₁	433	295	231	200	140	100	50	25	0	
	μ ₂₂	408	250	190	162	138	132	115	108	100	
1,00	μ ₁₁	62	40	27	18	9	4	1	0	0	
	μ ₂₁	84	45	28	18	8	4	0	0	0	
	μ ₃₁	347	193	125	90	40	15	0	0	0	
	μ ₂₂	350	195	140	125	110	100	100	100	100	
1,63	μ ₁₁	46	21	10	4	0	0	0	0	0	
	μ ₂₁	60	23	11	5	1	0	0	0	0	
	μ ₃₁	247	100	60	35	5	0	0	0	0	
	μ ₂₂	294	150	120	100	100	100	100	100	100	

q	μ	τ								
		1	2	3	4	6	8	12	16	24
$\gamma=1,4$										
0,25	μ_{11}	93	88	84	81	76	71	63	56	44
	μ_{21}	96	82	73	68	60	53	44	37	26
	μ_{31}	388	335	300	290	260	237	200	170	130
	μ_{22}	365	300	260	240	215	195	165	145	125
0,60	μ_{11}	80	67	57	49	36	27	15	8	2
	μ_{21}	70	49	37	30	21	15	8	5	0
	μ_{31}	290	217	175	150	110	80	41	21	0
	μ_{22}	271	189	152	134	120	115	106	103	100
1,00	μ_{11}	66	45	30	21	10	5	1	0	0
	μ_{21}	56	31	20	13	6	3	0	0	0
	μ_{31}	235	145	95	70	30	13	1	0	0
	μ_{22}	230	150	120	110	105	103	100	100	100
1,63	μ_{11}	49	24	11	5	0	0	0	0	0
	μ_{21}	41	16	8	4	1	0	0	0	0
	μ_{31}	172	80	40	20	2	0	0	0	0
	μ_{22}	200	125	110	104	100	100	100	100	100
$\gamma=1,8$										
0,25	μ_{11}	94	90	86	84	78	74	66	59	47
	μ_{21}	63	54	49	45	40	36	30	25	18
	μ_{31}	305	270	250	240	220	202	175	150	115
	μ_{22}	290	250	225	210	190	173	150	135	117
0,60	μ_{11}	82	70	60	51	39	29	16	9	2
	μ_{21}	46	33	25	20	14	10	5	3	0
	μ_{31}	238	185	153	132	98	73	38	19	1
	μ_{22}	223	166	138	123	112	108	103	101	100
1,00	μ_{11}	68	47	32	22	11	5	1	0	0
	ω_{21}	38	21	13	9	4	2	0	0	0
	μ_{31}	195	125	85	60	28	12	1	0	0
	μ_{22}	190	132	110	105	102	101	100	100	100
1,63	μ_{11}	51	25	12	5	0	0	0	0	0
	μ_{21}	28	11	5	2	1	0	0	0	0
	η_{31}	145	67	35	17	5	1	0	0	0
	μ_{22}	165	113	105	101	100	100	100	100	100

СПИСОК ЛИТЕРАТУРЫ

1. Анапольская Л. Е. Режим скорости ветра на территории СССР. — Л.: Гидрометеоздат, 1961. — 200 с.
2. Быков В. В. Цифровое моделирование в статистической радиотехнике. — М.: Советское радио, 1971. — 326 с.
3. Голенко Д. И. Моделирование и статистический анализ псевдослучайных чисел на электронных вычислительных машинах. — М.: Наука, 1965. — 227 с.
4. Каган Р. Л. О точности экстраполяции статистических характеристик weibullовских последовательностей по выборочным данным. — См. наст. сб.
5. Каган Р. Л., Сибир Е. Е., Хлебникова Е. И. О моделировании связанных случайных последовательностей с равномерным распределением. — См. наст. сб.
6. Каган Р. Л., Федорченко Е. И. О применении статистического моделирования при решении метеорологических задач. — В кн.: Применение статистических методов в метеорологии. Тр. II Всесоюз. симпозиума по применению статистических методов в метеорологии. Л., 11—13 ноября 1975 г. — Л.: Гидрометеоздат, 1977, с. 36—44.
7. Кошинский С. Д. Режимные характеристики сильных ветров на морях Советского Союза. Ч. 1. Каспийское море. — Л.: Гидрометеоздат, 1975. — 412 с.; Ч. 2. Север Японского, Охотское и Берингово моря. — Л.: Гидрометеоздат, 1978. — 391 с.
8. Марченко А. С. Аппроксимация двумерных распределений зависимых, неотрицательных и неограниченных сверху случайных величин. — В кн.: Статистические методы в метеорологии. Ч. 1. — Новосибирск, ВЦ СОАН, 1969, с. 156—178.
9. Марченко А. С., Романенко Т. П. Аппроксимация двумерных распределений модуля скорости ветра. — В кн.: Применение статистических методов в метеорологии. — Новосибирск, ВЦ СОАН, 1971, с. 30—48.
10. Марченко А. С., Романенко Т. П. Моделирование гамма-последовательностей и их использование для изучения выбросов скорости ветра. — Метеорология и гидрология, 1975, № 7, с. 54—62.
11. Митропольский А. К. Техника статистических вычислений. — М.: Наука, 1971. — 576 с.

О ТОЧНОСТИ ЭКСТРАПОЛЯЦИИ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК ВЕЙБУЛЛОВСКИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ПО ВЫБОРОЧНЫМ ДАННЫМ

1. Для аппроксимации данных наблюдений за ветром, особенно для статистической экстраполяции вероятных скоростей ветра малой обеспеченности, часто используется распределение Вейбулла [2, 7]. Точность такой экстраполяции оценивалась рядом авторов (см., например, [4, 7]), однако выводы при этом зачастую оказывались противоречивыми. Одной из причин этого является, по-видимому, трудность разделения погрешностей, связанных с ограниченностью объема выборки, и погрешностей, определяющихся особенностями используемых при наблюдениях за ветром приборов, ошибками наблюдений, а также возможной неадекватностью принятой модели распределения. С этой точки зрения представляется полезным использование метода статистического моделирования, который позволяет генерировать на ЭВМ именно вейбулловские ряды и исследовать отдельно возможную точность определения их характеристик. Получение оценок такого рода и является целью настоящей статьи.

