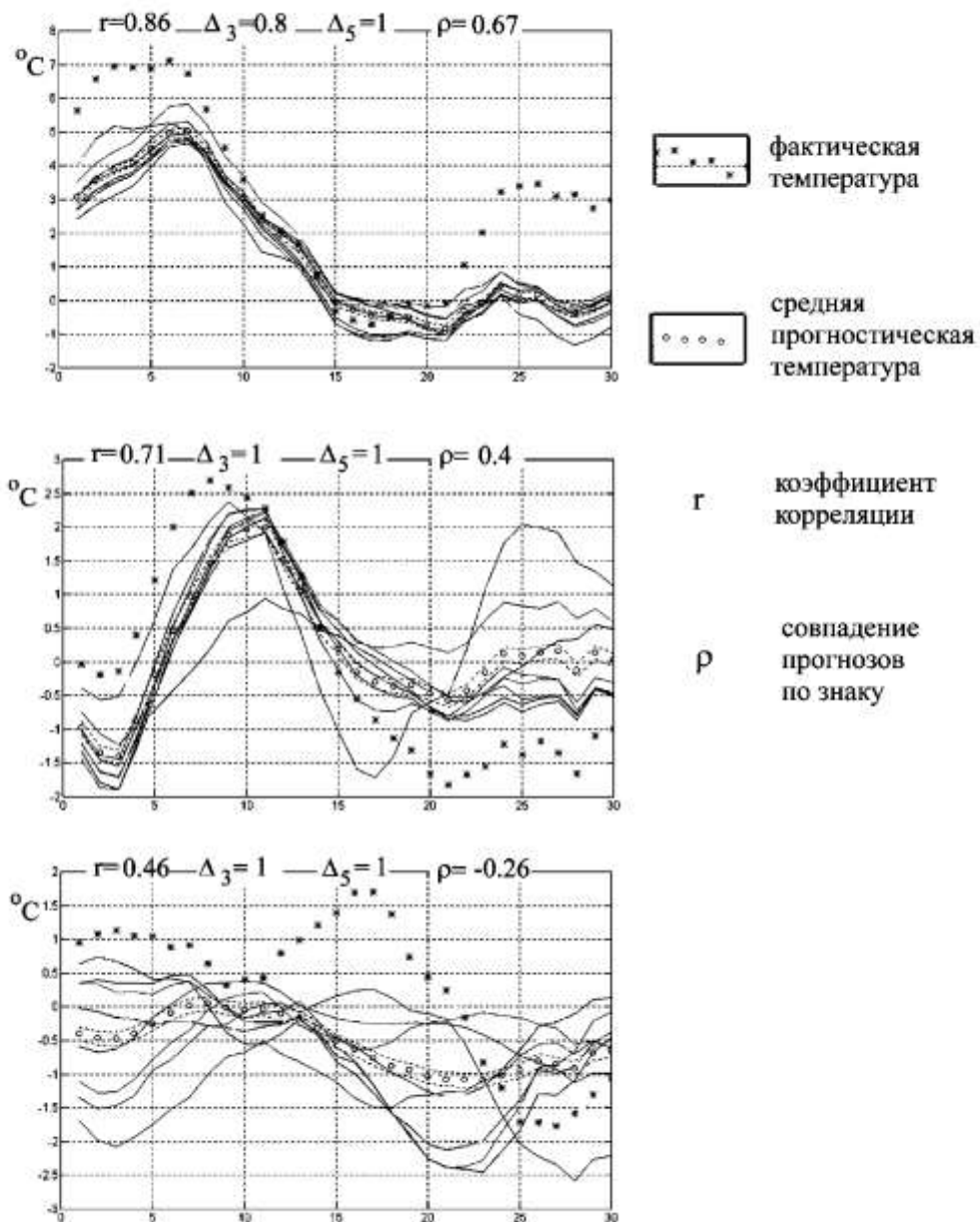


Априорное оценивание качества статистических прогнозов хода приземной температуры на основе анализа состояния циркуляции в средней тропосфере

