

ON GLOBAL WEATHER MODELING

(Empiric approach)

Romanov L.N.

Federal State Budgetary Institution “Siberian Regional Hydrometeorological Research Institute” (FSBI “SibNIGMI”)

Senior Researcher, Novosibirsk

Романов Л.Н.

Сибирский региональный научно-исследовательский
гидрометеорологический институт, *Главный научный сотрудник,*
Новосибирск

Science of Europe vol. 1, No 4(4) (2016)

ГЛОБАЛЬНОЕ МОДЕЛИРОВАНИЕ ПОГОДЫ

(Эмпирический подход)

Abstract

This paper deals with an approach which allows to construct large scale weather models as well as the modals of the global size. The construction of such models is being realized on the basis of average risk minimization method, using regular data. The scheme of representation of the initial data, which allows to obtain stable relations, connecting physical parameters in space and time is .described. The comparison of the suggested approach relative to the hydro dynamical approach is considered. Perspectives of the applying of the approach in practice are discussed. The results of the of the one step prediction of the global temperature are presented.

Аннотация

В работе рассматривается подход, позволяющий строить крупномасштабные, или глобальные модели без привлечения дифференциальных уравнений. Реализация таких моделей осуществляется на основе критерия минимума среднего риска с использованием регулярных данных. Разработана схема представления исходной информации, дающая возможность получать устойчивые соотношения, связывающие параметры атмосферы в пространстве и во времени. Приводится сравнительный анализ нового подхода с гидродинамическим подходом, и обсуждаются перспективы его применения на практике. Описаны некоторые эксперименты по прогнозированию глобальных полей температуры на один шаг вперед.

Key words: average risk, approximation, time series, multidimensional fields, process.

Ключевые слова: средний риск, аппроксимация, временной ряд, многомерные поля, процесс.

ВВЕДЕНИЕ

Современные глобальные модели погоды все без исключения существуют лишь в гидродинамическом исполнении. Это означает, что для решения задачи прогноза, или мониторинга глобальной атмосферы используется система гидродинамических уравнений с последующим применением для ее решения разностных методов. Однако сама исходная система уравнений в частных производных, является некоторой идеализацией процессов, происходящих в природе. Эта идеализация имеет место всякий раз, когда таким величинам как производные, градиенты или плотности, получаемым в результате предельного перехода, приписывается физический смысл. Для упрощения описания окружающего нас физического мира, такая идеализация в рамках гидродинамического подхода представляется достаточно естественной и по сути неизбежной [1].

Кроме того использование дифференциальных уравнений для целей моделирования предполагает известными граничные условия, задаваемые в виде непрерывных функций, которые при моделировании атмосферы, как правило, неизвестны. Преодоление же этих препятствий необходимо связано с издержками, которые в конечном итоге не могут не влиять на результат. При этом требуется переход к дискретной информации и конечно-разностному представлению производных функции, не говоря уже о разработке методов для решения разностных уравнений

Фактически, при создании численных моделей гидродинамического прогноза наблюдается двойной переход в прямом и обратном направлении: сначала мы идеализируем действительность, составляя, таким образом,

дифференциальные уравнения, а затем, отталкиваясь от этих уравнений, переходим к дискретному случаю и рассматриваем уравнения разностные, которые в конечном итоге приводят к алгебраическим. В условиях, когда исходная информация представляет собой дискретные данные, подобный переход может вызывать большого оптимизма. Более того, возникает желание начинать построения сразу, отталкиваясь от алгебраических систем, используя для этого данные, полученные непосредственно в точках наблюдений, а само решение искать уже в непрерывном виде. Тем более на практике решения в виде непрерывных функций наиболее предпочтительны. Возможность создания такой модели позволило бы говорить об общем эмпирическом подходе при моделировании, полностью основанном на данных наблюдений. Такой подход позволил бы при наличии экспериментальной базы моделировать процессы планетарного масштаба и прогнозировать физические поля без привлечения системы уравнений в частных производных. Для реализации такого подхода необходимы два условия: первое, - это механизм восстановления, который позволял бы определять функциональную зависимость по данным, второе, - наличие данных, удовлетворяющих определенным требованиям.