2. Двухпараметрическому закону Вейбулла некоторой величины v соответствует функция распределения

$$P(v \leq V) = 1 - F(V) = 1 - \exp\left(-\left(\frac{V}{\beta}\right)^\gamma\right) \quad (v > 0), \quad (1)$$

где γ и β представляют собой параметры, характеризующие форму и масштаб распределения.

Как это часто делается в метеорологии, мы будем далее пользоваться не функцией распределения P , а ее дополнением до единицы F , которое представляет собой вероятность превышения величиной v значения V .

Из формулы (1) получаются простые выражения для квантилей распределения Вейбулла. Мы их будем представлять в виде

$$V(F) = \beta(-\ln F)^{1/\gamma}. \quad (2)$$

Для статистической экстраполяции квантилей распределения в область редко встречающихся значений v удобно повторно прологарифмировать формулу (2), после чего мы имеем дело с линейной зависимостью

$$y = gx + b. \quad (3)$$

Здесь приняты следующие обозначения:

$$\begin{aligned} x &= \ln(-\ln F), & y &= \ln V(F), \\ g &= 1/\gamma, & b &= \ln \beta. \end{aligned} \quad (4)$$

Использование специальных вероятностных клетчаток, построенных в координатах $(\ln(-\ln F), \ln V)$, очень упрощает процедуру аппроксимации данных наблюдений законом Вейбулла, поскольку сводит ее к проведению прямой по нанесенным на клетчатку точкам, соответствующим эмпирическим данным. Более того, статистическая экстраполяция в этом случае может производиться графически, минуя этап определения параметров β и γ . Именно такая графическая процедура вероятностной оценки скорости ветра редкой повторяемости, описанная, например, в [2], использовалась при подготовке Справочника по климату СССР [12].

Во избежание субъективизма, возникающего при проведении прямой на клетчатке вручную, С. Д. Кошинский [7] предложил ручную процедуру аппроксимации заменить автоматизированным расчетом на ЭВМ с использованием метода наименьших квадратов. При этом по эмпирическим данным определяются величины x и y , после чего оценки параметров β и γ получаются по формулам:

$$\begin{aligned} g &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \\ b &= \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \\ \gamma &= 1/g, & \beta &= \exp(b), \end{aligned} \quad (5)$$

где n — означает объем выборки, x_i и y_i — эмпирические значения величин x и y . После определения β и γ легко определяются все интересующие нас характеристики, в том числе квантили распределения (2).

При расчетах по формулам (5) y_i определяется фактическими значениями наблюдаемых величин v_i . Что касается x_i , то для их определения, строго говоря, требуется знание вероятностей F_i , которые в действительности неизвестны. Эта трудность, которая

оказывается завуалированной при использовании сгруппированных данных, становится очевидной при использовании для аппроксимации индивидуальных значений исследуемой величины.

В этом случае данные наблюдений ранжируются, т. е. располагаются в порядке возрастания, и каждому наблюдению приписывается эмпирическая вероятность в зависимости от номера его i в ранжированной выборке. Однако для определения эмпирической вероятности по рангу i существуют различные формулы. Так, в гидрологической литературе (см., например, [1, 9]) широко обсуждались достоинства и недостатки формул

$$P_i = \frac{i - 0,5}{n}, \quad (6)$$

$$P_i = \frac{i}{n + 1}, \quad (7)$$

$$P_i = \frac{i - 0,3}{n + 0,4}, \quad (8)$$

первая из которых предложена А. Хазеном, вторая — С. Н. Крицким и М. Ф. Менкелем, а третья — Н. Н. Чегодаевым [11].

Как указал Е. Г. Блохинов [3], в зависимости от решаемой задачи может оказаться целесообразным для определения эмпирических вероятностей использовать разные формулы. Более того, наряду с формулами (6)—(8) могут быть использованы и другие формулы. Они могут быть представлены в общем виде

$$P_i = \frac{i - c}{n + 1 - 2c}. \quad (9)$$

При c , равном 0,5, 0 и 0,3, из (9) получаются соответственно формулы (6), (7) и (8).

Использование сгруппированных данных в соответствии со стандартной методикой [2], по-видимому, эквивалентно применению формулы Крицкого — Менкеля (7). Она официально рекомендуется и для использования при гидрологических расчетах [13]. Представляется, однако, целесообразным выяснить, насколько оправдано такое предпочтение применительно к рассматриваемой нами частной задаче статистической экстраполяции квантилей малой повторяемости для рядов, подчиняющихся распределению Вейбулла. Выбор наиболее подходящей для этой задачи формулы позволит наилучшим образом использовать метод наименьших квадратов.

Однако можно думать, что еще лучшие результаты должно дать использование метода максимального правдоподобия (см., например, [8]). В этом случае, вообще, отпадает необходимость в задании эмпирических вероятностей и соответственно в ранжировке данных, а оценки параметров β и γ могут быть получены

путем решения уравнений максимального правдоподобия, представленных в виде

$$Q = \frac{n}{\gamma} + \sum_{i=1}^n \ln v_i - \frac{n}{\sum_{i=1}^n v_i^\gamma} \sum_{i=1}^n v_i^\gamma \ln v_i = 0, \quad (10)$$

$$\beta = \left(\frac{1}{n} \sum_{i=1}^n v_i^\gamma \right)^{1/\gamma}. \quad (11)$$

Трансцендентное уравнение (10) легко решается методом последовательных приближений. Так, обозначая

$$Q_1 = \frac{\partial Q}{\partial \gamma} = -\frac{n}{\gamma^2} - \frac{n \left[\sum_{i=1}^n v_i^\gamma \sum_{i=1}^n v_i^\gamma \ln v_i - \left(\sum_{i=1}^n v_i^\gamma \ln v_i \right)^2 \right]}{\left(\sum_{i=1}^n v_i^\gamma \right)^2}, \quad (12)$$

можно, задавая некоторое предварительное значение γ_0 параметра γ , последовательно уточнять его в соответствии с формулой метода Ньютона

$$\gamma_{i+1} = \gamma_i - Q_i/Q_{1i} \quad (13)$$

до тех пор, пока невязка Q_i/Q_{1i} остается по абсолютной величине больше некоторого предписанного значения ε .

