УДК 519.6

ИНДУКТИВНЫЙ МЕТОД САМООРГАНИЗАЦИИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ВПРИЛОЖЕНИИ К ЗАДАЧАМ ПРИКЛАДНОЙ МЕТЕОРОЛОГИИ

К.Я. Синёва, В.С. Кичкильдинов*, В.Т. Калайда**, В.М. Белов***

Томский политехнический университет
*Отдел УМВД России по г. Томску
**Томский государственный университет
***Сибирский государственный университет телекоммуникаций и информатики, г. Новосибирск
E-mail: klarayakovlevna_@mail.ru

Разработан алгоритм восстановления, краткосрочного и долгосрочного прогнозирования для полей метеорологических величин с использованием идей индуктивного метода самоорганизации моделей на ЭВМ. Получены математические модели оптимальной сложности пространственно-временного краткосрочного и долгосрочного прогноза на небольшие периоды по времени, а также модели восстановления по высоте вертикальных профилей температуры в тропосфере по многолетним наблюдениям метеорологической станции Минск (53°11' с.ш., 27°32' в.д.)

Ключевые слова:
Метод группового учёта аргументов, самоорганизация прогнозирующих моделей, прикладная метеорология, статистика.

Key words:
Method of argument group account, self-organization of predictive models, applied meteorology, statistics.

Прогноз погоды относят к научным проблемам, которые пока ещё не получили точного решения, несмотря на объединённые усилия учёных и практиков всего мира. Проведён обзор отечественной и зарубежной литературы, представляющей анализ современных методов прогноза основных элементов погоды: температуры и осадков [1?3]. Цель работы: повышение точности восстановления и прогнозирования характеристик атмосферы путём разработки адаптивных методов и алгоритмов анализа и обработки метеорологических полей. Наиболее популярным и, вероятно, эффективным аппаратом является математическая статистика, особенно в случае прогноза погоды более чем на сутки. Однако в условиях сложных задач моделирования нужно синтезировать комбинированные подходы. В настоящей статье предложен комбинаторный полиномиальный алгоритм синтеза моделей восстановления, краткосрочного и долгосрочного прогнозирования для полей метеорологических величин. Проведён физико-статистический анализ вертикального профиля полей температуры в тропосфере с целью получения прогнозных и восстановительных математических моделей метеорологической величины для станции Минск.

Комбинаторный полиномиальный алгоритм синтеза моделей восстановления, краткосрочного и долгосрочного прогнозирования для метеорологических величин

Алгоритм индуктивного метода самоорганизации моделей можно представить в виде общей структурной схемы алгоритмов самоорганизации прогнозных моделей на рис. 1. Состоит из генератора моделей (ГМ), на выходе которого (в определённом порядке повышения сложности) получаются варианты (претенденты) моделей и селектирующего устройства (К), выбирающего по заданному внешнему критерию самую лучшую модель (В) [4, 5]. В зависимости от устройства генератора моделей-претендентов различают три основные структуры алгоритмов самоорганизации: комбинаторные, многорядные, многорядные пороговые.

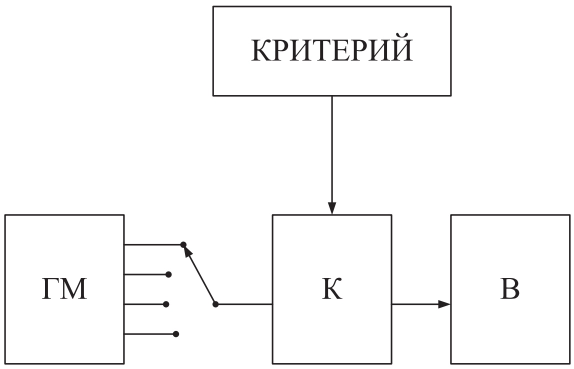

Рис. 1. *Общая структурная схема алгоритмов самоорганизации моделей на ЭВМ*

Комбинаторный алгоритм, представленный на рис. 2, соответствует генератору, который производит выбор модели из полного полинома приравниванием нулю («зануления») тех или других его слагаемых. Например, из полинома второй степени $y=a_0+a_1x+a_2x^2$ получаем три укороченных, частных полинома: $y=a_0+a_1x$; $y=a_0+a_2x^2$; $y=a_1x+a_2x^2$ и один полный полином. Происходит аналогичное «зануление» слагаемых для полинома третьей степени $y=a_0+a_1x+a_2x^2+a_3x^3$ и т. д.

