

**В.С. Комаров, В.И. Акселевич, А.В. Креминский**

### **МОДИФИЦИРОВАННЫЙ МЕТОД ГРУППОВОГО УЧЕТА АРГУМЕНТОВ КАК ЭФФЕКТИВНЫЙ МЕТОД СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ХАРАКТЕРИСТИК СВОБОДНОЙ АТМОСФЕРЫ В УСЛОВИЯХ ИНФОРМАЦИОННОЙ НЕОПРЕДЕЛЕННОСТИ**

Обсуждаются перспектива и возможности применения модифицированного метода группового учета аргументов (МГУА) в задачах прогноза (восстановления) вертикальных профилей температуры, зональной и меридиональной составляющих скорости ветра, проводимого в свободной атмосфере (при отсутствии экспериментальной информации) по данным нижележащих уровней. На конкретных примерах продемонстрированы преимущества и эффективность этого подхода, расширяющего в значительной степени возможности его использования в задачах зондирования атмосферы, в том числе и лидарными средствами локации.

В последние годы в связи с возросшими требованиями к данным атмосферного мониторинга со стороны различных отраслей народного хозяйства происходит бурное развитие новых методов и технических средств оперативного контроля за состоянием воздушной среды. Среди них несомненными преимуществами обладают методы активного дистанционного зондирования с использованием лазерных источников излучения [1, 2], поскольку они обеспечивают оценку атмосферных полей с исключительно высоким временным и пространственным разрешением и позволяют получить необходимые геофизические данные с очень большой оперативностью.

Однако несмотря на очевидные преимущества методов лазерного зондирования перед методами контактных измерений (например, радиозондовых), они имеют и ряд существенных недостатков. Главными из них являются недостаточная точность получаемых данных в свободной атмосфере (на высотах более 2–3 км) при ее зондировании лидарными системами наземного базирования и быстрый рост погрешностей измерения по мере увеличения протяженности контролируемой атмосферной трассы. Так, например, температура воздуха уже на расстоянии менее 2 км определяется лидаром с погрешностью больше чем  $2^\circ$  [2], что превышает критерий отбраковки значений температуры ( $1^\circ$  для тропосферы), установленный Всемирной метеорологической организацией для температурного зондирования атмосферы.

Поэтому в настоящее время для устранения этих недостатков (особенно при оценивании высотного распределения температуры и ветра – основных параметров атмосферы, используемых при прогнозировании погоды и моделировании климата) рекомендуется применять комплексный подход, который базируется на совместном использовании достаточно точных результатов лидарных измерений на малых высотах (до 1–3 км) и данных восстановления метеорологических величин на вышерасположенных уровнях, осуществляемого с помощью различного вида математических моделей (например, гидродинамических или физико-статистических).

Однако на практике решение задач восстановления высотных метеорологических полей на основе использования гидродинамических моделей прогноза не представляется возможным, поскольку эти модели достаточно громоздки, имеют существенные ограничения по разрешению и пределу предсказуемости и требуют для своей реализации данные наблюдений чуть ли не по всему полушарию, что чревато большими затратами машинного времени и необходимостью иметь мощные ЭВМ. К тому же вклад ошибок исходных данных в ошибки гидродинамических схем прогноза достигает 18–25 % [3].

Учитывая эти обстоятельства, для решения поставленной задачи целесообразнее использовать физико-статистические методы, которые получили широкое распространение в задачах восстановления и прогноза вертикальных профилей метеорологических величин (см., например, [4–6]). Среди подобных методов выделяется недавно разработанный модифицированный

метод группового учета аргументов [7], достаточно простой, не требующий большого объема исходных данных и значительных затрат машинного времени, не нуждающийся в предварительном статистическом осреднении многолетних рядов эмпирических наблюдений и, наконец, позволяющий на основе взятой информации, полученной до проведения численного эксперимента, синтезировать прогностическую модель в условиях частичной или полной неопределенности наших знаний о структуре моделируемого процесса и свойствах шумов в используемых данных. Этот метод и применен авторами настоящей статьи для решения задачи восстановления характеристик свободной атмосферы (и в частности, вертикальных профилей температуры и ветра) по данным пространственно-временных наблюдений.

