К.Г. Рубинштейн, И.Н.Новикова

Испытание метода статистической детализации сезонного прогноза погоды для региона Москвы

Введение

В настоящее время во многих центрах мира интенсивно разрабатываются методы численного прогноза на сезон. Результаты подобных прогнозов пока что не всегда хороши, в связи с этим разрабатываются различные методы, с помощью которых делается попытка уменьшения систематических ошибок.

Один из подобных подходов в англоязычной литературе называют статистический “downscaling”. Под термином “downscaling” (ДС) подразумевается более широкое понятие – это методы статистической, динамической [16] или вариационной [13] региональной интерпретации результатов прогнозов крупномасштабных гидродинамических моделей.

В отечественной терминологии нет установившегося аналога английского термина “downscaling”. Буквальный перевод английского термина *downscaling* – уменьшение масштаба – довольно точно, хотя и не полно, отражает суть этого метода. В работе [9] предполагается переводить его как “*детализация*” прогноза. Гораздо раньше, примерно лет на десять, в работах отечественных авторов появился термин ‘*телескопизация*’ прогноза, близкий к этому современному термину, хотя и имевший оттенок численной процедуры совмещения модельных расчетов на сетках различного пространственного разрешения.

Интерес к подобным методам в последние годы связан с недостатками в численном прогнозировании погоды на сроки месяц-сезон и попытками улучшить их результаты с помощью специально разработанных процедур.

Проведенные многочисленные эксперименты показали, что статистические модели позволяют восстановить более 70% изменчивости мелкомасштабного поля. Ошибка решения задачи детализации может быть значительная, причем, следует отметить, что с увеличением среднеквадратичной ошибки восстановления растет вероятность “выброса”, т.е. неоправданного прогноза экстремальной метеорологической величины, что может сделать неприемлемым использование ряда статистических моделей для решения практических задач. Для построения статистических моделей необходимо знать первые и вторые моменты рассматриваемых величин, оценка которых проводится на основе априорной информации, а именно, временных рядов наблюдений.

1. Краткий обзор статистических методов детализации прогнозов

Так как в данной работе использовались результаты статистических методов интерпретации, для обоснования выбранных методов кратко опишем существующие методы статистического ДС.

1.1. Метод аналогов

В метеорологии метод был внедрен Н.А.Багровым [1, 2, 3] и получил широкое распространение для статистического прогноза метеорологических полей [6, 7, 8, 29, 31] и для краткосрочного климатического прогноза [17, 36]. Для задач ДС метод применялся в работах [18, 21, 39].

Этот метод в [40] был иллюстрирован на примере ДС зимних осадков (декабрь-февраль) на Иберийском полуострове, представленных данными нерегулярной сети наблюдательных станций. Авторы считают, что региональная переменная (осадки) определяется изменчивостью атмосферной циркуляции в Европейско-Североатлантическом секторе. В качестве крупномасштабной переменной было взято поле давления на уровне моря (SLP).

Недостатком метода аналогов можно считать определенный субъективизм принятия решений. Весьма трудно обосновать использование первых пяти эмпирико-ортогональных функций (ЭОФ) и влияние уменьшения их количества на суммарную дисперсию и на конечный результат. Кроме того, ЭОФ более высоких порядков (выше 3) с трудом поддаются какой-либо интерпретации.

1.2. Линейные методы

Линейные методы наиболее популярны в задачах ДС в первую очередь в связи с их простотой. Основная идея линейных методов – использование линейной связи между аномалиями крупномасштабной циркуляции и аномалиями локальных переменных [15, 17, 19, 24, 33, 35; 37].

Особенности линейных методов ДС были изучены в [37] на рассмотренном выше примере зависимости между SLP над Северной Атлантикой и зимними осадками на Иберийском полуострове, но по более частой наблюдательной сети. Представленная модель была построена на основе метода канонических корреляций. Этот метод позволяет определить пары полей, временная эволюция которых оптимально коррелирована. Он широко применялся как отечественными авторами [4, 5, 11], так и зарубежными [17, 19].

В работе [37] канонические корреляции рассчитывались для первых двух ЭОФ поля SLP над Северной Атлантикой. В этой работе считается, что уже первые две ЭОФ описывают около 80% общей изменчивости. Полученные результаты позволили авторам оценить региональные аномалии осадков на наблюдательных станциях для данного поля крупномасштабной аномалии давления на уровне моря. Необходимо отметить, что анализ канонических корреляций дает оценку максимально возможной корреляции между двумя полями. Учитывая приведенные ими же оценки, нетрудно подсчитать, что в результате было воспроизведено около 60% суммарной дисперсии.

В работе [25] было показано, что методы множественной регрессии и дискриминантного анализа являются частными случаями метода канонических корреляций. Метод множественной регрессии, вероятно, является самым распространенным статистическим методом в метеорологии, в том числе и для задач статистического ДС. В работе [33] представлена регрессионная модель для прогнозирования региональной температуры по результатам модели общей циркуляции атмосферы (ОЦА) в концепции, подобной MOS (Model Output Statistics). Уравнения регрессии для средних сезонных величин были найдены для 8 городов США и сравнивались с результатами совместной климатической модели (CCM). Большинство этих уравнений описывают более чем 90% дисперсии температурных рядов. Аналогичные исследования были выполнены в [34]. Модель ДС также строилась на уравнениях регрессии для 31 пункта центрального района Аргентины. В качестве предикторов использовались значения моделируемой температуры на высоте 2 м, давление на уровне моря и составляющие скорости ветра на уровнях 700 и 200 гПа. Последние, по мнению авторов, необходимы для учета зоны струйного течения, влияющего на погоду в этих широтах. Как видно из представленных результатов, наилучшую корреляцию с локальной температурой имеют модельные значения температуры на уровне 2 м (0,64–0,93), с остальными параметрами наблюдается гораздо худшая отрицательная корреляция (-0,35 – -0,48). Все значения признаны статистически значимыми. Проверка на независимом материале показала существенное уменьшение описываемой дисперсии температуры. Данный пример хорошо иллюстрирует основную проблему регрессионных моделей: определение состава и размеров вектора-предиктора.