Таким образом, получаем таблицу полиномов с постепенным усложнением структуры. Каждый из полиномов должен быть оценён по избранному критерию селекции. Полином с наилучшим значением критерия и является полиномиальной моделью оптимальной сложности. Производится пересчёт оценок коэффициентов модели по всем точкам таблицы опытных данных, рекомендуется, когда число слагаемых полного полинома не превышает 20. При этом число сравниваемых моделей равно 2^{20}.

Критерий селекции достигает глубины своего минимума, если его значение становится равным 10^{-2}–10^{-5} [4]. Полином с первым наилучшим значением критерия и является полиномиальной моделью оптимальной сложности. Таблица исходных данных делится на две части, называемые обучающей (A) и проверочной последовательностями (B). При этом среднеквадратическая ошибка $\Delta(B)$, определяемая на проверочной последовательности, является одним из критериев для выбора структуры модели, синтезируемой по данным обучающей последовательности. Так создается первое внешнее дополнение. Таблицу исходных данных приходится делить на три части: обучающую, проверочную и две экзаменационные последовательности (A,B и C,D). Ошибка $\Delta(C)$, определяемая на экзаменационных данных, может служить вторым внешним дополнением, поэтому таблицу надо разделить так, чтобы $\Delta(C)\to\min$, требуется еще одно, третье, внешнее дополнение $\Delta(D)\to\min$ [3]. Мы приняли для критерия минимума смещения по формуле следующее разделение таблицы исходных данных:

$$N_A = 0,3;\; N_B = 0,3;\; N_C = 0,2;\; N_D = 0,2;$$

где N_A, N_B, N_C, N_D – число точек обучающей, проверочной, первой и второй экзаменационных последовательностей соответственно; N – общее число точек [5].

Целесообразно для определённости различать заблаговременность прогноза и прогностический период (срок его действия) как две самостоятельные характеристики [3–5]. Исследуется объект, описываемый параметрами $Y=(y_1,y_2,...,y_m)$, в моменты времени t_{y_j}, $j=\overline{1,m}$, где m–5, m–4, m–3, m–2, m–1, m – день. Например, объект удалось описать в терминах другого, более доступного во времени или в пространстве объекта $X=(x_1,x_2,...,x_n)$, как функционал $Y=f(x)$ на множестве реализаций X. При этом характеристики объекта X (наблюдаемые или рассчитанные), как правило, тоже относят к различным моментам времени t_{x_i}, $i=\overline{1,n}$. Такое исследование является прогностическим, если $\max\limits_i t_{x_i} < \min\limits_j t_{y_j}$. В противном случае экстраполяция во времени, а значит и прогноз отсутствует вообще. Далее условимся, что интервал времени $\tau = \min\limits_j t_{y_j} - \max\limits_i t_{x_i}$ – будем называть заблаговременностью прогноза, а момент времени $\max\limits_i t_{x_i}$ – моментом прогноза, а интервал времени $\tau_y = \max\limits_j t_{y_j} - \min\limits_j t_{y_j}$ означает прогностический период или срок действия прогноза. Прогностические работы подразделяются следующим образом: 1) краткосрочный прогноз (τ, τ_y в пределах суток, что соответствует радиусу инерции большинства метеорологических объектов); 2) прогноз на малые периоды (τ, τ_y в пределах радиуса корреляции метеорологических объектов, что в среднем составляет до 10 суток); 3) прогноз на большие периоды (τ, τ_y больше радиуса корреляции – месяц, сезон, год).

Внешние критерии, используемые в комбинаторном полиномиальном алгоритме синтеза моделей восстановления, краткосрочного и долгосрочного прогнозирования для полей метеорологических величин

Критерий минимума смещения, основанный на анализе решений

Модели, получаемые при использовании различных частей таблицы исходных данных, должны, по возможности, мало отличаться друг от друга, а при полном отсутствии смещения – совпадать. Выборка делится на две равные части. Модель на подвыборке A должна как можно меньше отличаться от модели на подвыборке B и наоборот по формуле:

$$n_{см}^2 = \frac{\sum\limits_{i=1}^{N}(y_{A_i} - y_{B_i})^2}{\sum\limits_{i=1}^{N} y_i^2},$$

где N – длина выборки; y_{A_i}, y_{B_i} – аппроксимация табличных данных на выборках A и B; y_i – значения выходной величины.