После определения γ параметр β определяется формулой (11).

3. Для оценки выборочных ошибок характеристик вейбулловских рядов были выполнены численные эксперименты методом статистического моделирования. В ходе каждого эксперимента моделировались выборки большого объема N из m независимых реализаций, длина каждой из которых составляла $n = N/m$ членов. При моделировании вейбулловских последовательностей использовался алгоритм преобразования связанных равномерно распределенных последовательностей (подробно рассмотренный в [5, 6] алгоритм А). Для заданных вейбулловских параметров β и γ предварительно моделированные равномерно распределенные величины z_i преобразовывались в вейбулловские по формуле

$$v_i = \beta(-\ln(1 - z_i))^{1/\gamma}. \quad (14)$$

Далее производилась статистическая обработка, в ходе которой каждая реализация рассматривалась как выборка объема n из генеральной вейбулловской совокупности с неизвестными параметрами β и γ , подлежащими определению в соответствии с описанными в п. 2 процедурами. В результате для каждой реализации номера j определялись выборочные значения β_j и γ_j , по которым вычислялись значения ряда статистических характери-

стик. Всего оценивалось более 20 характеристик, в числе которых рассчитывались:

среднее выборочное для реализации

$$\bar{v}_j = \beta_j \Gamma\left(1 + \frac{1}{\gamma_j}\right), \quad (15)$$

где Γ — означает гамма-функцию;
среднее квадратическое отклонение

$$\sigma_j = \beta_j \left[\Gamma\left(1 + \frac{2}{\gamma_j}\right) - \Gamma^2\left(1 + \frac{1}{\gamma_j}\right) \right]^{1/2}; \quad (16)$$

коэффициент изменчивости

$$c_{v_j} = \frac{\sigma_j}{\bar{v}_j} = \left[\frac{\Gamma(1 + 2/\gamma_j)}{\Gamma^2(1 + 1/\gamma_j)} - 1 \right]^{1/2}; \quad (17)$$

коэффициент асимметрии

$$A_j = \frac{\beta_j^3}{\sigma_j^3} \left[\Gamma\left(1 + \frac{3}{\gamma_j}\right) - 3 \Gamma\left(1 + \frac{2}{\gamma_j}\right) \Gamma\left(1 + \frac{1}{\gamma_j}\right) + 2 \Gamma^3\left(1 + \frac{1}{\gamma_j}\right) \right]; \quad (18)$$

коэффициент эксцесса

$$E_j = \frac{\beta_j^4}{\sigma_j^4} \left[\Gamma\left(1 + \frac{4}{\gamma_j}\right) - 4 \Gamma\left(1 + \frac{3}{\gamma_j}\right) \Gamma\left(1 + \frac{1}{\gamma_j}\right) + \right. \\ \left. + 6 \Gamma\left(1 + \frac{2}{\gamma_j}\right) \Gamma^2\left(1 + \frac{1}{\gamma_j}\right) - 3 \Gamma^4\left(1 + \frac{1}{\gamma_j}\right) \right] - 3; \quad (19)$$

средний куб значений v на реализации (третий начальный момент)

$$(\bar{v}^3)_j = \beta_j^3 \Gamma\left(1 + \frac{3}{\gamma_j}\right); \quad (20)$$

квантили распределения, соответствующие ряду заданных значений вероятности превышения F ,

$$V_j(F) = \beta_j (-\ln(F))^{1/\gamma_j}. \quad (21)$$

После получения указанных характеристик для всех реализаций по ним получались средние значения для всей выборки и среднее квадратическое отклонение. Например, для оценки средних значений, получающихся из m реализаций длины n , имеем

$$\bar{v} = \frac{1}{m} \sum_{j=1}^m \bar{v}_j, \quad (22)$$

$$\sigma_v = \left(\frac{1}{m-1} \sum_{j=1}^m (\bar{v}_j - \bar{v})^2 \right)^{1/2}.$$

Для каждой выборки при заданной длине реализации средние выборочные и стандарты характеристик f (\bar{f} и σ_f) сравнивались с истинными значениями f_0 этих характеристик, соответствующими задававшимся при моделировании параметрам β и γ .

Мы будем далее приводить не абсолютные значения \bar{f} и σ_f , а их отношения к теоретическим значениям f_0 , которые позволяют лучше судить о смещении и о рассеянии выборочных оценок. Основная часть расчетов выполнялась для выборок объема