Другая проблема заключается в разной дисперсии поля локальной переменной (предиктанта) и крупномасштабного поля (предиктора). В работе [38] предлагается использовать технику «инфляции», которая, по мнению автора, делает прогнозируемую изменчивость более реалистичной.

1.3. Методы классификации

Методы классификации не обязательно могут быть линейными, но мы ограничимся только ими. Классификационные схемы атмосферной циркуляции строятся на исторических наблюдениях в районе, для которого производится ДС, распределенных по определяемым классам. Полученные критерии классификации применяются затем к крупномасштабной циркуляции; таким образом, для каждого класса локальной переменной находятся синхронные наблюдения крупномасштабной переменной. Величина локальной переменной может быть выбрана как любая средняя из всех региональных наблюдений элементами одного класса крупномасштабной переменной или принадлежать только одному элементу выбранного класса. Применение конкретной стратегии зависит от поставленной задачи. Оценка выполненной классификации должна базироваться на показателях, включающих временную структуру, такую как продолжительность.

Практической проблемой является выбор метода для классификации крупномасштабной переменной. Существует множество методов классификации, и все они в той или иной степени субъективны, хотя, однажды определенные, они позволяют создать программируемую схему классификации объектов крупномасштабной атмосферной циркуляции. В большинстве схем классификации берется только область, над которой исследуются данные, и число классов определяется субъективно на стадии ее разработки.

Гораздо большей сложностью отличаются схемы объективной классификации [например, 14]. Одним из общих свойств объективной классификации является необходимость задания метрики, т.е. численной меры различия между любыми парами одноименных признаков, а также метрики пространства признаков, т.е меры различия между любыми парами классифицируемых объектов с учетом значения их признаков.

Для задач ДС используются:

- схемы классификации, которые зависят только от данных крупномасштабной циркуляции;
- схемы классификации, зависящие только от локальной переменной;
- схемы, которые используют оба типа данных.

Типичным примером анализа из первой и второй групп является применение кластер – анализа типов атмосферной циркуляции [20].

1.4. Кригинг

В работе [18] исследуется возможность использования ДС как процедуры интерполяции, известной как кригинг. Основная идея метода заключается в том, что значение переменной в точке интерполяции ищется как взвешенная сумма значений этой переменной в известных точках. Этот метод может быть использован практически для всех данных, обладающих пространственной корреляцией. Авторы [18] представили пример использования этого метода для реконструкции месячных сумм зимних осадков на Иберийском полуострове по полю приземного давления в зимний сезон над Северной Атлантикой. Пространственная размерность поля была уменьшена с помощью ЭОФ. Количество осадков, связанное с этой точкой, оценивалось кригинг-интерполяцией в пространстве ЭОФ.

В целом кригинг и метод аналогов реалистично воспроизводят средние месячные значения осадков. В некоторой степени кригинг лучше описывает пространственные изменения, но в отличие от метода аналогов он занижает временную дисперсию, т.к. обладает хорошо известным сглаживающим свойством. Ограничением в применении кригинга является требование стационарности. Однако на практике такое сильное предположение неприменимо. Вместо этого используют более слабые предположения.

1.5. Нейронные сети

Нейронные сети (НС) в последние годы нашли широкое применение как статистический инструмент. Достаточно полный обзор на эту тему можно найти в [30]. В метеорологии этот метод был, по-видимому, впервые применен в 1992 г. [23]. Технология НС имеет достаточно широкие возможности, но известно сравнительно мало работ, в которых она применяется задаче ДС [26, 27].

Была построена НС для описания суточных осадков в зимний сезон в Испании за период 1978-1983 гг. [40]. Входными параметрами, как уже было сказано, являются коэффициенты первых пяти ЭОФ приземного поля давления, рассчитанные на суточных значениях. Согласие с наблюдениями получается лучше для сглаженных рядов, что указывает на существование высокочастотной изменчивости, не описываемой рассмотренными методами. Однако осталось неясным, какой вклад вносит каждый параметр и как использование последних по времени наблюдений улучшает прогноз.

Весь проведенный анализ методов ДС показывает, что основной проблемой ДС является проблема уровня моделируемой изменчивости. Линейные модели и нейронные сети, в целом описывают частичную связь между «независимыми» переменными, представляющими крупномасштабную изменчивость и «зависимыми» локальными переменными. Часть локальных переменных, не описанных независимыми переменными, обычно считаются шумом. Некоторые авторы [28] используют завышенные коэффициенты регрессии линейной модели, чтобы увеличить изменчивость на выходе. Чему отдать предпочтение – воспроизведению средней или воспроизведению изменчивости, зависит от специфики конкретной задачи.

2. Реализованный метод множественной регрессии

2.1. Исходные данные

В работе использовались сезонные прогнозы температуры, полученные с помощью моделей AGCM2 (Second Generation Atmospheric General Circulation Model – модель общей циркуляции атмосферы второго поколения) и глобальной спектральной модели SEF Канадского Климатического Центра (Canadian Centre for Climate Modelling et Analysis – CCCMA) (www.cccma.bc.ec.gc.ca/models/gcm2.shtml) [32, 22].

Для анализа были взяты прогнозы для четырех сезонов за 26 лет (1969-1995 гг.). Модели AGCM2 и SEF являются глобальными моделями с пространственным разрешением 3,75° по долготам и широтам, т.е. слишком грубым для описания локальных или региональных особенностей.