Критерий баланса переменных и его применение для синтеза моделей

На интервале интерполяции устанавливается некоторая математическая связь между переменными, образующими систему. Критерий баланса требует, чтобы связь, называемая «законом», сохранялась на интервале прогноза. Если такой «закон» действительно существует и остается постоянным, то критерий баланса способен обеспечить достаточно точные долгосрочные прогнозы. Пусть $g_1(t),g_2(t),...,g_s(t)=0$ – функция баланса – «закон», связывающий переменные $g_j(t)$, $j=1,...,S$. Из множества всех прогнозирующих моделей для переменных $g_i(t)$ нужно выбрать такую систему, для которой на интервале экстраполяции (в районе точки прогноза) эта связь выполняется наилучшим образом. Нарушение баланса переменных характеризуют величинами небаланса $b_i=f(g_1(t_i),g_2(t_i),...,g_s(t_i))$, которые рассчитываются для t_i принадлежащих интервалу экстраполяции для каждой переменной $g_j(t)$. На интервале интерполяции строятся несмещенные законы: $g_1=f_{1обр}(g_2,g_3,...,g_s)$; $g_2=f_{2обр}(g_1,g_3,...,g_s)$; $g_s=f_{s\,обр}(g_1,g_2,...,g_{s-1})$. Причем, i-я обратная функция находится из j-й прямой ($i\neq j$). Величины $b_{1i}=(f_1-f_{1обр})_i^2$, $b_{2i}=(f_2-f_{2обр})_i^2,...,b_{si}=(f_s-f_{s\,обр})^2$, рассчитанные на интервале экстраполяции по прогнозирующим моделям для каждой переменной, характеризуют небаланс системы прогнозов в точке. Критерий баланса по формуле

$$B = \sum_{j=1}^{s} \frac{\sum_i b_{ij}}{\sum_i f_i^2(t)}$$

используется для выбора оптимальной модели.

Комбинированный критерий селекции

Для гармонических и алгебраических моделей, использующих функции времени (среднесрочный и долгосрочный прогнозы), в случаях, когда известен или может быть открыт закон физического баланса переменных, рекомендуется критерий «несмещение плюс баланс переменных». При формировании комбинированного критерия приняты единичные веса для нормированных составляющих баланса и несмещенности по формуле:

$$K = \sqrt{\left(\frac{B}{B_{\max}}\right)^2 + \left(\frac{n_{см}^*}{n_{см\max}^*}\right)^2},$$

где B, $n^*_{см}$ – небаланс и смещение для конкретной комбинации моделей системы; $B_{\max}$, $n^*_{см\max}$ – их максимальные значения на множестве возможных комбинаций.

Краткосрочное и долгосрочное прогнозирование и восстановление по высоте вертикальных профилей температуры

С помощью разработанного комбинаторного полиномиального алгоритма синтеза моделей для восстановления, краткосрочного и долгосрочного прогнозирования полей метеорологических величин по методу группового учёта аргументов было принято решать задачи пространственно-временного прогнозирования и восстановления высотного распределения вертикальных профилей температуры по формуле на основе пространственно-временных наблюдений вида:

$$\{T_{h,t}\ \ h = 0,1,...,h_k;\ t = 1,...,N\},$$

где h – высота; h_k – максимальная высота наблюдений; t — время наблюдений; N – длина выборки. В качестве исходных данных принимали значения многолетних (1961–1978 гг.) радиозондовых измерений температуры T, °С метеорологической станции Минск (53°11′с.ш., 27°32′в.д.). Измерения метеорологических величин производятся по стандартным изобарическим поверхностям. В существующей практике радиозондовые наблюдения обычно проводятся в два срока (в 0 и 12 ч по Гринвичу), т. е. с интервалом в 12 ч. Наблюдения были интерполированы (метод линейной интерполяции) на сетку стандартных высот, являющуюся наиболее удобной и используемой для решения прикладных задач прогноза температуры, в тринадцати точках (0,0 (уровень земли); 0,1; 0,2; 0,4; 0,8; 1,2; 1,6; 2,0; 2,4; 3,0; 4,0; 5,0; 8,0 км).