При проведении работы по созданию схемы прогноза сглаженного хода температуры внутри месяца для различных пунктов на территории России были отмечены случаи, когда прогнозы, рассчитанные по разным версиям схемы, оказывались несогласованными и имели очень низкую оправдываемость. На рис. 1 показаны в качестве наглядного примера случаи удачных, удовлетворительных и несогласованных прогнозов для Москвы. Прогностическая схема [4], использованная в работе, представляет собой кусочно-линейную регрессионную модель, предикторами в которой являются ортогонализированные ансамблевые декадные прогностические поля $H-500$, P_0 и $T-850$, являющиеся выходной продукцией гидродинамических моделей общей циркуляции атмосферы (ОЦА) Гидрометцентра России [5,6] и Национального Центра по Прогнозам Окружающей Среды (НЦПОС) США. Ход приземной температуры (фактической и прогностической) представляется несколькими первыми коэффициентами разложения по естественным ортогональным составляющим (е.о.с.). Выходная продукция предложенной статистической модели представляет собой расширенный ансамбль, реализации которого были получены в результате расчетов модели по различным совокупностям информативных точек предикторов, а также по всем отдельным членам ансамблей входных данных.



Целью настоящего исследования является поиск признаков и "предшественников" развития региональных аномалий, с помощью которых можно было бы рассчитать априорную оценку качества прогноза при статистической интерпретации ансамблевых среднесрочных гидродинамических прогнозов.

Моделирование подобных "индикаторов", предшествующих долгопериодным аномалиям, обычно находится за пределами возможностей современных динамических моделей. Эмпирические методы могут быть более эффективными в этом смысле. С помощью статистического анализа удастся обнаружить дальние связи между предиктантом и предикторами. Главная задача состоит в разработке критерия для априорного предсказания результативности прогноза в зависимости от параметров, характеризующих крупномасштабное состояние атмосферы.

Одним из направлений для выявления причин низкой оправдываемости прогнозов хода температуры внутри месяца, может служить анализ типов объективной классификации метеополей в первой декаде прогностического месяца. Расчеты на архивном материале продемонстрировали, что наиболее информативными для воспроизведения динамики среднесуточной температуры внутри месяца являются сглаженные поля геопотенциала в течение первой декады месяца. По этой причине мы ограничились для анализа оправдываемости прогнозов только рассмотрением полей геопотенциала.

Автоматическая классификация средних декадных полей $H-500$ в каждой декаде года для территории первого естественного синоптического района (е.с.р.) была проведена по методике [2], сочетающей черты алгоритма Ланда и метода k -средних [3]. Классификация представляет собой один из алгоритмов «самообучения», когда ни число классов (типов), ни сами типы заранее не известны и определяются только свойствами рассматриваемых полей. Объекты классификации - среднедекадные поля $H-500$, предварительно разлагались по е.о.с.. Критерием сходства (различия) объектов служила евклидова метрика.

$$D_{\alpha,\beta} = \left\{ \sum_i P_i [(X_i(\alpha) - X_i(\beta))]^2 \right\}^{\frac{1}{2}},$$

где P_i - весовой коэффициент i -го признака, X_i - i -й признак объектов α и β .

Изначально задается число классов, равное k . Далее рассчитывается матрица расстояний между ситуациями $D_{\alpha\beta}$ и выбирается пороговое значение расстояния d , такое, что если расстояние $D_{\alpha\beta} < d$, то ситуации считаются аналогичными. Каждой ситуации (строке матрицы расстояний) присваивается ранг r , который представляет собой число похожих на нее ситуаций. Ситуация с максимальным рангом принимается за эталон первого класса. Эталоном второго класса считается ситуация, наиболее удаленная от первого эталона. Затем вычисляется сумма расстояний каждой ситуации до обоих эталонов и в качестве третьего эталона выбирается объект, максимально удаленный от первых двух. Когда число эталонов становится равным k , проводится новая классификация по минимуму расстояния до эталона, при этом каждый объект относится к тому классу, эталон которого оказывается ближе всего. Далее рассчитываются новые эталоны, представляющие собой средние значения всех элементов класса. Эта процедура повторяется до тех пор, пока очередная итерация не перестанет изменять эталоны и состав классов. Обычно требуется 10-15 итераций. Для определения оптимального числа классов использовался критерий качества классификации, который представляет собой отношение средней внутриклассовой дисперсии к общей дисперсии всей обучающей выборки, и имеет вид:

$$\Delta J = \sum_{j=1}^k m_j \sum p_i \sigma_{ji}^2 / \sum \sigma_i^2,$$

где k - число классов, m_j - частота j -го класса, p_i - вес i -го компонента вектора, σ_i - дисперсия

i -й компоненты всего архива, σ_{ij} - дисперсия i -й компоненты внутри j -го класса. Как показано в [1], точка перегиба кривой указывает на оптимальное число классов.