В качестве потенциальных предикторов исследовались параметры выходной продукции AGCM2 и SEF: прогноз следующих метеовеличин – приземной температуры воздуха в °С на высоте 2 м (T), давление на уровне моря в гПа (P), температура воздуха на изобарической поверхности 700 гПа (T700), высота изобарической поверхности 500 гПа в метрах (H500). Прогностические величины являются средними для стандартных сезонов: зима, весна, лето, осень. Для каждой из этих величин имеется по шесть членов ансамбля, которые получены за счет интегрирования при разных начальных условиях со сдвигом на шесть часов (6 ч, 12 ч, 18 ч, …... 36 ч). В тексте члены ансамбля будут обозначаться соответственно как первый, второй, ….. , шестой члены ансамбля (01, 02, 03, 04, 05, 06).

Для России были выбраны 28 объектов, для которых предполагалось проводить ДС. В их числе наиболее крупные города России и отдельные важные экономические объекты. Обязательным условием проведения статистического ДС является наличие длинных и надежных рядов наблюдений. В данной работе мы демонстрируем результаты ДС для точки с координатами г. Москвы. В качестве эмпирических были использованы данные синоптической станции ВДНХ в г. Москве.

Все ряды сезонных прогнозов температуры билинейно интерполировались в точку с координатами г. Москвы (интерполяция производилась по каждому члену ансамбля).

2.2. Описание используемого метода ДС–метода множественной регрессии

На основе проведенного анализа методов ДС на данном этапе, был выбран один из наиболее простых и эффективных методов – метод множественной регрессии.

Суть разработанного алгоритма состоит в расчете регрессионной модели для выбранной области по зависимой выборке и использовании выбранной регрессионной модели для описания прогноза приземной температуры в регионе.

Один из наиболее распространенных подходов к оцениванию параметров регрессионной модели, который применялся в работе [10, 12] заключается в следующем: вводится мера рассогласования отклика и регрессионной функции, и оценки параметров регрессии определяются так, чтобы сделать это рассогласование наименьшим. Достаточно простые расчетные формулы для оценок получают при выборе критерия в виде суммы квадратов отклонений значений отклика от значений регрессионной функции (сумма квадратов остатков):

$$Q\ (\beta_0,\ \beta_1) = \sum_{i=1}^{n} \xi^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \to \min_{\beta_0,\ \beta_1}. \tag{1}$$

Задача минимизации квадратичного критерия (1) носит название метода наименьших квадратов (МНК).

В работе применялась простая линейная и множественная линейная регрессии.

$$Y=A_0+A_1X_1+A_2X_2+A_3X_3+\ldots+A_nX_n+\xi \tag{2}$$

Уравнение (2) представляет собой уравнение регрессии, где X_1, …, X_n являются предикторами, Y – предиктантом, $A_0,\ldots, A_n$, – коэффициентами, ξ – погрешность.

При оценке уравнений регрессии использовался метод корреляционного анализа. Применение методов корреляционного анализа в задачах регрессионного анализа позволяет произвести отбор влияющих факторов и выявить среди них наиболее информативные.

В качестве оценок рядов рассчитывались:

- коэффициенты корреляции;
- абсолютная ошибка:

$$A = \frac{\sum_{1}^{N}\left(T - T_Э\right)}{N},$$

где $T_Э$ – эмпирическая (наблюдаемая) температура, T – прогностическая температура или температура, полученная по уравнению регрессии, N – длина ряда;

– нормированная ошибка:

$$E = \sqrt{\frac{\sum_{1}^{N}\left(T - T_Э\right)^2}{D}},$$

где D – дисперсия $T_Э$;

– средняя квадратическая ошибка:

$$RMSE = \sqrt{\frac{\sum_{1}^{N}\left(T - T_Э\right)^2}{N}}\ .$$

В качестве потенциальных предикторов использовались приземная температура (T), давление на уровне моря (P), температура воздуха на изобарической поверхности 700 гПа (T700), высота изобарической поверхности 500 гПа в метрах (H500). Для каждой из этих величин имелось по шесть членов ансамбля, которые получены путём расчета сезонных прогнозов при разных начальных данных.

Кроме ансамбля, связанного со сдвигом начальных данных, в нашем распоряжении имелись результаты второй канадской модели и, таким образом, мы имели не только результаты ансамблевых расчетов, но и результаты суперансамбля.

Весь ряд прогнозов был разделен на два временных интервала: 15-летний интервал (1969-1983 гг.) и 10-летний интервал (1984-1993 гг.); интервал I был использован как обучающий (зависимая выборка), а II – как интервал для тестирования (независимая выборка).

2.3. Статистический отбор предикторов

Для построения оптимальных уравнений регрессии использовалась процедура пошагового отбора наиболее информативных предикторов. Критериями отбора на каждом шаге являлись: множественный коэффициент корреляции, его уточненное значение с учетом числа степеней свободы и абсолютная ошибка. Использование пошаговой схемы дало возможность отобрать для каждого уравнения регрессии не более двух независимых предикторов, при которых множественный коэффициент корреляции достигал статистически значимого максимума, а абсолютная ошибка была минимальной. В результате, для каждого сезона по данным модели AGCM2 были построены уравнения регрессии, связывающие прогноз приземной температуры воздуха в точке с отобранными параметрами.

Опишем метод статистического отбора более подробно. Отбор предикторов осуществлялся следующим образом. На начальном этапе для зависимой выборки был найден парный коэффициент корреляции (R) эмпирической температуры ($T_Э$) с каждым из шести членов ансамбля метеовеличин T, P, T700, H500, а также со средним значением по каждому ансамблю. По каждой величине (T, P, T700, H500) получилось по семь значений парных коэффициентов корреляции (например, для T – $R_{T_Э/T_{01}}$, $R_{T_Э/T_{02}}$, $R_{T_Э/T_{03}}$, $R_{T_Э/T_{04}}$, $R_{T_Э/T_{05}}$, $R_{T_Э/T_{06}}$, $R_{T_Э/T_{ср}}$, аналогично и для других величин). По всем четырем сезонам из каждого ансамбля отбирался наилучший коэффициент корреляции с эмпирической температурой (результаты представлены в табл. 1).