Сетка стандартных высот позволяет описать с большим вертикальным разрешением почти всю тропосферу [1–3]. В качестве входных данных для группы «обучения» (интервал интерполяции) приняли температуру по первым 10 суткам одного месяца из сезона (например, август 1975 г). Вводим в программу 10 значений температуры (с 1 по 10 августа) на высотах 0,0...8,0 км. Данные аэрологических наблюдений взяты с временным интервалом в 24 ч., а для высот h>3 км их целесообразно брать с интервалом в 12 ч. В качестве прогнозной функции задаём степенной полином 9 степени. Рекомендуем задавать степень полинома до 14, учитывая пространственно-временное распределение высотного вертикального профиля температуры в тропосфере. Прогноз делаем на следующие 10 суток (с 11 по 20 августа) соответственно по каждой стандартной высоте сетки. В результате программа выдает математические прогнозные модели, их коэффициенты, значения комбинированного критерия, смещения и баланса.

Прогноз делается по всей выборке – по 20 точкам. Модель оптимальной сложности, дающая объективный прогноз, выбирается по минимуму комбинированного внешнего критерия, его значение стало равным $K \le 10^{-3}$. Комбинированный критерий выдает модель оптимальной сложности – полином третьей степени. Абсолютную погрешность определяют как разность $\Delta T = (T_ф - T_{пр.})$ – отклонение между фактическим и прогнозным значениями. В качестве допустимой средней абсолютной ошибки прогноза (погрешности) метеорологических величин примем величину менее ±1,5 °С для краткосрочного и менее ±2,5 °С – для долгосрочного прогнозов, что обычно используется в практике статистических прогнозов, восстановлений и допускается Всемирной метеорологической организацией для радиозондовых наблюдений [1–3].

Таблица 1. *Модели оптимальной сложности для краткосрочного и долгосрочного пространственно-временного прогнозирования по времени 1–10 суток для температуры в тропосфере для станции Минск (53°11′ с.ш., 27°32′ в.д.), август 1975 г.*

h, км	Модели Y для температуры T, °С	Критерий K	Абсолютная погрешность T, °С
0,0	$26{,}88+0{,}45X-0{,}22X^2+0{,}40X^3$	$4{,}21\cdot10^{-4}$	±2,35
0,1	$23{,}58+0{,}70X-0{,}22X^2+0{,}50X^3$	$4{,}15\cdot10^{-4}$	±2,45
0,2	$21{,}24+0{,}79X-0{,}23X^2+0{,}40X^3$	$4{,}23\cdot10^{-4}$	±2,60
0,4	$16{,}58+0{,}98X-0{,}24X^2+0{,}30X^3$	$4{,}41\cdot10^{-5}$	±2,02
0,8	$11{,}86+1{,}17X-0{,}26X^2+0{,}20X^3$	$4{,}12\cdot10^{-5}$	±1,48
1,2	$7{,}16+1{,}58X-0{,}29X^2+0{,}30X^3$	$5{,}02\cdot10^{-5}$	±2,43
1,6	$4{,}72+1{,}47X-0{,}27X^2+0{,}21X^3$	$6{,}16\cdot10^{-5}$	±2,94
2,0	$2{,}73+1{,}29X-0{,}25X^2+0{,}50X^3$	$6{,}95\cdot10^{-5}$	±2,96
2,4	$2{,}66+0{,}84X-0{,}23X^2+0{,}70X^3$	$8{,}34\cdot10^{-5}$	±2,03
3,0	$-9{,}40+0{,}66X-0{,}18X^2+0{,}60X^3$	$3{,}18\cdot10^{-5}$	±2,47
4,0	$-9{,}41+0{,}67X-0{,}10X^2+0{,}80X^3$	$2{,}30\cdot10^{-4}$	±3,78
5,0	$-22{,}33+0{,}34X-0{,}11X^2+0{,}40X^3$	$4{,}42\cdot10^{-4}$	±2,97
8,0	$-35{,}08-0{,}20X+0{,}12X^2-0{,}30X^3$	$6{,}90\cdot10^{-4}$	±3,58

До высоты 3,0 км включительно мы имеем хорошие результаты прогнозирования. Здесь абсо-