В результате вычислительных экспериментов выбиралось оптимальное число классов метеорологических полей, которое в большинстве случаев равнялось 12 (рис.2).

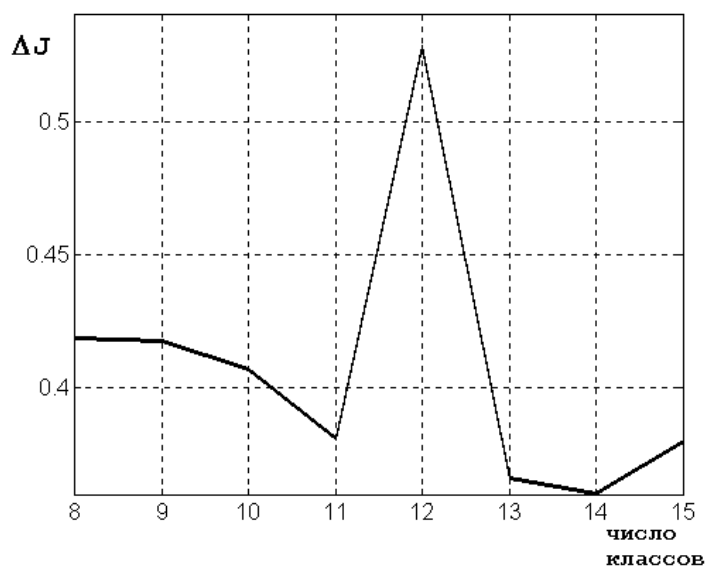


Рис. 2. Изменение величины ΔJ , определяющей оптимальное число классов в зависимости от числа классов.