Таблица 1

Парные коэффициенты корреляции между эмпирической температурой и членом ансамбля сезонных прогнозов приземной температуры, давления на уровне моря, температуры воздуха на изобарической поверхности 700 гПа и высоты изобарической поверхности 500 гПа по гидродинамической модели AGCM2 для точки с координатами г. Москвы (зависимая выборка – I). Представлен член ансамбля с максимальной корреляцией.

№ п/п	Сезон							
	весенний		летний		осенний		зимний	
1	$R_{T_Э/T_{06}}$	0,67	$R_{T_Э/P_{01}}$	0,55	$R_{T_Э/T_{01}}$	-0,59	$R_{T_Э/T700_{04}}$	-0,63
2	$R_{T_Э/P_{01}}$	0,55	$R_{T_Э/T700_{05}}$	0,50	$R_{T_Э/T700_{01}}$	-0,43	$R_{T_Э/T_{04}}$	-0,55

3	$R_{T_Э/T700_{06}}$	0,43	$R_{T_Э/H_{05}}$	0,49	$R_{T_Э/P_{05}}$	0,37	$R_{T_Э/H_{CP}}$	-0,54
4	$R_{T_Э/H_{01}}$	0,41	$R_{T_Э/T_{05}}$	0,42	$R_{T_Э/H_{06}}$	-0,19	$R_{T_Э/P_{06}}$	-0,48

Затем для каждого сезона из коэффициентов корреляции составлялись четыре блока (блок состоит из коэффициентов корреляции между $T_Э$ и членом ансамбля метеовеличины); первый блок для какого-либо сезона состоит из корреляций метеовеличины, которая соответствует № 1 из табл. 1, второй блок - № 2 из табл. 1 и т.д.; первым членом блока оказывался член ансамбля с наибольшей корреляцией (блоки по каждому сезону в табл. 1 выделены жирной рамкой). Далее ранжирование этих четырех блоков производилось от максимального к минимальному R. Приведём пример ранжирования летнего сезона: первый блок будет состоять из коэффициентов корреляции $T_Э$ с P, второй блок – из коэффициентов корреляции $T_Э$ с T700, третий блок – из коэффициентов корреляции $T_Э$ с H и четвёртый блок – из коэффициентов корреляции $T_Э$ с T. В каждый блок отбирались не все семь парных коэффициентов корреляции, а только те, которые имели с максимальным R один и тот же знак и наибольшие по абсолютной величине (см. табл. 2).

Таблица 2

Парные коэффициенты корреляции между эмпирической температурой и членами ансамбля сезонных прогнозов давления на уровне моря (I блок), температуры воздуха на изобарической поверхности 700 гПа (II блок), высоты изобарической поверхности 500 гПа (III блок) и приземной температуры (IV блок) по гидродинамической модели AGCM2 для точки с координатами г. Москвы (зависимая выборка – I; летний сезон).

I блок		II блок		III блок		IV блок	
элемент	R	элемент	R	элемент	R	элемент	R
P_{01}	0,55	$T700_{05}$	0,50	H_{05}	0,49	T_{05}	0,42
P_{CP}	0,36	$T700_{CP}$	0,33	H_{01}	0,46	T_{CP}	0,28
P_{05}	0,34	$T700_{04}$	0,24	H_{CP}	0,44	T_{02}	0,26
P_{04}	0,32	$T700_{02}$	0,19	H_{04}	0,34	T_{06}	0,14
P_{06}	0,11	$T700_{06}$	0,16	H_{06}	0,18	–	–
P_{03}	0,09	–	–	H_{02}	0,14	–	–
–	–	–	–	H_{03}	0,05	–	–

Примечание. В таблице указан элемент, с которым рассчитывался коэффициент корреляции с $T_Э$.

Для дальнейшего отбора была применена следующая схема (на примере летнего сезона). Из каждого блока (летнего сезона) был взят элемент с максимальной корреляцией, а из третьего блока были взяты три первых элемента с максимальной корреляцией, т.к. её значение убывает медленно (т.е. были отобраны следующие элементы: P_{01}, $T700_{05}$, H_{05}, H_{01}, H_{CP}, T_{05}). Затем из этих элементов отбирались два «наилучших» предикторов. P_{01} – фиксированный элемент с наибольшей корреляцией из всех ансамблей четырех метеовеличин. Для отыскивания второго предиктора из вышеупомянутых элементов было построено шесть статистических регрессионных моделей: с одним предиктором (P_{01}), с двумя предикторами (P_{01} и $T700_{05}$), с тремя (P_{01}, $T700_{05}$, H_{05}), ….. , с шестью предикторами (P_{01}, $T700_{05}$, H_{05}, H_{01}, H_{CP}, T_{05}). По результатам этих статистических моделей были рассчитаны множественные коэффициенты корреляции (R_M), их уточненные значения с учетом числа степеней свободы (R_A) и средние абсолютные ошибки результатов регрессии (A). Оценки этих шести регрессионных моделей представлены на рис. 1. По рис. 1 видно, что при добавлении в регрессионную модель второго предиктора (II) увеличивается уточнённый коэффициент корреляции (R_A) и уменьшается средняя абсолютная ошибка. Итак, для дальнейшего анализа, в качестве предикторов отбираем P_{01} и $T700_{05}$.

Аналогичным образом предикторы отбирались для всех остальных сезонов.