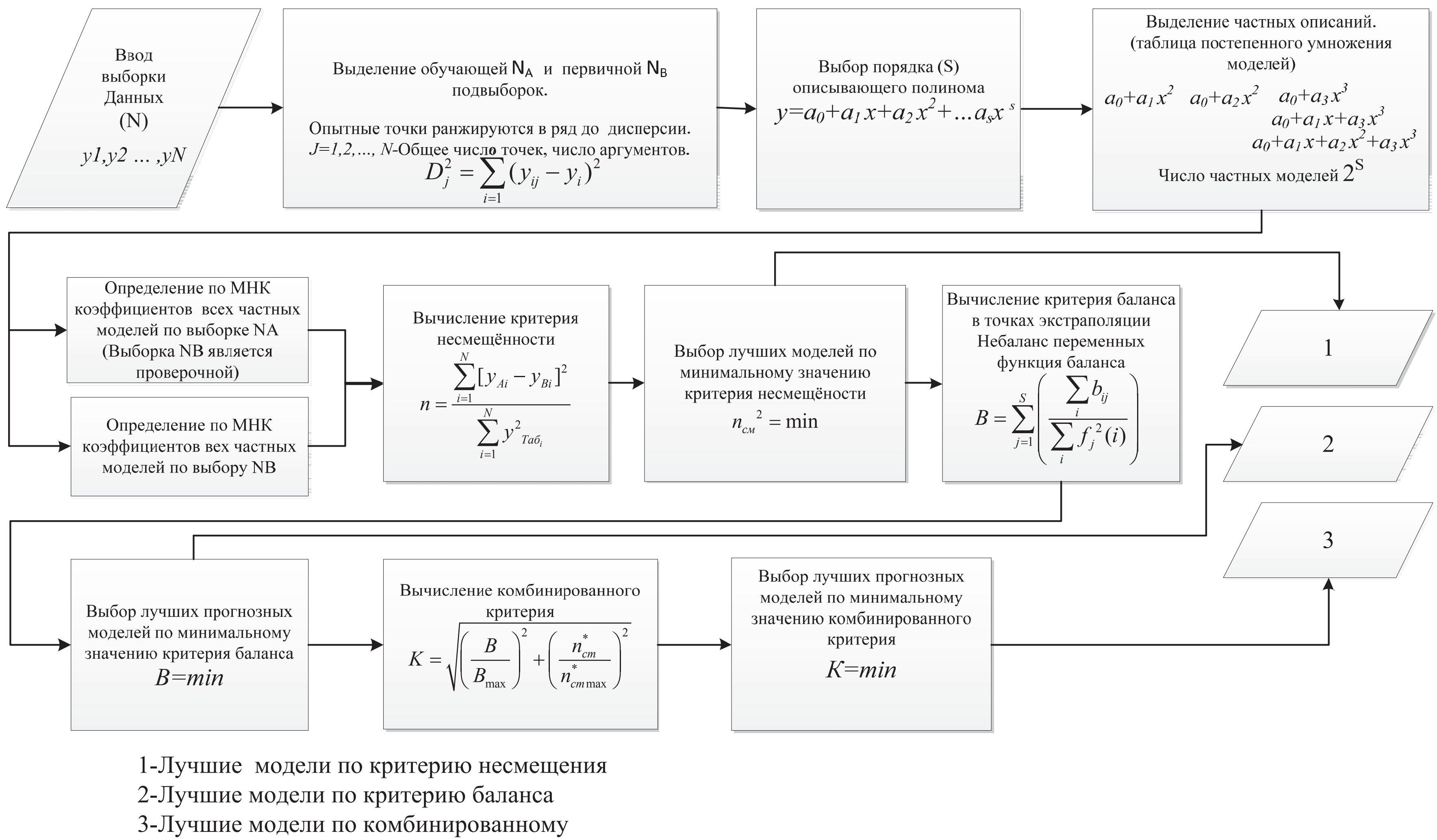

Рис. 2. *Структурная схема комбинаторного полиномиального алгоритма синтеза моделей восстановления, краткосрочного и долгосрочного прогнозирования для метеорологических величин. МНК – метод наименьших квадратов*