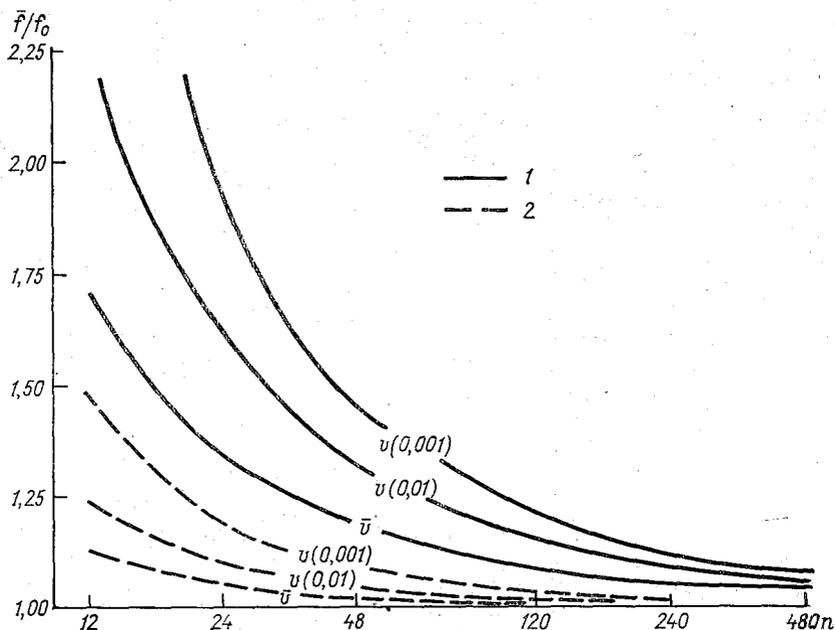


Рис. 1. Смещение характеристик распределения бесвязной вейбулловской последовательности, определенных методом наименьших квадратов. $\gamma=0,6$.

1) $c=0$; 2) $c=0,5$.

120 000 членов. Длина реализаций варьировалась от $n=12$ до $n=4800$. Моделирование осуществлялось для четырех значений вейбулловского параметра γ (0,6; 1,0; 1,4 и 1,8) охватывающих большую часть диапазона изменений этого параметра применительно к данным наблюдений за скоростью ветра [7]. Поскольку параметр β характеризует лишь масштаб размерных характеристик, а интересующие нас относительные величины от него не зависят, он в ходе расчетов не варьировался, а задавался равным 1.

При расчетах методом максимального правдоподобия в качестве начального приближения задавалось $\gamma_0=\gamma$, что позволи-

Таблица 1

Выборочные значения (в процентах от теоретических значений) характеристик распределения при различном определении эмпирических вероятностей и при использовании метода максимального правдоподобия (ММП) (бесвязные последовательности, $n=120$)

Характеристика f	Среднее выборочное \bar{f}					Среднее квадратическое отклонение σ_f				
	c				ММП	c				ММП
	0	0,3	0,5	0,7		0	0,3	0,5	0,7	
$\gamma=0,6$										
\bar{v}	109	104	101	96	100	21	20	19	18	15
σ	116	108	102	95	100	34	31	29	26	20
c_v	106	102	100	97	100	12	12	12	11	9
A	109	104	101	96	100	20	19	18	17	13
E	127	115	106	96	101	54	48	44	40	30
$\gamma=1,0$										
\bar{v}	103	101	100	98	100	10	10	10	10	9
σ	107	103	100	96	99	17	16	16	15	11
c_v	104	102	100	97	99	10	9	9	9	7
A	106	103	100	96	99	15	14	14	14	11
E	107	109	102	94	100	37	35	33	31	24
$\gamma=1,8$										
\bar{v}	101	100	100	99	100	5	5	5	5	5
σ	104	102	100	97	99	10	10	10	9	7
c_v	104	101	100	98	99	9	8	8	8	6
A	107	103	99	95	99	18	18	18	17	13
E	131	115	104	89	99	66	62	59	56	44

ло заметно сократить число итераций по формуле (13), особенно при больших n . Выборочные значения этого параметра определялись с точностью до третьего знака после запятой ($\varepsilon = 0,001$).

4. И систематические и случайные выборочные ошибки существенно зависят от значений параметра γ , особенно велики они при $\gamma < 1$. Они могут достигать очень больших значений, особенно при малых выборках (несколько десятков членов), уменьшаясь с увеличением объема выборки n .

В качестве примера на рис. 1 представлена зависимость от объема выборки систематических ошибок определения методом наименьших квадратов характеристик \bar{v} , $V(0,01)$ и $V(0,001)$ для $\gamma=0,6$ при задании эмпирических вероятностей формулами (6)

Таблица 2

Выборочные оценки (в процентах от теоретических значений) квантилей распределения при различном определении эмпирических вероятностей и при использовании метода максимального правдоподобия (ММП) (бесвязные последовательности, $n=120$)

Вероятность F	Среднее выборочное \bar{f}					Среднее квадратическое отклонение σ_f				
	c				ММП	c				ММП
	0	0,3	0,5	0,7		0	0,3	0,5	0,7	
$\gamma=0,6$										
0,1	108	104	100	96	99	21	20	19	18	16
0,01	115	108	102	95	100	33	30	28	26	20
0,001	121	111	103	94	100	42	38	35	32	24
0,0001	125	113	105	94	100	49	44	41	36	26
$\gamma=1,0$										
0,1	104	102	100	97	99	12	12	11	11	9
0,01	108	104	100	96	99	18	17	16	16	12
0,001	110	105	101	95	99	22	21	20	19	14
0,0001	112	106	101	95	99	26	24	23	21	16
$\gamma=1,4$										
0,1	103	101	100	98	100	8	8	8	8	7
0,01	105	102	100	97	99	12	12	12	11	8
0,001	107	103	100	96	99	15	15	14	13	10
0,0001	108	104	100	96	99	17	17	16	15	11
$\gamma=1,8$										
0,1	102	101	100	98	100	6	6	6	6	5
0,01	104	102	100	97	99	9	9	9	9	7
0,001	105	102	100	97	99	12	11	11	10	8
0,0001	106	103	100	97	99	13	13	12	12	9

и (7). Из рисунка видно, что выборочные значения этих характеристик являются завышенными, особенно при $c=0$, т. е. при использовании формулы Крицкого—Менкеля (7). С увеличением n смещение выборочных характеристик уменьшается, однако при неудачном выборе c оно остается заметным и при сравнительно больших объемах выборки.