Для каждого сезона отбирались два «наилучших» предиктора, затем эти отобранные предикторы применялись ко всем сезонам. Таким образом, к каждому сезону было применено четыре регрессионные модели. Наиболее устойчивый результат имели регрессионные модели с предикторами P_{01} и $T700_{04}$.

Далее в тексте P_{01} и $T700_{04}$ будут указываться без индексов как P и T700.

Для каждого сезона с помощью стандартных пакетов получены следующие уравнения регрессии:

a) для весеннего сезона: $T = 5{,}18 - 0{,}03 \cdot T700$, $T = 10{,}08 - 0{,}25 \cdot P$, $T = 13{,}73 + 0{,}32 \cdot T700 - 0{,}28 \cdot P$;

b) для летнего сезона: $T = 16{,}27 + 0{,}55 \cdot T700$, $T = 11{,}02 + 0{,}60 \cdot P$, $T = 10{,}93 + 0{,}30 \cdot T700 + 0{,}57 \cdot P$;

c) для осеннего сезона: $T = 2{,}65 - 0{,}38 \cdot T700$, $T = 3{,}29 + 0{,}12 \cdot P$, $T = 0{,}73 - 0{,}39 \cdot T700 + 0{,}12 \cdot P$;

d) для зимнего сезона: $T = -23{,}26 - 1{,}15 \cdot T700$, $T = -1{,}90 - 0{,}31 \cdot P$, $T = -18{,}32 - 1{,}05 \cdot T700 - 0{,}21 \cdot P$.

Были рассчитаны оценки температуры, полученной по уравнениям регрессии. В табл. 3 приведены результаты расчетов по различным уравнениям регрессии с одним предиктором – температурой на уровне 700 гПа (T7) или давлением на уровне моря (P), а также по уравнению регрессии с двумя предикторами (T7, P) для всех сезонов.

Из табл. 3 видно, что для всех сезонов максимальный коэффициент корреляции для прогнозов температуры по модели AGCM2 имеет уравнение регрессии с двумя предикторами (T700 и P). Для второй прогностической модели этот результат не всегда подтверждается – это связано, возможно, с тем, что уравнения регрессии строились по результатам первой модели. Заметно улучшаются и остальные оценки при учете двух предикторов.

Таблица 3

Оценки расчетов приземной температуры воздуха по различным уравнениям регрессии, построенные по данным двух гидродинамических моделей для точки с координатами г. Москвы.

(зависимая выборка – I)

Сезон	Весенний			Летний			Осенний			Зимний		
Регрессионная модель	**T7**	**P**	**T7, P**	**T7**	**P**	**T7, P**	**T7**	**P**	**T7, P**	**T7**	**P**	**T7, P**
C (AGCM2)	0,02	0,55	0,59	0,24	0,55	0,57	0,18	0,27	0,32	0,63	0,39	0,68
C (SEF)	-0,03	0,30	0,31	-0,11	0,28	0,25	0,01	0,13	0,07	0,02	0,03	0,04
A (AGCM2), °C	1,28	1,10	1,03	0,93	0,74	0,80	0,79	0,80	0,81	1,30	1,68	1,22
A (SEF), °C	1,29	2,18	1,84	1,89	2,21	1,44	1,24	0,93	1,05	2,66	2,87	3,68
E (AGCM2), °C	3,87	3,23	3,14	3,76	3,23	3,19	3,81	3,73	3,66	3,00	3,56	2,83
E (SEF), °C	3,88	5,91	5,23	6,43	6,73	4,69	4,95	4,04	4,50	5,59	6,54	8,07
RMSE (AGCM2), °C	1,62	1,35	1,32	1,36	1,17	1,15	1,14	1,12	1,10	1,63	1,94	1,54
RMSE (SEF), °C	1,99	3,03	2,68	2,85	2,98	2,08	1,81	1,48	1,65	3,72	4,36	5,38

Примечание. C – коэффициент корреляции между эмпирической и прогностической температурой; A – средняя абсолютная ошибка ряда прогностической температуры; E – нормированная ошибка; RMSE – средняя квадратическая ошибка.

Таким образом, в результате использованной процедуры для уравнения регрессии были отобраны следующие предикторы: температура на уровне 700 гПа и давление на уровне моря, т.к. они имели наиболее устойчивое влияние для всех сезонов. Малоинформативными оказались поля высоты поверхности 500 гПа и поле самой приземной температуры.

На основании анализа данных для зависимого периода в дальнейшем (при проведения ДС для независимого периода) использовались уравнения множественной регрессии с учетом двух предикторов – температуры воздуха на уровне 700 гПа и приземного давления P.

3. Анализ численных сезонных прогнозов приземной температуры воздуха

Прежде чем описывать алгоритм и результаты ДС, приведем оценки рядов сезонных прогнозов для точки с координатами г. Москвы по первому 15-летнему интервалу (1969-1983 гг.). Для этого были рассчитаны коэффициенты корреляции между рядами эмпирической температуры и каждым членом ансамбля прогностической температуры обеих моделей. Из шести коэффициентов корреляции для двух моделей был выбран максимальный (C_M). Кроме того, все шесть членов ансамбля для каждого сезона осреднялись с равными весами, и рассчитывался коэффициент корреляции эмпирической температуры со средним по ансамблю прогнозом (C_A). Эта операция проводилась для каждой модели. Результаты приведены в табл. 4. В этой же таблице представлены средние абсолютные ошибки прогностической температуры для члена ансамбля, имеющего максимальную корреляцию с эмпирической температурой (A_M), а также для этого же члена ансамбля представлены нормированные ошибки (E_M) и RMSE ($RMSE_M$). Те же характеристики приведены для среднего по ансамблю прогноза (A_A, E_A, $RMSE_A$). Из анализа таблицы видно, что наиболее высокие корреляции прогноза с наблюдениями по расчетам с помощью обеих моделей были в весенний период. Причем в осенний и зимний сезоны по обеим моделям корреляция отрицательная. Видно также, что для обеих моделей и для всех сезонов находился член ансамбля с более высокой корреляцией, чем средняя по ансамблю.