лютная погрешность прогноза ΔT не превышает величину ±1,5 °С для прогноза 1–10 суток. Начиная с высоты 4,0...8,0 км, некоторые прогнозные значения превышают допустимую в практике статистических прогнозов величину абсолютной погрешности ±1,5 °С. На рис. 3 представлены прогнозные модели Y температуры и графики пространственно-временного прогноза температуры за период 10 дней соответственно своей высоте. Значения внешнего комбинированного критерия K, абсолютной погрешности прогнозирования на небольшие периоды температуры по времени для соответствующей модели Y представлены в табл. 1. Максимальное отклонение ΔT прогнозного значения от фактического (абсолютная погрешность) составило ±3,58 °С, минимальное составило ±2,35 °С. Для критериев селекции использовалась функция баланса, учитывающая высотное распределение вертикального профиля температуры по формуле:

$$f = g_2^2 + 6 \cdot g_2 + 6{,}4 \cdot g_1 - 3{,}8,$$

где g_1 – значения температуры по высоте 0,0...3,0 км по первым 10 точкам месяца, сезона, года; g_2 – по высоте 4,0...8,0 км по первым 10 точкам месяца, сезона, года.

Далее, используя те же данные, которые использовались при пространственно-временном прогнозе на небольшие периоды по времени 1–10 суток, решаем задачу восстановления по высоте вертикального профиля температуры соответственно своему прогнозируемому дню (τ, τ_y в пределах суток), т. е. решаем задачу восстановления на вышележащих уровнях по данным наблюдений на нижележащих уровнях. Восстановление температуры осуществляется по стандартным высотам сетки (0,0 (уровень земли); 0,1; 0,2; 0,4; 0,8; 1,2; 1,6; 2,0; 3,0; 4,0; 5,0; 8,0 км) за каждые сутки из 10 дней месяца. На интервале интерполяции в качестве данных для группы «обучения» используются значения температуры, начиная с высоты 0,0...3,0 км. Интервал, где расположены точки, соответствующие температуре по высотам 4,0...8,0 км – интервал экстраполяции. Алгоритм выдаёт математические модели восстановления – решает задачу восстановления по высоте вертикальной структуры профиля температуры на всём интервале 0,0...8,0 км. Модель оптимальной сложности выбирается по первому минимуму внешнего комбинированного критерия селекции, равному $K \leq 10^{-3}$ – это алгебраический полином 3 степени. Модели восстановления по высоте вертикальных профилей температуры Y представлены соответственно своему дню (суткам) на рис. 4, а так же значения внешнего комбинированного критерия K, абсолютные погрешности восстановления показаны в табл. 2. При восстановлении максимальное отклонение ΔT значения температуры от исходного (абсолютная погрешность) составило ±1,38 °С, минимальное составило ±0,01 °С.

На всем интервале восстановления (интерполяции и экстраполяции) абсолютная ошибка восстановления температуры в тропосфере не превышает принятое допустимое значение в практике статистических восстановлений ±1,5 °С [1–3].