Прежде чем перейти к результатам статистической оценки (восстановления) характеристик свободной атмосферы с помощью алгоритмов модифицированного МГУА рассмотрим коротко некоторые теоретические основы, на которых базируется данный метод.

Основная идея модифицированного МГУА, согласно [7], состоит в том, что в заданном классе функций на основе выборки экспериментальных данных автоматически генерируется некоторое множество прогностических моделей различной структуры, из которого выбирается одна или несколько лучших моделей (в смысле некоторого критерия качества), а затем с помощью выбранной модели реализуется процедура самого прогноза (или восстановления).

Очевидно, что в приведенной формулировке решение поставленной задачи требует: задания вида и объема выборки экспериментальных данных; задания класса базисных функций (операторов), из которых формируется множество прогностических моделей; определения способа генерации структур различных моделей; выбора метода оценивания параметров генерируемых моделей и метода поиска минимума критерия качества.

В нашем случае в качестве исходных экспериментальных данных, как уже было сказано, использованы пространственно-временные наблюдения

$$\{ Y_{h,t}, h = 0, 1, \dots, h^*; t = 1, 2, \dots, N \},$$

$$\{ Y_{h,t}, h = 0, 1, \dots, \bar{h} \leq h^*; t = N + 1 \} \quad (1)$$

(здесь  $h$  – высота и  $t$  – время наблюдения), а в качестве базисных функций взяты смешанные разностные динамико-стохастические модели вида:

$$Y_{h,N+1} = \sum_{s=1}^{N^*} A_{h,\tau} Y_{h,N+1-\tau} + \sum_{j=0}^{h-1} B_{h,j} Y_{j,N+1} + \varepsilon_{h,N+1},$$

$$(h = \tilde{h} + 1, \dots, h^*), \quad (2)$$

где  $N^*$  – порядок запаздывания по времени ( $N^* < [N - h - 1] / 2$ );  $A_{h,1}, \dots, A_{h,N}$  и  $B_{h,0}, \dots, B_{h,h-1}$  – неизвестные параметры модели;  $\varepsilon_{h,N+1}$  – невязка модели.

Для определения наилучшей модели (2) и успешного прогнозирования на ее основе использованы (в соответствии с [7]) все исходные данные (1), предварительно разбиваемые на выборку  $A$  (она содержит наблюдения по времени  $t = N - 1$  включительно) и выборку  $B$ , содержащую наблюдения только в момент времени  $t = N$ , и два специальных метода, а именно:

1) Метод направленного группового перебора для оптимизации структуры модели с двухэтапной селекцией моделей, проводимой с помощью:

– ф и н а л ь н о й о ш и б к и прогнозирования (Х. Акаике) вида

$$FRE = \frac{(N - N^* - 1) + s}{(N - N^* - 1) - s} RSS(s), \quad (3)$$

где  $RSS(s) = \sum_{j=1}^{N-N^*-1} [Y_{h,N-j} - \hat{Y}_{h,N-j}(s)]^2$  – остаточная сумма квадратов для текущей модели  $\hat{Y}_{h,N-j}(s)$ , содержащей  $s$  ненулевых оценок параметров. При этом оценка значения  $\hat{Y}_{h,N-j}$  осуществляется с помощью выражения

$$\hat{Y}_{h,N-j} = X \hat{\Theta}, \quad X \in M_{(N-N^*-1) \times (N^*+h)}, \quad \hat{\Theta} \in R^{N^*+h}, \quad (4)$$

где  $\hat{\Theta} = [\hat{A}_{h,1} \dots \hat{A}_{h,N^*} \hat{B}_{h,0} \dots \hat{B}_{h,h-1}]^T$  – минимальная оценка параметров по выборке  $A$ , вычисляемая по специальным формулам ( $T$  – операция транспонирования);  $R^k$  – евклидово пространство  $k$ -мерных векторов;  $M_{m \times p}$  – пространство матриц размера  $m \times p$ ;

– среднеквадратической ошибки прогноза на контрольной выборке (выборка  $B$ ):

$$|Y_{h,N} - \hat{Y}_{h,N}(s)|^2 \rightarrow \min, \quad (5)$$

где минимум берется по всем  $N^* + h$  структурам, каждой из которых соответствует своя модель  $\hat{Y}_{h,N}(s)$ .