Это хорошо видно и из табл. 1 и 2, в которых представлена зависимость ряда выборочных характеристик распределения бесвязных вейбулловских последовательностей от задания парамет-

за c в формуле (9) для определения эмпирических вероятностей при длине реализации $n=120$.

Наименьшее смещение практически всех рассмотренных характеристик достигается при значениях c , близких к 0,5, наибольшее — при $c=0$. Существенно также, что с увеличением c уменьшается и среднее квадратическое отклонение выборочных характеристик. Величина σ_f при $c=0,5$ не достигает минимума, уменьшаясь далее с увеличением c . Следует однако иметь в виду, что σ_f характеризует лишь рассеяние выборочных оценок от-

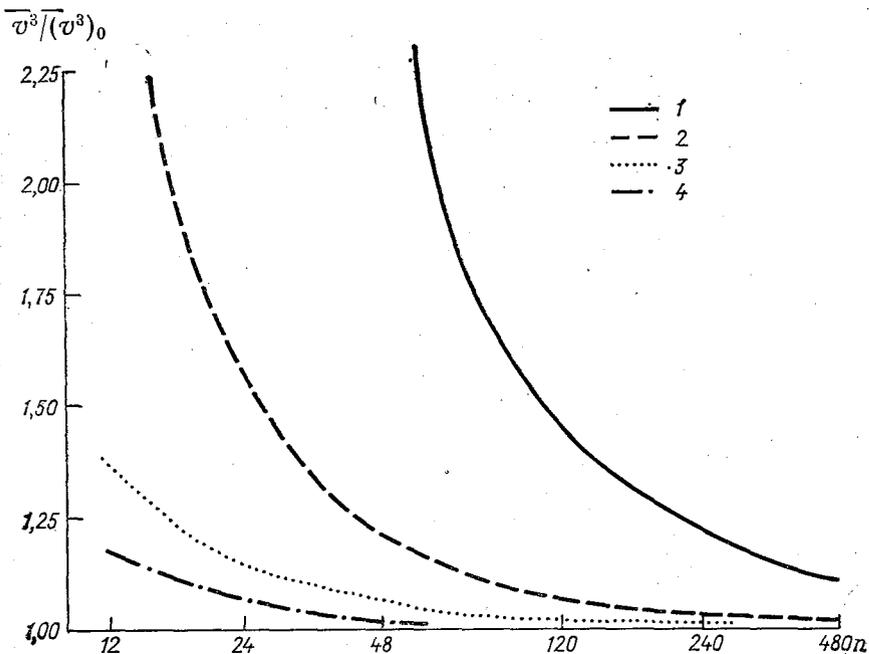


Рис. 2. Смещение выборочных значений \bar{v}^3 для бесвязной вейбулловской последовательности, полученных методом наименьших квадратов при $c=0,5$.

1) $\gamma=0,6$; 2) $\gamma=1,0$; 3) $\gamma=1,4$; 4) $\gamma=1,8$.

носителю f , а не ошибку оценки истинного значения f_0 , дисперсия которой при наличии смещения величины f составляет $\sigma_f^2 + (f-f_0)^2$. С учетом этого случайные выборочные ошибки определения f также оказываются минимальными при $c=0,5$; т. е. при использовании формулы Хазена (6). В связи с этим дальнейшие оценки, относящиеся к методу наименьших квадратов, приводятся нами именно для этого варианта.

В этом случае при $n > 100$ смещением выборочных характеристик практически можно пренебречь. Исключение составляют некоторые существенно нелинейные характеристики, смещение которых оказывается заметно больше. Характерно в этом смысле

поведение выборочных значений \bar{v}^3 , которые для скорости ветра представляют интерес как параметр, определяющий потенциальные ветроэнергоресурсы. Из рис. 2, на котором представлена зависимость $\bar{v}^3/(\bar{v}^3)_0$ для различных объемов выборки, видно, что эта величина при $\gamma < 1$ оказывается существенно завышенной при объемах выборки, достигающих 200—300 членов. При больших n и при $\gamma \geq 1$ смещением этой характеристики также можно пренебречь.

Таблица 3

Средние квадратические отклонения (в процентах от теоретических значений) выборочных характеристик распределения, определенных методом наименьших квадратов ($c=0,5$) для бессвязных рядов при различном объеме выборки

f	n					n				
	12	24	48	120	480	12	24	48	120	480
	$\gamma=0,6$					$\gamma=1,0$				
\bar{v}	94	48	31	19	10	32	23	16	10	5
σ	415	90	50	29	14	59	37	25	16	8
c_v	41	27	19	12	6	30	21	15	9	5
A	81	45	30	18	9	48	32	22	14	7
E	2172	199	90	44	20	200	89	56	33	16
β	55	37	26	16	8	31	22	15	9	5
γ	34	22	15	9	5	34	22	15	9	5
\bar{v}^3	—	—	—	180	24	—	344	105	50	23
	$\gamma=1,4$					$\gamma=1,8$				
\bar{v}	22	15	11	7	4	17	12	8	5	3
σ	37	26	18	12	6	30	22	15	10	5
c_v	27	19	14	9	4	26	18	13	8	4
A	47	33	24	15	7	55	39	28	18	9
E	160	93	62	38	19	213	143	97	59	29
β	22	15	11	7	3	17	12	8	5	3
γ	34	22	15	9	5	34	22	15	9	5
\bar{v}^3	401	86	48	28	14	94	48	31	19	10

Оценки, полученные методом максимального правдоподобия, оказываются практически несмещенными, что хорошо видно из табл. 1 и 2. Случайные выборочные погрешности σ_f этих оценок заметно меньше ошибок оценок, полученных методом наименьших квадратов. Особенно это относится к оценкам коэффициентов асимметрии и эксцесса и квантилей малой повторяемости, для которых выборочные оценки методом максимального правдоподобия оказываются примерно в 1,5 раза точнее.