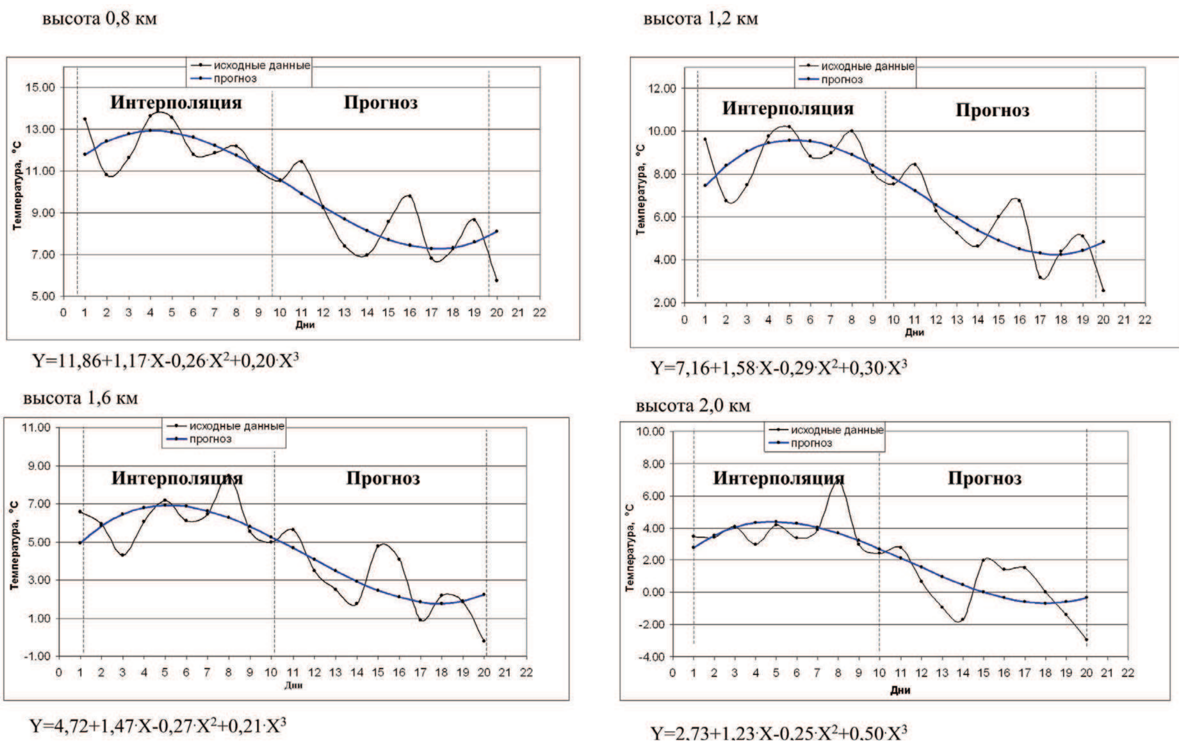

Рис. 3. *Графики и модели пространственно-временного прогнозирования температуры по времени 1–10 суток в тропосфере по стандартным высотам сетки для станции Минск (53°11' с.ш., 27°32' в.д.), август 1975 г.*

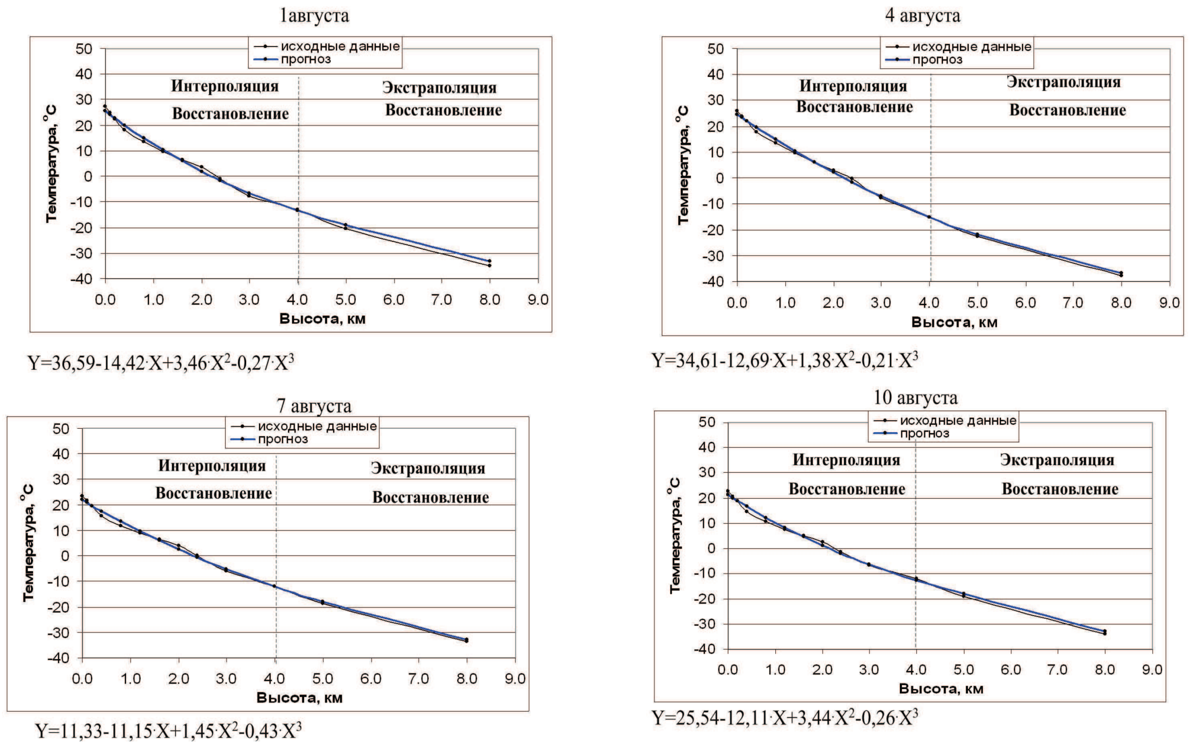

Рис. 4. *Графики и модели восстановления вертикальных профилей температуры соответственно своему прогнозируемому дню по высоте в тропосфере для станции Минск (53°11' с.ш., 27°32' в.д.), август 1975 г.*