Таблица 4

Оценки сезонных прогнозов приземной температуры по двум гидродинамическим моделям для точки с координатами г. Москвы (зависимая выборка – I).

Сезон	Весенний		Летний		Осенний		Зимний	
Модель	AGCM2	SEF	AGCM2	SEF	AGCM2	SEF	AGCM2	SEF
C_M	0,67	0,72	0,42	0,55	-0,59	-0,36	-0,55	-0,46
C_A	0,38	0,62	0,28	0,43	-0,50	-0,08	-0,32	-0,31
A_M, °C	6,71	3,25	1,38	1,93	1,28	1,01	4,70	3,34
A_A, °C	7,73	3,29	1,28	2,11	1,14	0,96	4,38	2,30
E_M, °C	16,37	8,25	4,41	6,26	5,81	5,08	10,52	6,86
E_A, °C	19,09	8,51	4,16	6,83	4,92	4,25	9,30	4,82
$RMSE_M$, °C	6,85	3,45	1,60	2,26	1,47	1,52	5,72	3,73
$RMSE_A$, °C	7,99	3,56	1,50	2,47	1,47	1,27	5,06	2,62

Из сравнения табл. 3 и 4 видно также, что температура, вычисленная по уравнениям регрессии для всех сезонов, имеет заметно меньшую среднюю абсолютную ошибку, чем по результатам модельных прогнозов. Особенно значимы изменения для весеннего и зимнего сезонов: ошибка снизилась в 4 и 5 раз соответственно. В целом коэффициент корреляции вычисленной температуры, полученной по уравнению множественной регрессии, выше, чем для модельной для всех сезонов, кроме весеннего. Причем для осени и зимы коэффициент корреляции сменил знак на положительный (см. табл. 3 и табл. 4).

Наибольшие абсолютные и нормированные ошибки прогноза наблюдались также в весенний и летний сезоны, что связано с малой величиной дисперсии в эти периоды. Ошибки обеих моделей близки друг к другу. Отметим также, что в отличие от коэффициентов корреляции, ошибки средних по ансамблю для ряда сезонов (например, для лета) ниже, чем для лучших по коэффициенту корреляции прогнозов.

4. Проверка результатов ДС на независимом интервале II

Прежде чем анализировать результаты ДС на независимой выборке, рассмотрим оценки сезонных прогнозов по обеим моделям в точке с координатами в г. Москва (табл. 5).

Таблица 5

Оценки сезонных прогнозов приземной температуры в точке с координатами г. Москвы по двум гидродинамическим моделям (независимая выборка – II).

Сезон	Весенний		Летний		Осенний		Зимний	
Модель	AGCM2	SEF	AGCM2	SEF	AGCM2	SEF	AGCM2	SEF
C_M	0,44	0,42	0,52	0,62	0,67	-0,37	0,58	-0,58
C_A	0,28	0,07	0,45	0,42	0,11	0,31	0,36	0,15
A_M, °C	8,41	3,63	1,24	1,92	1,32	1,12	5,32	3,46
A_A, °C	7,83	3,67	1,06	1,94	1,29	0,82	5,54	2,68
E_M, °C	24,37	10,78	4,42	6,11	4,01	3,71	6,93	4,27
E_A, °C	22,28	10,86	11,97	6,31	4,35	3,22	4,18	3,51
$RMSE_M$, °C	8,73	3,86	1,25	2,12	1,65	1,53	4,77	3,59
$RMSE_A$, °C	7,98	3,89	3,39	2,19	1,79	1,32	2,88	2,96

Примечание. Обозначения те же, что и в табл. 4.

Видно, что оценки по этому ряду количественно несколько отличаются от оценок, приведенных в табл. 4, но они достаточно близки. Первая модель (AGCM2) позволила получить прогнозы с более высокими коэффициентами корреляции для всех сезонов, кроме весны, с другой стороны, в весенний период остальные оценки во второй модели выше, чем в первой. Это связано с меньшими систематическими ошибками во второй модели, как видно на рис. 1-3, к анализу которых обратимся ниже.

В табл. 6 приведены основные результаты ДС с помощью уравнения множественной регрессии. В весенний и летние сезоны коэффициенты корреляции расчетов по уравнению регрессии с наблюдениями меньше, чем были в прогнозах, но значительно уменьшились (в 2-3 раза) ошибки прогноза. В осенне-зимний периоды коэффициенты корреляции для первой модели (AGCM2) также заметно уменьшились, а для второй модели (SEF) увеличились, одновременно увеличились ошибки прогнозов.

Таблица 6

Оценки расчетов ДС приземной температуры воздуха в точке с координатами г. Москвы (независимая выборка – II)

Сезон	Весенний		Летний		Осенний		Зимний	
Модель	AGCM2	SEF	AGCM2	SEF	Модель	AGCM 2	SEF	AGCM 2

C	0,35	0,06	0,51	0,30	-0,04	0,36	-0,54	0,35
A	1,32	1,21	0,94	1,48	1,11	1,54	3,11	3,22
E	4,02	4,25	3,29	5,18	4,11	4,70	4,14	4,78
RMSE	1,44	1,52	1,14	1,80	1,69	1,93	3,49	4,03

Примечание. Обозначения те же, что и в табл. 5

В табл. 7 представлены отношения оценок сезонных прогнозов (из табл. 5) к оценкам расчетов ДС приземной температуры воздуха (из табл. 6). Например, C_M из табл. 7 соответствует отношению C_M (из табл. 5) к C (из табл. 6), а C_A из табл. 7 – C_A (из табл. 5) к C (из табл. 6), остальные показатели в табл. 7 – соответственно.