2) Метод минимаксного оценивания для получения оценок параметров модели, позволяющей гарантировать качество соответствующего прогноза, которое может быть оценено с помощью неравенства:

$$E |E(Y_{h,N+1}) - \hat{Y}_{h,N+1}|^2 \leq \delta_{h,N+1}, \quad (h = \bar{h} + 1, \dots, h^*), \quad (6)$$

где  $E(\cdot)$  – оператор математического ожидания, позволяющий производить усреднение по всевозможным реализациям ошибок наблюдений, а  $Y_{h,N+1}$  и  $\delta_{h,N+1}$  – минимаксные оценки, зависящие от дисперсии ошибок наблюдений и от априорной информации о максимально допустимых значениях величины прогноза.

Описанная выше методика и была использована в задаче восстановления вертикальных профилей температуры ( $T$ ), зональной ( $V_x$ ) и меридиональной ( $V_y$ ) составляющих скорости ветра в свободной атмосфере. Оценка качества и эффективности модифицированного МГУА при этом проводилась на основе многолетних (1961–1975 гг.) наблюдений четырех аэрологических станций: Кефлавик (63°57'с.ш., 22°37'з.д.), Ставангер (58°33'с.ш., 05°38'в.д.), Рим (41°48'с.ш., 12°38'в.д.) и Майами (25°49'с.ш., 80°17'з.д.), представляющих различные физико-географические районы северного полушария, а точность восстановления указанных физических параметров оценивалась с помощью стандартных (среднеквадратических) ошибок  $\delta$  и относительных ошибок  $\delta/\sigma$ , выраженных в процентах (здесь  $\sigma$  – среднее квадратическое отклонение, характеризующее естественную изменчивость каждого исследуемого параметра).

На первом этапе решения поставленной задачи были проведены численные эксперименты по оценке зависимости эффективности модифицированного МГУА от взятой метеорологической величины, числа задаваемых структур прогностических моделей и порядка матриц входных переменных, зависящего от числа используемых профилей.

Численные эксперименты показали, что:

1. Модифицированный метод группового учета аргументов является достаточно эффективным методом численной оценки характеристик свободной атмосферы (в нашем случае температуры, зональной и меридиональной составляющих скорости ветра) по данным ниже лежащих уровней, в том числе и тех уровней (<Земля> и 850 гПа), для которых имеются достаточно надежные лидарные измерения. Это хорошо видно из табл. 1, в которой приведены в качестве примера (только для ст. Кефлавик, Рим и Майами) значения абсолютных ( $\delta$ ) и относительных ( $\delta/\sigma$ , %) ошибок восстановления величин  $T$ ,  $V_x$ ,  $V_y$  на различных уровнях тропосферы по данным двух уровней: <Земля> и 850 гПа (~1,5 км).

2. Данный метод дает наилучшие результаты (за редким исключением) при восстановлении характеристик ветра (особенно зональной составляющей скорости ветра  $V_x$ ), причем почти во всех случаях (независимо от местоположения станции, сезона и уровня) стандартная относительная ошибка восстановления величин  $V_x$ , как это видно из табл. 1, находится преимущественно в пределах 18–60%, а погрешность численного оценивания величин  $V_y$  не превышает в основном 60–70%.

Таблица 1

Стандартные ошибки восстановления (по данным уровня <Земля> и 850 гПа)–  $\delta$  и средние квадратические отклонения  $\sigma$  температуры ( $T, ^\circ\text{C}$ ), зональной ( $V_x, \text{м/с}$ ) и меридиональной ( $V_y, \text{м/с}$ ) составляющих скорости ветра