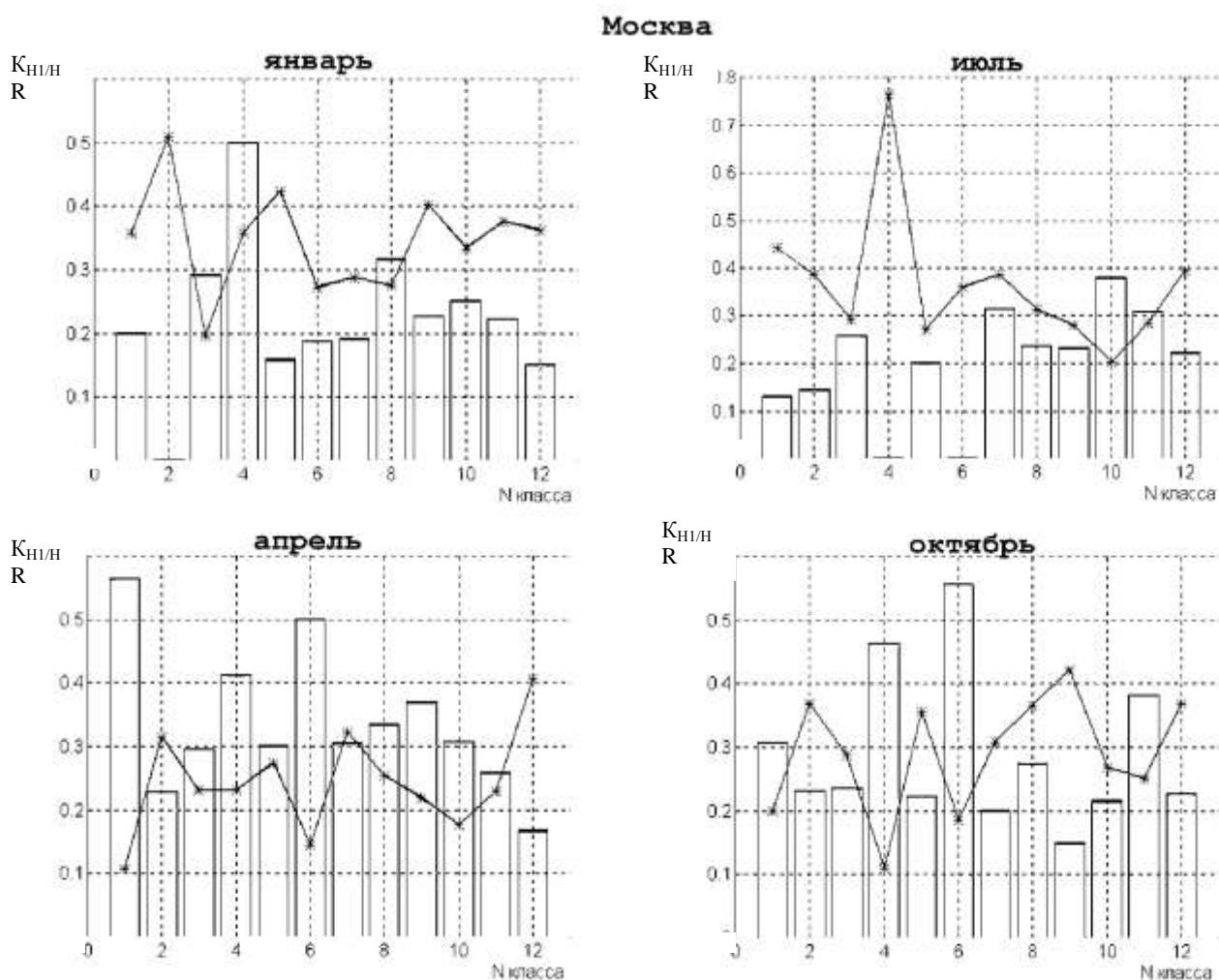


Рис.3. Показатель несогласованности прогнозов ($K_{n1/n}$ - столбики) и средние коэффициенты корреляции R (кривые) прогнозов хода температуры для случаев, попавших в каждый класс $H-500$.

На рис.3 представлено распределение по классам следующих показателей: относительной величины несогласованных прогнозов - отношения числа случаев в каждом классе, когда кривые прогностической температуры для Москвы оказались несогласованными, к числу случаев, попавших в каждый класс (N_{i1}/N_i) и средних по выделенным классам коэффициентов корреляции ($r_{кл}$) между ходом прогностической и фактической температуры в Москве. Видно, что для каждого представленного месяца минимальное значение $r_{кл}$ соответствует классу, в котором число несогласованных прогнозов максимально (по относительной величине). Следовательно, при определенном типе (типах) циркуляции в средней тропосфере предложенная прогностическая схема должна быть изменена.

Обнаружено, что если циркуляция в средней тропосфере соответствует выявленному таким образом одному из классов, априорная вероятность неудачного прогноза составляет 50-80%. При других типах циркуляции априорная вероятность неудачного прогноза составляет лишь 10-40%.

Выводы, полученные на архивной выборке за 1949-1994 гг., подтверждались на независимом материале за 1997-2000 гг. Анализ результатов на независимом материале

показал, что выявленные классы полей, при которых повторяемость несогласованных прогнозов максимальна, соответствуют классам на обучающей выборке.

Кроме того, проводилась валидация прогностической схемы на текущих ансамблевых данных (НЦПОС) за период с 1 октября 2000 г. по 1 ноября 2001 г. для города Москвы. Конечно, объем последней выборки недостаточен для получения устойчивых результатов, однако эти результаты представляют определенный интерес. Попытки выделения наиболее информативных составляющих в ансамблевом пучке выходных данных оказались малоуспешными, поэтому целесообразней использовать осредненные значения прогностических данных по всем реализациям ансамбля. В среднем, коэффициент корреляции между фактическими и прогностическими данными составил 0.31, а параметр c - 0.25. Ниже приведена информация (табл.1 и табл.2) о прогностических показателях для декад, соответствующих наилучшим и наихудшим прогнозам.