В табл. 3 представлена зависимость от объема выборки случайных ошибок оценок, полученных методом наименьших квадратов. Аналогичные результаты для метода максимального правдоподобия представлены в табл. 4.

Из этих таблиц видно, что значения σ_f при малых объемах выборки убывают довольно быстро с ростом n , а при $n > 100$ практически выполняется «закон корня из n », так что для больших объемов выборки можно пользоваться приближенными оценками

$$\sigma_f(n) = \sigma_f(120) \sqrt{\frac{120}{n}}. \quad (23)$$

Исключение составляют оценки $\overline{v^3}$, для которых при $\gamma < 1$ «закон корня из n » начинает выполняться лишь при $n > 300$.

Аналогичная зависимость от объема выборки имеет место и для случайных ошибок квантилей распределения. Значения их,

Таблица 4

Средние квадратические отклонения (в процентах от теоретических значений) выборочных характеристик распределения, определенных методом максимального правдоподобия для бесвязных рядов при различном объеме выборки

f	n					n				
	24	48	120	240	600	24	48	120	240	600
	$\gamma=0,6$					$\gamma=1,0$				
\overline{v}	36	25	15	11	7	20	14	9	6	4
σ	47	32	20	14	10	25	18	11	8	5
c_v	20	14	9	6	4	16	12	7	5	3
A	31	21	13	9	6	24	17	11	7	5
E	85	52	30	21	13	57	39	24	17	10
β	37	25	15	10	7	21	15	9	6	4
γ	18	12	7	5	3	18	12	7	5	3
$\overline{v^3}$	577	185	80	50	29	92	57	33	23	15
	$\gamma=1,4$					$\gamma=1,8$				
\overline{v}	15	10	6	4	3	12	8	5	3	2
σ	19	13	8	6	4	16	11	7	5	3
c_v	15	11	7	5	3	14	10	6	5	3
A	25	18	11	8	5	30	21	13	9	6
E	63	44	27	19	12	99	70	44	31	19
β	15	11	7	4	3	12	8	5	3	2
γ	18	12	7	5	3	18	12	7	5	3
$\overline{v^3}$	50	34	21	14	10	36	25	15	11	7

соответствующие различным значениям F для оценок, выполняемых методом наименьших квадратов, представлены в табл. 5. Относительные погрешности оказываются сравнительно малыми вблизи медианы распределения, при $F=0,3$ они минимальны. Для квантилей малой вероятности F относительные погрешности являются довольно большими. Заметно уменьшаются они при расчетах методом максимального правдоподобия, оценки случайных ошибок выборочных квантилей для которого приведены в табл. 6.

5. Приведенные выше оценки относились к бесвязным рядам. В реальном случае связанных рядов можно ожидать существенных отличий в этих оценках. Для выяснения влияния связности были выполнены аналогичные вышеописанным эксперименты. При

Таблица 5

Средние квадратические отклонения (в процентах от теоретических значений) квантилей распределения, определенных методом наименьших квадратов ($c=0,5$) для бесвязных рядов при различном объеме выборки

F	n					n				
	12	24	48	120	480	12	24	48	120	480
	$\gamma=0,6$					$\gamma=1,0$				
0,5	64	42	29	17	9	35	24	17	13	5
0,3	53	36	25	16	8	30	21	15	9	5
0,1	69	46	31	19	10	37	26	18	11	6
0,05	91	55	36	22	12	43	30	21	13	7
0,02	135	69	43	26	13	53	35	24	15	8
0,01	181	79	48	28	14	61	39	26	16	9
0,005	240	90	53	31	15	68	42	29	18	9
0,001	447	117	63	35	17	88	50	33	20	10
0,0005	570	129	67	37	18	96	53	34	21	11
0,0001	953	157	76	41	19	117	59	38	23	11
	$\gamma=1,4$					$\gamma=1,8$				
0,5	24	17	12	7	4	19	13	9	6	3
0,3	22	15	11	7	3	17	12	8	5	3
0,1	25	18	13	8	4	19	14	10	6	3
0,05	29	21	15	9	5	22	16	11	7	4
0,02	34	24	17	11	6	26	18	13	8	4
0,01	38	27	18	12	6	28	20	14	9	5
0,005	42	28	20	12	6	31	21	15	10	5
0,001	50	33	22	14	7	36	25	17	11	6
0,0005	53	34	23	15	8	38	26	18	11	6
0,0001	61	38	26	16	8	42	28	20	12	6

этом задавались такие значения определяющего связность ряда параметра q , которые, по-видимому, можно считать характерными для данных наблюдений за скоростью ветра. В самом деле из сравнения данных работы [5], в которой приводятся корреляционные функции вейбулловских рядов, моделированных в соответствии с нашим алгоритмом А, с данными о временной структуре скорости ветра, приведенными в [10], видно, что эти функ-

Таблица 6

Средние квадратические отклонения (в процентах от теоретических значений) квантилей распределения, определенных методом максимального правдоподобия для бессвязных рядов при различном объеме выборки

F	n					n				
	24	48	120	240	600	24	48	120	240	600
	$\gamma=0,6$					$\gamma=1,0$				
0,1	35	25	16	11	—	21	15	9	6	—
0,05	38	27	17	12	—	22	16	10	7	—
0,02	43	30	19	13	—	25	18	11	8	—
0,01	46	32	20	14	9	26	19	12	8	5
0,005	49	34	21	15	9	28	20	13	9	6
0,001	56	38	24	16	11	31	22	14	10	6
0,0005	59	40	25	17	11	33	23	15	10	7
0,0001	66	43	26	18	12	35	25	16	11	7
	$\gamma=1,4$					$\gamma=1,8$				
0,1	15	11	7	5	—	12	8	5	4	—
0,05	16	11	7	5	—	13	9	6	4	—
0,02	18	13	8	6	—	14	10	6	4	—
0,01	19	13	8	6	4	15	10	7	5	3
0,005	20	14	9	6	4	16	11	7	5	3
0,001	22	16	10	7	5	17	12	8	5	4
0,0005	23	16	10	7	5	18	13	8	6	4
0,0001	25	18	11	8	5	19	14	9	6	4

ции при $q=1$ приближенно соответствуют временной корреляции рядов наблюдений за скоростью ветра при четырех наблюдениях в сутки. Корреляция вейбулловских последовательностей, моделированных при $q=0,6$, близка к корреляционным функциям скорости ветра при восьмиразовых наблюдениях.