Уровень, гПа	T			$V_x$			$V_y$		
	$\delta$	$\sigma$	$\delta/\sigma, \%$	$\delta$	$\sigma$	$\delta/\sigma, \%$	$\delta$	$\sigma$	$\delta/\sigma, \%$
1	2	3	4	5	6	7	8	9	10
ст. Кефлавик									
Зима									
700	2,5	5,4	46	1,9	8,3	23	5,1	9,6	53
500	3,5	5,6	62	3,0	11,3	26	6,9	13,0	53
400	3,3	4,6	72	4,4	13,6	32	8,7	15,6	56
300	3,2	3,5	91	4,4	14,7	30	8,9	16,9	53
Лето									
700	1,9	3,4	56	2,5	6,4	39	3,6	6,6	55
500	2,5	3,8	66	3,1	9,1	34	3,7	9,1	40
400	2,9	4,1	70	3,8	11,5	33	7,2	11,3	63
300	3,1	3,8	82	4,0	14,6	27	6,8	14,0	48
ст. Рим									
Зима									
700	2,6	4,4	59	5,0	9,1	55	6,7	9,2	73
500	3,0	4,5	66	6,0	11,5	52	8,8	13,4	66
400	2,9	4,4	66	8,4	14,4	58	9,8	16,5	59
300	3,0	3,3	91	11,1	17,9	62	13,8	19,4	71
Лето									
700	2,5	3,5	71	4,3	7,5	57	4,3	6,6	65
500	2,1	3,3	64	5,7	8,4	67	6,0	8,7	69
400	2,3	2,8	82	7,2	10,6	68	8,2	10,5	78
300	2,5	2,6	96	9,0	13,2	69	10,6	13,6	78
ст. Майами									
Зима									
700	1,9	2,5	76	1,8	5,9	30	3,8	6,4	60
500	1,5	2,4	62	2,2	7,8	28	4,9	7,2	68
400	1,4	2,5	56	2,7	9,5	28	5,8	9,0	64
300	1,4	2,2	64	4,0	11,7	34	6,5	11,0	59
Лето									
700	0,9	1,3	69	0,9	4,3	21	2,7	3,7	73
500	1,0	1,3	77	1,2	4,9	24	2,2	3,7	59
400	0,9	1,3	69	1,3	5,8	22	3,1	4,9	63
300	1,0	1,6	62	1,3	7,4	18	4,3	5,9	73

3. Наиболее успешный прогноз (восстановление) вертикальных профилей температуры и ветра в тропосфере осуществляется в основном при задании 10 структур, определяющих лучшую (с точки зрения качества восстановления) структуру прогностической модели, и при использовании статистической выборки в объеме от 7 до 16 профилей (при временном интервале в 12 ч).

На втором этапе решения поставленной задачи были проведены численные эксперименты по дополнительной оценке эффективности минимаксного подхода к проблеме статистического оценивания характеристик свободной атмосферы. С этой целью была использована процедура сопоставления результатов восстановления температуры, зональной и меридиональной составляющих скорости ветра, проведенного как с помощью алгоритма модифицированного МГУА, так и на основе метода многомерной экстраполяции (ММЭ), получившего широкое распространение при оценке параметров свободной атмосферы по данным ниже- и вышележащих уровней (см., например, [4, 9, 10]). При этом восстановление вертикальных профилей метеорологических величин по данным нижележащих уровней (при использовании ММЭ) осуществлялось с помощью выражения вида

$$\hat{a}_0 = \bar{a} + \omega' (\mathbf{a}_i - \bar{\mathbf{a}}_i), \quad \omega' = S^{-1} \mathbf{s}, \quad (7)$$

где  $\hat{a}_0$  и  $\bar{a}_0$  – восстановленное и среднее значения метеорологической величины на каком-либо уровне восстановления;  $\mathbf{a}_i$  и  $\bar{\mathbf{a}}_i$  – вектор индивидуальных и средних значений той же величины на нижележащих уровнях;  $\omega'$  – параметр множественной регрессии, оцениваемой с помощью

выборочной ковариационной матрицы  $\|S\|$  и выборочного  $k$ -мерного вектора связи  $s$  между предиктором и предиктантом.