Таблица 1

Прогнозы с самой низкой оправдываемостью

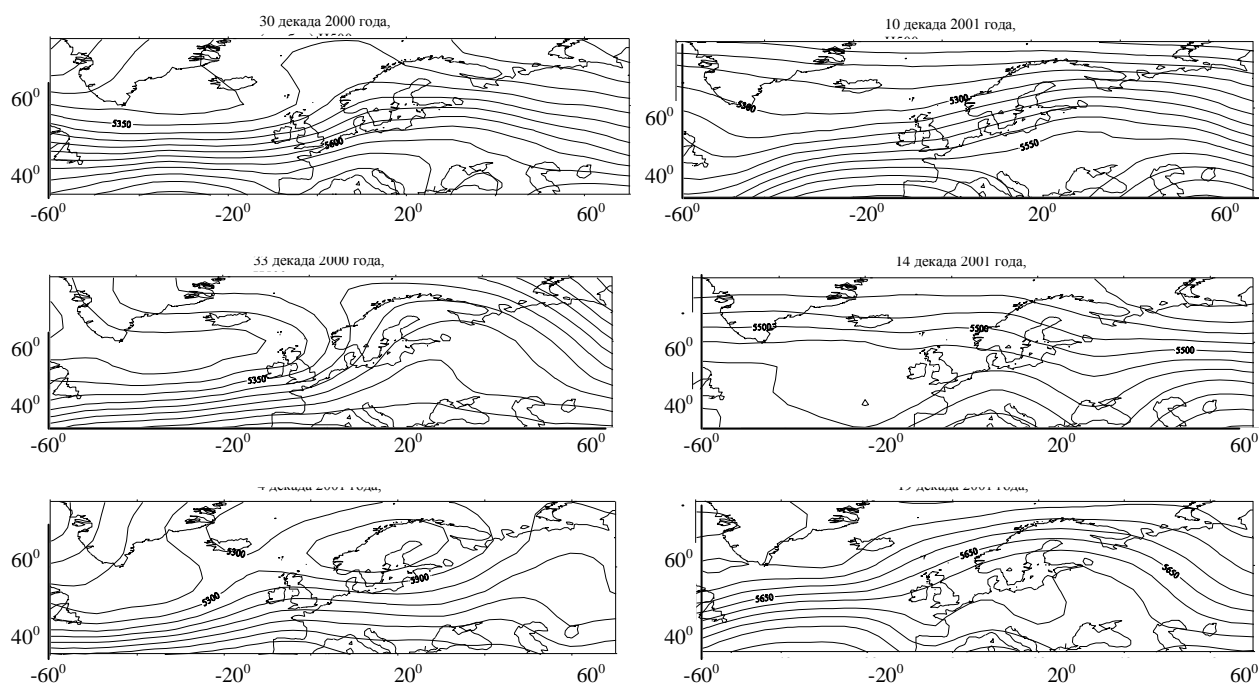
Декада	30- 2000г.	33- 2000г.	4- 2001г.	10- 2001г.	14- 2001г.	19- 2001г.	29- 2001г.
Коэф. кор.	-0.17	-0.25	-0.69	-0.22	-0.04	-0.65	-0.69
c	-0.43	0.00	-0.11	-0.14	-0.29	1	-0.57

Таблица 2

Прогнозы с самой высокой оправдываемостью

Декада	34- 2000г.	7- 2001г.	12-2001г.	13-2001г.	15- 2001г.	23- 2001г.	27- 2001г.
Коэф. кор.	0.65	0.69	0.70	0.70	0.78	0.69	0.85
c	0.64	0.93	0.50	0.36	0.57	0.71	0.79

а)



б)