Таблица 7

Отношения оценок сезонных прогнозов к оценкам расчетов ДС приземной температуры воздуха (независимая выборка – II) в точке с координатами г. Москвы по двум гидродинамическим моделям.

Сезон	Весенний		Летний		Осенний		Зимний	
Модель	AGCM2	SEF	AGCM2	SEF	AGCM2	SEF	AGCM2	SEF
C_M	1,26	7,00	1,02	2,07	16,75	1,03	1,07	1,66
C_A	0,80	1,17	0,88	1,40	2,75	0,86	0,67	0,43
A_M	6,37	3,00	1,32	1,30	1,19	0,73	1,71	1,07
A_A	5,93	3,03	1,13	1,31	1,16	0,53	1,78	0,83
E_M	6,06	2,54	1,34	1,18	0,98	0,79	1,67	0,89
E_A	5,54	2,56	3,64	1,22	1,06	0,69	1,01	0,73
$RMSE_M$	6,06	2,54	1,10	1,18	0,98	0,79	1,37	0,89
$RMSE_A$	5,54	2,56	2,97	1,22	1,06	0,68	0,83	0,73

Примечание. Обозначения те же, что и в табл. 5.

Из табл. 7 видно, что оценки A_M, A_A, E_M, E_A, E_M, $RMSE_M$, $RMSE_A$ удалось улучшить: ошибки расчетов ДС меньше ошибок сезонных прогнозов практически всегда (отношение больше 1), кроме осеннего сезона (особенно для модели SEF), также не улучшился результат для зимнего сезона, полученного при помощи данных модели SEF (однако, отношение близко к 1). Для коэффициентов корреляций наоборот: если отношение меньше 1, то оценки расчетов ДС удалось улучшить. В расчетах ДС в большинстве случаев коэффициент корреляции ухудшался.

Для наглядной демонстрации результатов ДС рассмотрим рис. 2-5.

На рис. 2-5 представлены прогнозы приземной температуры на независимом периоде для всех четырёх сезонов. На каждом рисунке имеется пять графиков: ход эмпирической среднесезонной приземной температуры воздуха, интерполированных модельных прогнозов температуры (взят член ансамбля с максимальной корреляцией) и результаты расчета по уравнению регрессии для двух моделей.

Модельные данные для весеннего сезона (рис. 2) систематически занижают температуру в точке с координатами г. Москвы. Кривые хода температуры, полученные по уравнениям регрессии, не имеют систематического сдвига, но не для всех лет описывают ход температуры достаточно близко.

В летний сезон (рис. 3) видно, что модель AGCM2 систематически завышает температуру, а модель SEF – занижает; наиболее близка к данным наблюдений температура, восстановленная по уравнению регрессии для первой модели. На кривой хода температур, полученной по уравнению регрессии для второй модели, в 1986 и 1992 гг. наблюдаются заметные выбросы, которые могли исказить оценки результатов.

На рис. 4 и 5 приведены те же характеристики для осени и зимы. По сути, все кривые для осени за большую часть лет лежат выше кривой реальной температуры.

ЗАКЛЮЧЕНИЕ

1. В работе приведен аналитический обзор по статистическим методам детализации численного прогноза (ДС).
2. Разработан, реализован и испытан алгоритм детализации температуры приземного воздуха, основанный на методе множественной регрессии для точки с координатами г. Москвы.
3. Испытание ансамблей прогноза приземной температуры по двум гидродинамическим моделям Канадского Климатического центра показало, что с помощью этого алгоритма для ряда сезонов удаётся заметно уменьшить систематические ошибки численного гидродинамического прогноза приземной температуры.

Авторы выражают благодарность В.В. Оганесяну за предоставленный в их распоряжение обзор методов статистической детализации прогноза.

Работа выполнена при частичной поддержке грантами РФФИ N 03-05-64312, 04-05-64151, 04-05-65099, 05-05-08018_ОФИ_а, а также гранта INTAS - 03-51-5296.

Список литературы

1. Багров Н.А. Аналитическое представление метеорологических полей посредством естественных ортогональных составляющих // Труды ЦИП.–1959.–Вып.74.–С. 3-24.
2. Багров Н.А. Аналогичность метеорологических полей и оценка прогнозов // Труды ЦИП. – 1959. – Вып. 74. – С. 56-68.
3. Багров Н.А. О некоторых вопросах подыскания аналога для данного образа // Труды ГМЦ СССР. – 1973. – Вып. 106. – С. 3-12.
4. Багров Н.А. О канонической корреляции векторов // Труды ГМЦ СССР. – 1980. –Вып. 226. – С. 3-9.
5. Багров Н.А., Оганесян В.В. К вопросу о тепловом взаимодействии океана и атмосферы // Труды ГМЦ СССР. – 1978. – Вып. 195. – С. 57-62.
6. Груза Г.В. Прогностические модели в метеорологии и статистические прогнозы // Труды ВНИГМИ-МЦД. – 1977. – Вып. 35. – С. 3-10.
7. Груза Г.В., Клещенко Л.К., Ранькова Э.Я. О методе метеорологической интерпретации численного долгосрочного прогноза // Труды ВНИГМИ-МЦД. – 1977. – Вып. 58. – С. 3-20.
8. Груза Г.В., Ранькова Э.Я. Метод статистического прогноза погоды на основе динамической климатологии // Труды III Всес. симпоз. по применению статистических методов в метеорологии. – 1978. – С. 15-26.
9. Дмитриев Е.В., Рубинштейн К.Г., Чавро А.И. Детализация крупномасштабного поля приземной температуры для Московского региона // Метеорология и гидрология. – 2003. – № 7. – С. 19-29.
10. Езекиэл М., Фокс К.А. Методы анализа корреляций и регрессий линейных и криволинейных // Из-во «Статистика». – Москва. – 1966. – 558с.
11. Оганесян В.В. Каноническая корреляция температурных полей атмосферы и океана // Метеорология и гидрология. – 1978. – № 2. – С. 45-51.
12. Пановский Г.А., Брайер Г.В. Статистические методы в метеорологии // Гидрометеорологическое издательство. Л.: 1967. – 242 с.