Для связных рядов смещение выборочных оценок оказывается гораздо большим, чем для бессвязных. В этом случае значительное смещение для выборок, объем которых меньше 100—200 чле-

нов, имеет место и при использовании метода максимального правдоподобия. В качестве примера на рис. 3 приведены оценки смещения квантили $v(0,001)$ для случая $\gamma=0,6$. При $n > 300$ при указанных значениях параметра q смещением выборочных характеристик можно пренебречь.

Случайные погрешности выборочных оценок представлены в табл. 7, из которой следует, что для связанных рядов эти погрешности существенно повышаются по сравнению с рассмотренным ранее случаем бессвязных рядов ($q = \infty$).

Метод максимального правдоподобия и для связанных рядов обеспечивает уменьшение случайных выборочных погрешностей по сравнению с методом наименьших квадратов, особенно для квантилей малой обеспеченности.

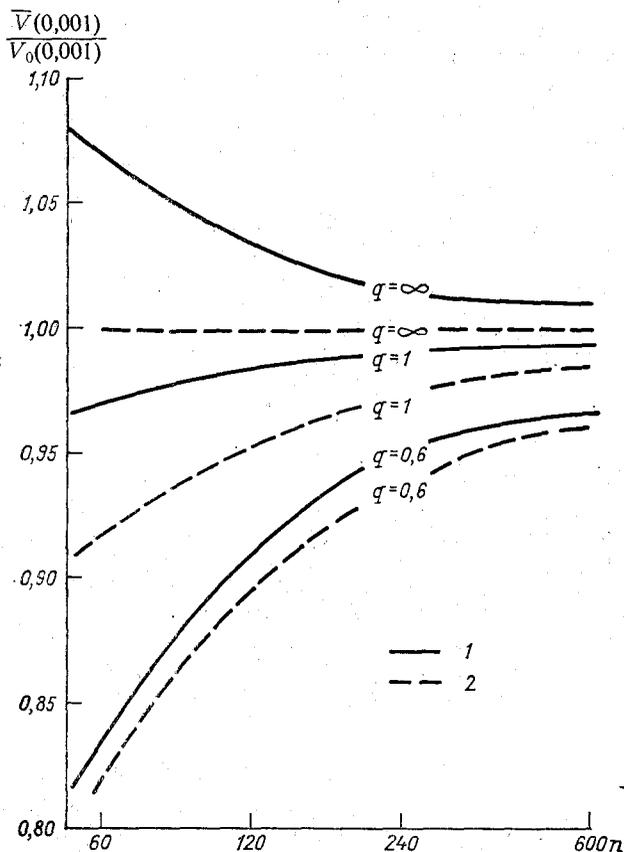


Рис. 3. Смещение выборочных значений квантили $V(0,001)$ вейбулловской последовательности, полученной методом наименьших квадратов при $s=0,5$ (1) и методом максимального правдоподобия (2). $\gamma=0,6$.

Таблица 7

Средние квадратические отклонения (в процентах от теоретических значений) выборочных характеристик связанных weibullовских последовательностей при $q = \infty$; 1; 0,6

f	Метод наименьших квадратов ($c=0,5$)						Метод максимального правдоподобия					
	n=120			n=240			n=120			n=240		
	∞	1	0,6	∞	1	0,6	∞	1	0,6	∞	1	0,6
$\gamma=0,6$												
\bar{v}	19	35	47	14	25	35	15	30	42	11	21	30
σ	29	43	50	21	31	38	20	31	38	14	23	28
A	18	25	29	13	18	21	13	21	28	9	15	20
E	44	65	76	30	43	49	30	50	73	21	35	46
V(0,001)	35	51	56	25	35	42	24	35	41	16	25	30
V(0,0001)	41	58	64	29	40	46	26	38	46	18	28	33
$\gamma=1,0$												
\bar{v}	10	20	30	7	15	22	9	19	28	6	13	20
σ	16	23	28	11	17	22	11	18	23	8	13	17
A	14	20	24	10	14	17	11	17	23	7	12	16
E	33	46	53	23	33	37	24	38	51	17	27	35
V(0,001)	20	28	32	14	21	25	14	21	25	10	15	18
V(0,0001)	23	31	35	16	23	27	16	22	27	11	16	20
$\gamma=1,8$												
\bar{v}	5	10	18	4	8	13	5	12	17	3	8	12
σ	10	14	17	7	11	13	7	11	14	5	8	10
A	17	25	31	12	18	22	13	21	14	9	13	20
E	59	82	93	42	60	68	43	69	29	31	50	64
V(0,001)	11	15	18	8	11	14	8	12	15	5	8	11
V(0,0001)	12	17	19	8	12	15	9	13	16	6	9	11

Увеличение объема выборки при заданной связности, т. е. за счет увеличения длины ряда, приводит к уменьшению погрешностей, которое при $n > 120$ происходит в соответствии с «законом корня из n». Однако увеличение объема выборки за счет учащения измерений, при котором связность ряда повышается, приводит к гораздо меньшему увеличению информации и соответственно к меньшему повышению точности выборочных характеристик. Это видно из сравнения точности оценок при $n=120$ и $q=1$ с точностью их при $n=240$ и $q=0,6$.