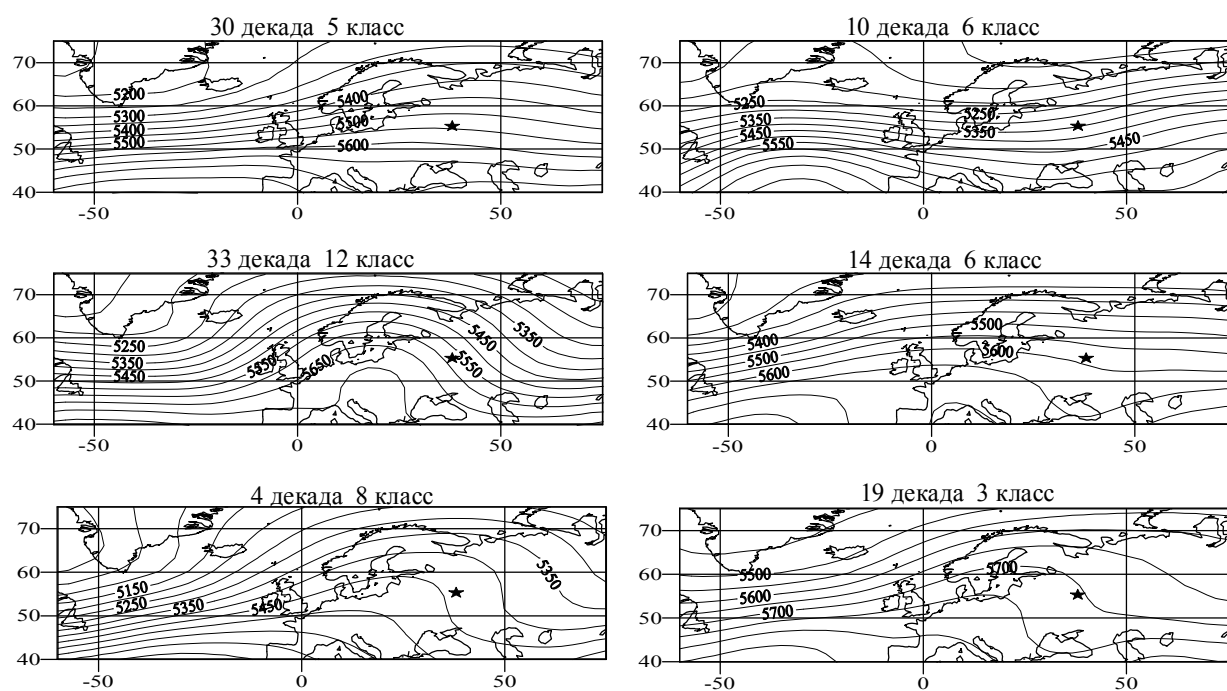
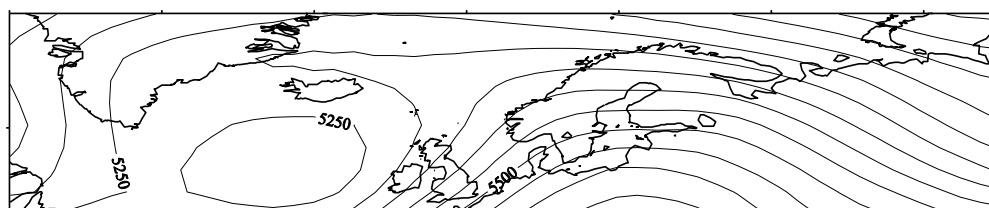
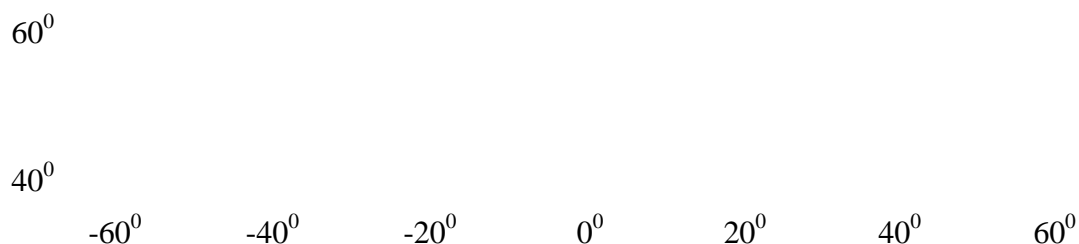
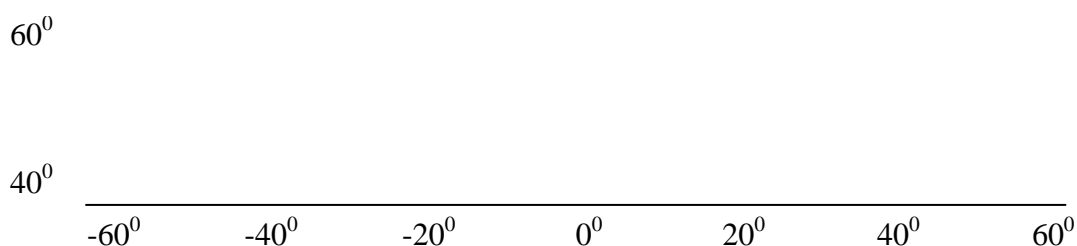
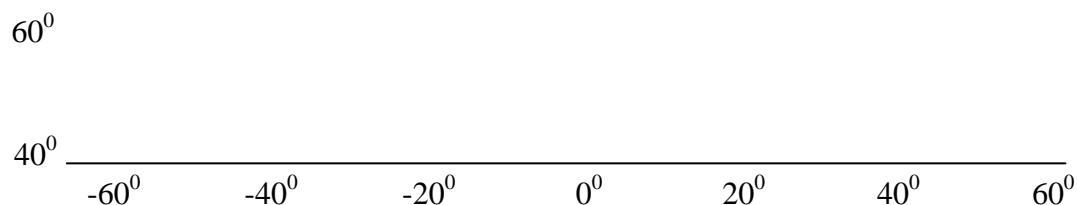