ВОССТАНОВЛЕНИЕ ФУНКЦИЙ ПО ДАННЫМ

Решению задачи восстановления функций по данным уделяется большое место в приложениях. Мы не будем подробно останавливаться на этой задаче, будем лишь предполагать, что если механизм ее решения существует, и если есть данные, удовлетворяющие определенным требованиям, то задача может быть эффективно решена. На практике, чтобы восстановить функцию по данным надо проделать следующие шаги:

- а) сформировать матрицу $X(N, n)$ и вектор $Y(N)$ (строки – ситуации, столбцы – параметры). Y – столбец, представляющий известные значения неизвестной функции,

b) сформировать функционал, представляющий собой средний риск. Для случая, когда минимизируется средняя квадратичная ошибка, он выглядит следующим образом

$$I(\varphi) = \int (y - \varphi(x))^2 P(x, y) dx dy (1)$$

с) выбрать оценку среднего риска, которая уже не зависит от неизвестной плотности вероятностей $P(x, y)$

$$J(\tilde{\varphi}) = \sum_{i=1}^N \left(\frac{y_i - \tilde{\varphi}(x_i)}{F(x_i)} \right)^2 (2)$$

d) найти минимум оценки среднего риска (2) по всем подвыборкам из исходной выборки ситуаций (X, Y) , и всем видам функций, принадлежащих заданному классу, или группе классов. В результате будем иметь функцию $\hat{\varphi}$,

$$J(\hat{\varphi}) = \min J(\tilde{\varphi}), (3)$$

которая некоторым оптимальным образом приближает неизвестную зависимость:

Существующие методы восстановления непрерывных функций по данным формально вписываются в перечисленные выше три этапа. Различия могут состоять лишь в выборе оценки (2), которая может быть получена как теоретически [2], так и с помощью экспериментальных данных [3,4]. В дальнейшем, для проверки эффективности предложенного подхода, при минимизации (2) использовалась оценка, основанная на данных, причем сама минимизация этой оценки осуществлялась с помощью алгоритма полного перебора [5], позволяющего достигнуть точки экстремума функционала (2) за приемлемое время. Ограничения в данном случае могут быть связаны лишь с размерами совокупностей, которые не должны включать в себя более 12-13 параметров. При решении практических задач, такое ограничение не является серьезным препятствием, поскольку все основные существующие в природе физические законы связывают, как правило, небольшое число переменных. Заметим, что в полном переборе

участвуют не только измеряемые параметры, но и значения функции от этих параметров, вычисляемых в тех же точках, причем число таких значений, для достижения более глубокого экстремума, должно многократно превосходить число исходных параметров.

На практике, однако, нет необходимости перебирать широкие классы функций для целей аппроксимации, поскольку многочлен с целочисленными степенями аппроксимирует любую непрерывную функцию с любой наперед заданной точностью. Главное при этом, чтобы степени параметров и наборы коэффициентов при одночленах доставляли минимальное значение оценке среднего риска (2). Таким образом, если подставить в функционал (2) многочлен некоторой максимально допустимой степени

$$\varphi^{(n)} = \sum_{i=1}^m \alpha_i (x_1^{k_1}, \dots, x_p^{k_p}),$$

то после оптимизации будем иметь некоторый другой многочлен, существенно более простой, в котором коэффициенты $\alpha_1, \dots, \alpha_m$ представляют собой рациональные числа, могущие принимать также и нулевые значения. При этом каждый сомножитель $x_i^{k_i}$ будет входить в состав какого-либо одночлена $x_i^{k_i}, \dots, x_p^{k_p}$ с любой целочисленной степенью k_i , удовлетворяющей неравенству

$$(k_1 + k_2 + \dots + k_p \leq n),$$

где n — максимально допустимая степень полинома. Последние два замечания имеют принципиальное значение, поскольку именно это отличает рассматриваемую полиномиальную аппроксимацию, от обычной полиномиальной регрессии. Фактически, вхождение какого-либо из исходных параметров в одночлен с нулевой степенью означает его там не присутствие, а нулевые значения каких-либо коэффициентов α_i — означает неинформативность соответствующих одночленов. Подобный процесс оптимизации с помощью многочлена можно рассматривать как разложение функции в полиномиальный ряд, порядок, состав и число одночленов которого определяет критерий среднего риска.

ПРЕДСТАВЛЕНИЕ ДАННЫХ ДЛЯ ГЛОБАЛЬНОГО МОДЕЛИРОВАНИЯ.