13. Рубинштейн К.Г. Обеспечение метеоролгической информацией моделей дальнего переноса примеси в атмосфере // Метеорология и гидрология. – 2002. – № 9. – С. 17-30.
14. Сонечкин Д.М. Об объективной классификации метеорологических явлений и ситуаций с помощью ЭВМ // Метеорология и гидрология. – 1968. – № 5. – С. 12-22.
15. Чавро А.И., Дмитриев Е.В. Статистическая модель восстановления региональной структуры геофизических полей // Метеорология и гидрология. – 2002. – № 6. – С. 39-49.
16. Школьник И.М. О моделировании климата на ограниченной территории // Труды ГГО. – 2001. – Вып. 550. – С. 110-127.
17. Barnett T., and R. Preisendorfer. Multifield analog prediction of short-term climate fluctuations using a climate state vector. // J.Atmos.Sci. 1978. 35, C. 1771-1787.
18. Biau G., E. Zorita H. von Storch, H. Wackernagel. Estimation of Precipitation by Kriging in the EOF Space of the Sea Level Pressure Field. // J. Climate. 1999. vol.12, p. 1070-1085.
19. Bretherton C., C. Smith, and J. Wallace. An intercomparison of methods to find coupled patterns in climate data. // J. Climate. 1992. vol. 5, p. 541-560.
20. Cheng X., and J. M. Wallace. Cluster analyses of the Northern Hemisphere Wintertime 500hPa Heit fields: Spatial patterns. // J. Atmos. Sci. 1993. 50, p. 2647-2696.
21. Cubasch U., H. von Storch, J. Waszkewtz, and E. Zorita. Estimates of climate changes in Southern Europe using different downscaling techniques. // Climate Res. 1996. 7, p. 129-149.
22. Derome J., Brunet G., Plante A., Gagnon N., Boer G.J., Zwiers F.W., Lambert S.J., Sheng J. and Ritchie H. Seasonal Prediction Based on Two Dynamical Models. // Canadian Meteorological and Oceanographic Society. Atmosphere-Ocean. 2001. **39,** No 4, p. 485-501.
23. Elsner J. B., and A. A. Tsonis. Nonlinear predicting, chaos, and noise. // Bull. Amer. Meteor. Soc. 1992. 73, p. 49-60.
24. Gershunov A., T. Barnett, D. R. Cayan, T. Tubbs, L. Goddard. Predicting and Downscaling ENSO Impacts on Intraseasonal Precipitation Statistics in California: The 1997/98 Event. // Journal of Hydrometeorol. 2000. vol. 1, p. 201-210.
25. Glahn H. R. Canonical Correlation and its Relationship to Discriminant Analysis and Multiple Regression. // J. Atmos. Sci. 1968. vol. 25, No 1, p. 23-31.
26. Hewitson B. Climate downscaling: Techniques and application. // Climate Res. 1996. 7, p. 85-95.
27. Hewitson B., and R. G. Crane. Large-scale atmospheric controls on local precipitation in Tropical Mexico. // Geophys. Res. Lett. 1992. 19, p. 1835-1838.
28. Karl T. M., W.-C. Wang, M. E. Schlesinger, D. E. Knight, and D. Portman. A method of relating General Circulation Model simulated climate to observed local climate. Part 1: Seasonal statistics. // J. Climate. 1990. 3, 1 p. 053-1079.
29. Kruizinga S., and A. Murphy. Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands. MWR, 1983. 111, p. 2244-2254.
30. Lau C., and B. Widrow, Eds. Special issue on neural networks. Proc. IEEE. 1990. 78, p. 1411-1414.
31. Lorenz E. N. Atmospheric predictability as revealed by naturally occurring analogs. // J. Atmos. Sci. 1969. 26, p. 639-646.
32. McFarlane N.A., G.J. Boer, J.-P. Blanchet, and M. Lazare. The Canadian Climate Centre second-generation general circulation model and its equilibrium climate. // J. Climate. 1992. 5, p. 1013-1044.
33. Sailor D. Y., and X. Li. A Semiempirical Downscaling Approach for Predicting Regional Temperature Impacts Associated with Climate Change. // Journal of Climate. 1999. 12, p. 103-114.
34. Sinclair M. R. A diagnostic model for estimating orografic precipitation. // J. Appl. Meteor. 1994. 33, p. 1163-1174.
35. Solman S. A., and M. N. Nunez. Local estimates of Global Climate Change: A Statistical Downscaling Approach. Int. // J. Climatol. 1999. 19, p. 835-861.

36. Van den Dool H. Searching for analogs, how long must we wait? Tellus. 1994. 46A, p. 314-324.
37. Von Storch H., E. Zorita, and U. Cubasch. Downscaling of climate change estimates to regional scales: An application to Iberian winter rainfall in winter time. // J. Climate. 1993. 6, p. 1161-1171.
38. Von Storch H. On the Use of “Inflation” in Statistical Downscaling. // Journal of climate. 1999. 12, p. 3505-3506.
39. Zorita E., J. Hughes, D. Lettenmaier, and H. von Storch. Stochastic downscaling of regional circulation patterns for climate model diagnosis and estimation of local precipitation. // J. Climate. 1995. 8, p. 1023-1042.
40. Zorita E., and H. von Storch. The Analog Method as a simple Statistical Downscaling Technique: Comparison with More Complicated Methods. // Journal of Climate. 1999. 12, p. 2474-2489.