Рис.4. Осредненные поля геопотенциала за декады, соответствующие наименее удачным статистическим прогнозам для Москвы (а). Соответствующие эталонные поля $H-500$ для указанных декад (б).

34 декада 2000 года, $H500$





60°
 5. Осредненные поля геопотенциала за декады, соответствующие наилучшим
 истическим прогнозам для Москвы.

40°
 В качестве примера на рис. 4(а) приведены осредненные поля геопотенциала за декады, соответствующие наилучшим прогнозам (рис. 4(б)), кроме того, приведены стационарные поля геопотенциала, относящиеся к наилучшим прогнозам. Все номера классов этих полей соответствуют полученным ранее на архивном материале классам для несогласованных прогнозов и в большинстве случаев относятся к полям с зональной циркуляцией. Таким образом, можно сделать заключение, что выявленные классы полей, имеющие наибольшую повторяемость, при несогласованности прогностических кривых являются достаточно устойчивыми.

Показано, что в случае блокирующих ситуаций в северной части Атлантики, когда атмосферная ситуация стабильна и развивается во времени относительно медленно, долгосрочные прогнозы имеют самые высокие показатели качества. В качестве примера на

рис.5 приведены осредненные поля геопотенциала за декады с наиболее высокими показателями оправдываемости статистических прогнозов для Москвы.

Результаты проведенного исследования свидетельствуют о том, что для возможности априорной оценки качества прогноза и повышения качества прогнозов имеет смысл использовать классификацию метеорологических полей в средней тропосфере. Планируется для проведения классификации метеорологических полей привлечь большее количество параметров, а также различные статистические подходы.

Работа выполнена при поддержке гранта РФФИ 01-05-64665.

Список литературы

1. Батырева О.В., Вильфанд Р.М., Рудичева Н.И. Прогноз аномалии среднемесячной температуры воздуха с помощью канонической корреляции и автоматической классификации // Метеорология и гидрология. - 1994. - № 12. - С. 34-45.
2. Батырева О.В., Васильев А.А., Вильфанд Р.М., Волобуев Н.М., Воробьев В.П., Тищенко В.А. Автоматическая классификация ежедневных полей геопотенциала // Тр. Гидрометцентра России. - 1999. - Вып. 330. - С. 84-89.
3. Боннер Р.Е. Некоторые виды классификации. В сб.: Автоматический анализ сложных изображений. М.: Мир, 1969. С. 205-235.
4. Тищенко В.А. Развитие схемы прогноза хода температуры в течение месяца на территории России // Тр. Гидрометцентра России. - 2000. - Вып. 333. - С. 53-60.
5. Филатов А.Н., Муравьев А.В., Реснянский Ю.Д. Долгосрочный метеорологический прогноз: математические проблемы и возможности гидродинамических моделей. В сб. 70 лет Гидрометцентру России. СПб.: Гидрометеиздат, 1999. С. 141-165.
6. Фролов А.В., Важник А.И. Интегрированная схема гидродинамического среднесрочного прогноза и объективного анализа глобальных метеорологических полей. В сб. 70 лет Гидрометцентру России. СПб.: Гидрометеиздат, 1999. С. 25-43.