Остановимся подробнее на данных. Какими должны быть данные, чтобы восстанавливаемая по ним функция, адекватно описывала процесс? Во-первых, если представить эти данные в виде таблицы $A(N, n + 1)$, где строки – ситуации, а столбцы – параметры, то таблица должна быть вытянута в вертикальном направлении. Иначе говоря, число строк должно быть, по крайней мере, в несколько раз больше числа столбцов. Общий размер матрицы должен соответствовать сложности восстанавливаемой функции: если зависимость по предположению сложна, то размеры матрицы должны быть увеличены. Основное же требование к данным заключается в том, что все строки матрицы A должны поступать случайно и независимо согласно некоторому неизвестному, но фиксированному распределению плотности вероятности $P(x, y)$. Очевидно, при постановке задач моделирования атмосферы, последнему условию можно удовлетворить лишь с большой натяжкой, поскольку метеорологические ситуации поступают не случайно и независимо, а в хронологическом порядке. Принимая во внимание тот факт, что атмосфера постоянно подвергается циклическому воздействию, обусловленному внешними признаками, предположение о независимости исходных ситуаций может быть сделано весьма условно.

В общем виде задачу можно неформально представить следующим образом: имеется глобальная информация, снятая на некотором интервале времени (временной ряд), требуется восстановить пространственно-временные зависимости, которые бы с удовлетворительной точностью выполнялись не только внутри этого временного интервала, но и в определенной степени за его пределами.

Но как охватить всю имеющуюся глобальную информацию? Как учесть весь опыт, накопленный в данных, доступных нам для моделирования в настоящий момент? В рамках исходного временного ряда рассмотрим два типа ситуаций: первый тип, - это большие ситуации, состоящие из параметров, измеренных в один определенный момент времени, второй тип, - это малые ситуации, состоящие из набора измерений в одной географической точке. Очевидно, ситуации первого типа состоят из ситуаций второго типа, число которых может быть очень велико. Прямое использование ситуаций первого типа для восстановления функций в данном случае неприемлемо ввиду того, что ситуации огромны, а длина ряда относительно мала. При этом, как число переменных, так и число больших ситуаций может измеряться десятками, или даже сотнями тысяч. Матрица A в данном случае будет вытянута как в длину, так и в ширину. Подобное представление исходных данных непременно должно нас привести не только к необходимости оперировать плохо обусловленными матрицами, но и к серьезной потере информации, обусловленной ее нерациональным использованием.

Тем не менее, постараемся задействовать всю имеющуюся информацию в той степени, в какой она может быть полезна для моделирования. Для этой цели в качестве строк исходной матрицы A будем рассматривать малые ситуации, представляющие собой совокупности измерений в точке, а уравнение для определения аппроксимирующей функции ψ

$$T_{t_0+\Delta t} = \psi(Q(V, t), \quad (t \leq t_0)$$

распишем для всех таких точек. При этом число таких уравнений N будет многократно превосходить число возможных параметров m ($N \gg m$), от которых потенциально может зависеть аппроксимирующая функция. Такой переход моментально решает проблему соотношения числа параметров и ситуаций в исходных данных. Однако если брать все временные слои исходных данных подряд (в хронологическом порядке), то распределение ситуаций, каждая из которых представляет набор данных в одной географической точке, должно отличаться большой сложностью, и это,

несомненно, повлечет за собой усложнение самой восстанавливаемой функции. Сложность же аппроксимирующей функции в условиях ошибок измерений, которые всегда присутствуют в экспериментальных данных, никак нельзя рассматривать как положительный фактор.

Для упрощения распределения ситуаций, каждая из которых представляет набор данных водной географической точке, используем циклический характер изменения состояния атмосферы.

Лучше всего это рассмотреть на примере: пусть имеется многолетний ряд глобальных данных, измеренных через каждый час, и мы собираемся прогнозировать погоду на час вперед. Для этой цели мы очевидно должны сделать 1 шаг по времени вперед. Зафиксируем определенный час, день и месяц во временном ряду, а все остальные элементы удалим из этого ряда. В результате получим ряд, число членов которого будет во много раз меньше исходного ряда. Очевидно, что многомерное распределение полученного таким образом ряда будет существенно более простым, поскольку в нем отсутствуют суточные, месячные, или сезонные циклы. Однако число ситуаций, каждая из которых представляет одну географическую точку, будет по-прежнему очень велико, поскольку даже одному временному слою могут соответствовать десятки тысяч таких ситуаций. Поэтому можно ожидать, что на основе сформированных таким образом ситуаций, может быть получено устойчивое решение ψ , позволяющее спрогнозировать соответствующий элемент погоды на один шаг вперед.