Таблица 2. *Модели оптимальной сложности восстановления вертикальных профилей температуры по высоте (0,0...8,0 км) для станции Минск (53°11' с.ш., 27°32' в.д.), август 1975 г.*

№ суток	Модели Y для температуры T, °С	Критерий K	Абсолютная погрешность ΔT, °С
1	$36{,}59-14{,}42X+3{,}46X^2-0{,}27X^3$	$2{,}03\cdot10^{-4}$	±0,01
2	$45{,}38-13{,}82X+1{,}83X^2-0{,}35X^3$	$2{,}05\cdot10^{-5}$	±1,13
3	$14{,}83-12{,}42X+1{,}35X^2-0{,}33X^3$	$3{,}04\cdot10^{-4}$	±0,24
4	$34{,}61-12{,}69X+1{,}38X^2-0{,}21X^3$	$3{,}13\cdot10^{-5}$	±1,35
5	$33{,}31-12{,}47X+1{,}45X^2-0{,}21X^3$	$2{,}09\cdot10^{-5}$	±1,36
6	$12{,}43-12{,}69X+3{,}25X^2-0{,}21X^3$	$3{,}11\cdot10^{-5}$	±0,78
7	$11{,}33-11{,}15X+1{,}45X^2-0{,}43X^3$	$4{,}22\cdot10^{-4}$	±1,45
8	$39{,}08-10{,}86X+2{,}74X^2-0{,}12X^3$	$4{,}15\cdot10^{-5}$	±0,69
9	$21{,}33-12{,}15X+2{,}21X^2-0{,}46X^3$	$5{,}13\cdot10^{-4}$	±1,39
10	$25{,}54-12{,}11X+3{,}44X^2-0{,}26X^3$	$4{,}23\cdot10^{-4}$	±0,98

Выводы

Построены математические прогностические модели высотных профилей температуры, позволяющие прогнозировать и восстанавливать изменения температуры в период времени 1–10 суток по сетке стандартных высот полей метеорологической величины температуры.

Абсолютная погрешность прогнозирования температуры по времени на высотах 0,0...3,0 км не превышает ±2,5 °С. На высоте 4,0...8,0 км максимальное отклонение некоторых прогнозных значений от исходных составило ±3,58 °С.

Абсолютная погрешность восстановления температуры на высотах 0,0...8,0 км не превышает на большинстве уровней допустимое в практике статистического восстановления и прогнозирования значение ±1,5 °С.

Разработанный алгоритм эффективен для решения задач прогнозирования и восстановления полей температуры в тропосфере, и его можно рекомендовать для определённого класса прикладных задач по метеорологии.

СПИСОК ЛИТЕРАТУРЫ

1. Зуев В.Е., Комаров В.С. Метеорологические исследования в ИОА СО РАН в период с 1980–1999 гг. // Оптика атмосферы и океана. – 2000. – Т. 13. – № 1. – С. 18–30.
2. Комаров В.С., Дубовик К.Ю., Попов Ю.Б., Лавриненко А.В. Пространственная интерполяция метеорологических полей с помощью малопараметрической динамико-стохастической модели с вертикальной компонентой // Оптика атмосферы и океана. – 2010. – Т. 23. – № 12. – С. 1075–1079.
3. Комаров В.С. Статистика в приложении к задачам прикладной метеорологии. – Томск: Изд-во СО РАН, 1997. – 256 с.
4. Ивахненко А.Г., Мюллер И. Самоорганизация прогнозных моделей. – Киев: Техніка, 1985. – 350 с.
5. Справочник по типовым программам моделирования / под ред. А.Г. Ивахненко. – Киев: Техніка, 1980. – 182 с.

Поступила 29.06.2012 г.