Таблица 2

Величины средних квадратических отклонений ( $\sigma$ ) и стандартные погрешности восстановления (по данным уровня <Земля> и 850 гПа) ( $\delta$ ) температуры ( $T$ , °С), зональной ( $V_x$ , м/с) и меридиональной ( $V_y$ , м/с) составляющих скорости ветра

Уровень, гПа	$T$			$V_x$			$V_y$		
	$\delta_1$	$\delta_2$	$\sigma$	$\delta_1$	$\delta_2$	$\sigma$	$\delta_1$	$\delta_2$	$\sigma$
700	2,6	3,0	4,4	5,0	8,0	9,1	6,7	7,3	9,2
500	3,0	5,7	4,5	6,0	8,9	11,5	8,8	9,8	13,4
400	2,9	6,3	4,4	8,4	9,2	14,4	9,8	9,9	16,5
300	3,0	3,6	3,3	11,1	12,4	17,9	15,3	15,8	19,4

В табл. 2 в качестве примера приведены погрешности восстановления вертикальных профилей температуры, зональной и меридиональной составляющих скорости ветра, осуществленного с помощью алгоритмов ММГУА (1) и ММЭ (2) по данным ст. Рим, взятым для зимы 1970–1975 гг.

Анализ данных табл. 2 показывает, что величины стандартных погрешностей  $\delta$ , рассчитанных на основе двух методических подходов, вполне сопоставимы между собой и более того почти во всех случаях (независимо от уровня и метеорологической величины) точность восстановления профилей  $T$ ,  $V_x$  и  $V_y$  существенно выше при использовании модифицированного метода группового учета аргументов, что говорит в пользу данного метода статистического оценивания характеристик свободной атмосферы.

Таким образом, на основе проведенных статистических оценок можно сделать общий и важный вывод о том, что использование алгоритма ММГУА в задаче восстановления характеристик свободной атмосферы по данным нижележащих уровней вполне допустимо (в том числе и при использовании лидарных измерений в слое <Земля> и 850 гПа) и к тому же не требует (как метод многомерной экстраполяции) предварительного статистического обобщения большого массива многолетних аэрологических наблюдений и расчета по ним параметров множественной регрессии. Следует также подчеркнуть, что еще большую эффективность алгоритм ММГУА может дать при использовании исходных данных с большим пространственно-временным разрешением и при учете синоптической ситуации в период эксперимента (когда формируется обучающая выборка), но все эти вопросы являются предметом наших будущих исследований.

1. Зуев В.Е., Зуев В.В. Дистанционное оптическое зондирование атмосферы. Санкт-Петербург: Гидрометеиздат, 1992. 232 с.
2. Межерис Р. Лазерное дистанционное зондирование. М.: Мир, 1987. 550 с.
3. Миякода К. // Теоретические основы прогноза погоды на средние сроки. Л.: Гидрометеиздат, 1979. С. 5–78.
4. Комаров В.С. // Метеорология и гидрология. 1970. № 4. С. 16–19.
5. Багров А.И., Локтионова Е.А. // Тр. ГМЦ. 1978. Вып. 212. С. 42–46.
6. Калинин А.Н., Теущикова Т.Д. // Оптика атмосферы. 1990. Т. 3. № 1. С. 60–67.
7. Кочерга Ю.Л. // Автоматика. 1991. № 5. С. 80–86.
8. Зуев В.Е., Комаров В.С. Статистические модели температуры и газовых компонент атмосферы. Л.: Гидрометеиздат, 1986. 264 с.
9. Комаров В.С. // Тр. ВНИИГМИ-МЦД, 1977. Вып. 42. С. 22–26.
10. Марченко А.С., Минакова Л.А., Семочкин А.Г. // Применение статистических методов в метеорологии. Новосибирск: Сибирское отделение АН СССР, 1971. С. 82–121.

Институт оптики атмосферы  
СО РАН, Томск

Поступила в редакцию  
9 апреля 1993 г.

V.S. Komarov, V.I. Akselevich, A.V. Kreminskii. **Modified Method of Group Allowing for Arguments as an Efficient Technique for Statistical Estimation of Characteristics of Free Atmosphere Under Deficit of Information.**

Prospects and potentialities of the modified method of group allowing for arguments (MGAA) for forecasting (restoring) the vertical temperature profiles and zone and meridian components of wind rate based on data of lower atmospheric levels at deficit of experimental information are discussed. The advantages and efficiency of the approach are demonstrated by concrete examples, which show its applicability to the problems of atmosphere sounding.