Приведенные выше результаты дают представление о точности, с которой могут быть получены оценки характеристик вейбулловских последовательностей по данным выборок ограниченного объема. Из них следует, что для расчета этих характеристик желательно использовать метод максимального правдоподобия. При использовании метода наименьших квадратов желательно для оценки эмпирических вероятностей применять формулу Хазена (6).

Указанные выводы относятся к случаю, когда обработке подвергаются индивидуальные данные наблюдений, а не сгруппированные данные, которые используются в климатологических обобщениях. Представляется, что эти выводы в качественном плане справедливы и для сгруппированных данных, однако соответствующие количественные оценки для этого случая должны быть уточнены, равно как требуют уточнения и способы применения методов наименьших квадратов и максимального правдоподобия для таких данных.

СПИСОК ЛИТЕРАТУРЫ

1. Алексеев Г. А. О формулах для вычисления обеспеченностей гидрологических величин. — Метеорология и гидрология, 1955, № 6, с. 40—43.
2. Анапольская Л. Е. Режим скоростей ветра на территории СССР. — Л.: Гидрометеиздат, 1961. — 200 с.
3. Блохинов Е. Г. О выборе формулы для оценки вероятной повторяемости характеристик речного стока. — В кн.: Сборник работ по гидрологии, № 2. — Л.: Гидрометеиздат, 1961, с. 115—121.
4. Заварина М. В. Расчетные скорости ветра на высотах нижнего слоя атмосферы. — Л.: Гидрометеиздат, 1971. — 163 с.
5. Каган Р. Л., Сибир Е. Е. О статистических характеристиках некоторых моделей связанных вейбулловских последовательностей. — См. наст. сб.
6. Каган Р. Л., Сибир Е. Е., Хлебникова Е. И. О моделировании связанных случайных последовательностей с равномерным распределением. — См. наст. сб.
7. Кошинский С. Д. Режимные характеристики сильных ветров на морях Советского Союза. Ч. 1. Каспийское море. — Л.: Гидрометеиздат, 1975. — 412 с.; Ч. 2. Север Японского, Охотское и Берингово моря. — Л.: Гидрометеиздат, 1978. — 391 с.
8. Крамер Г. Математические методы статистики. — М.: Мир, 1975. — 648 с.
9. Крицкий С. Н., Менкель М. Ф. К вопросу об оценке вероятной повторяемости гидрологических величин. — Метеорология и гидрология, 1957, № 3, с. 52—53.
10. Марченко А. С., Романенко Т. П. Моделирование гамма-последовательностей и их использование для изучения выбросов скорости ветра. — Метеорология и гидрология, 1975, № 7, с. 54—62.
11. Рождественский А. В., Чеботарев А. И. Статистические методы в гидрологии. — Л.: Гидрометеиздат, 1974. — 424 с.
12. Справочник по климату СССР. Ч. III. Ветер. — Л.: Гидрометеиздат, 1968.
13. Указания по определению расчетных гидрологических характеристик. СН 435—72. — Л.: Гидрометеиздат, 1972. — 19 с.

СОДЕРЖАНИЕ

Г. В. Груза, И. И. Поляк, В. А. Шахмейстер. К анализу статистической структуры широтно-временных полей OT_{1000}^{500} и зональной составляющей скорости геострофического ветра поверхности 500 мбар северного полушария	3
И. И. Поляк, О. А. Смирнова. Сглаживание коррелированных наблюдений	16
А. П. Соколов. Построение модели авторегрессии — скользящего среднего для стационарного временного ряда	21
В. А. Шахмейстер. Программа спектрального анализа однородного случайного поля	27
Б. Г. Вагер. Использование одномерных сплайнов при обработке экспериментальных данных	33
Б. Г. Вагер, Е. Д. Надежина. Об изменении верхней границы планетарного пограничного слоя	39
Т. В. Носова. К анализу корреляционной структуры сезонных временных рядов температуры	46
Б. Г. Вагер. Сравнение сплайновых и конечно-разностных решений дифференциальных уравнений на примерах простейших течений в пограничном слое	49
А. И. Полищук, Т. Н. Юринова. Архивация данных в автоматизированной системе обработки информации о загрязнении атмосферы (АСОИЗА)	(54)
Л. П. Афиногенов, Г. А. Абашев. Количество информации, передаваемой в системе длительного хранения, использующей алфавит с фиксированными ограничениями	67
Л. Г. Печискер. Организация диалога при решении задач океанологии	77
Л. Г. Печискер. Диалоговое взаимодействие в задачах океанологии	84
Р. Л. Каган, Е. Е. Сибир, Е. И. Хлебникова. О моделировании связанных случайных последовательностей с равномерным распределением	95
Р. Л. Каган, Е. Е. Сибир. О статистических характеристиках некоторых моделей связанных вейбулловских последовательностей	104
Р. Л. Каган. О точности экстраполяции статистических характеристик вейбулловских последовательностей по выборочным данным	119

Труды ГГО, вып. 437.

АНАЛИЗ ГИДРОМЕТЕОРОЛОГИЧЕСКИХ ПРОЦЕССОВ С ПОМОЩЬЮ ЭВМ

Н/К

Редактор В. И. Кузьменко. Техн. редактор Г. В. Ивкова. Корректор И. А. Динабург. Сдано в набор 26.01.81. Подписано в печать 08.06.81. М-21522. Формат 60×90^{1/16}. Бум. тип. № 1. Лит. гарн. Печать высокая. Печ. л. 9.0. Кр.-отт. 9,25. Уч.-изд. л. 9,27. Тираж 650 экз. Индекс МЛ-115. Заказ 136. Цена 65 коп. Гидрометеиздат. 199053. Ленинград, 2-я линия, д. 23.

Сортавальская книжная типография Управления по делам издательств, полиграфии и книжной торговли Совета Министров Карельской АССР. Сортавала, Карельская, 42.