С вычислительной точки зрения мы теперь имеем желаемый эффект налицо. А именно, мы имеем много ситуаций (много экспериментов) и относительно немного параметров. Это гарантирует нам с одной стороны отсутствие вычислительных трудностей, связанных с обращением плохо обусловленных матриц, и в то же время возможность использовать всю полезную информацию исходного ряда. Однако, используемые при этом синхронные ситуации, т.е. ситуации, измеряемые в какой-то определенный момент времени, противоречат классической постановке задачи

восстановления функции по данным, которая требует их статистической независимости. Действительно, ситуации соответствующие определенному сроку (год, месяц, день, час) связанные, прежде всего по всем трем координатам географического пространства, никак не могут считаться статистически независимыми. Но если принять во внимание, что требование независимости связано лишь с оцениванием результатов моделирования, то препятствие может быть в значительной степени устранено с помощью процедуры группового скользящего контроля.

СКОЛЬЗЯЩИЙ КОНТРОЛЬ

Процедура скользящего контроля, состоящая в последовательном исключении ситуаций из материала обучения и вычислении на этих исключенных ситуациях ошибок как на независимом материале, известна давно (см., например [6]). Однако, долгое время эта процедура не находила себе должного применения ввиду большого объема вычислений, требуемых для ее реализации. Фактически, речь могла идти лишь об оценивании конечного результата построений с помощью скользящего контроля, но никак не о минимизации с его помощью среднего риска. Компактные формулы, позволяющие оценивать средний риск в условиях больших объемов исходной информации, представлены в работе [3]. Там же была получена формула и для группового скользящего контроля

$$\Delta_{\text{гр}} = \frac{1}{N} \sqrt{\sum_{j=1}^m \sum_{k=1}^r (\lambda_k^{(j)})^2},$$

где

$$\lambda^j = (I - K_j^T B^{-1} K_j)^{-1} (Y_j - K_j^T B^{-1} Z_j), (1 \leq j \leq m)$$

которая по существу представляет собой оценку среднего риска. Здесь внутреннее суммирование осуществляется по числу элементов в группе,

внешнее – по числу групп, B – ковариационная матрица, полученная по всем ситуациям, K – матрица исключенных ситуаций ($n \times r$).

Первоначально, цель получения этой формулы состояла, прежде всего, в сокращении времени счета, которое требуется для минимизации среднего риска. Действительно, согласно этой формуле, матрица B обращается лишь один раз, а в дальнейшем всего лишь корректируется. В наши дни машинное время уже не имеет такого сакраментального значения как ранее, но полученный функционал приобретает в данном случае новый, более глубокий смысл. Осуществляя групповой скользящий контроль, при котором в каждую группу входят только синхронные ситуации, мы будем иметь возможность оценивать результаты построений не по каждой отдельной ситуации, а по группам ситуаций, которые в первом приближении можно считать статистически независимыми. Разнося группы ситуаций на определенное расстояние по времени, можно добиться большей степени независимости групп, чем это имеет место в случае, когда группы следуют в матрице одна за другой в хронологическом порядке. В результате будем иметь исходный материал для восстановления функций в виде независимых групп ситуаций, а это и есть и есть основное требование при статистическом оценивании результатов построений.

АНАЛОГИИ

Просеивая, таким образом, исходный ряд глобальных наблюдений, мы сможем восстановить и другие функции ψ_i для шагового прогноза других прогнозируемых элементов T^i . В результате будем иметь формулы

$$T_{t_0+\Delta t}^i(V) = \psi_i(Q(t, t_0 - \Delta t, V)), (4)$$

где $i = 1, \dots, l$, а время $t \leq t_0$. может принимать дискретные значения, входящие в интервал предыстории соответствующих процессов. Получив таким образом прогностические значения T^1, \dots, T^l , будем иметь условия, для того чтобы сделать шаг по времени вновь и получить новые

прогностические значения. При этом после получения очередной группы прогнозируемых параметров, мы должны, вновь осуществлять просеивание членов исходного временного ряда, и таким образом, на основе упрощенного распределения, находить новые аппроксимирующие функции для каждого из прогнозируемых элементов.

Заметим, что если в уравнениях (4) положить $t = t_0$, то будем иметь всего лишь три значения дискретного времени, и в этом случае аналогия с гидродинамическими схемами прослеживается наиболее явно. Действительно, формулы (4) будут при этом соответствовать прогностическим формулам в гидродинамических схемах, в которых производная по времени представляется в виде центральных разностей по времени. При эмпирическом подходе начальный интервал предыстории, который в гидродинамических схемах либо фиксирован, либо минимален, желательно задавать достаточно продолжительным, оставляя тем самым возможность для критерия среднего риска выбрать оптимальный вариант. Очевидно, начальный интервал предыстории должен быть согласован с масштабами рассматриваемых процессов и, соответственно, с шагом по времени. В реальных моделях этот шаг должен выбираться исходя из реальных условий сбора и хранения данных, учитывая при этом масштабы прогнозируемого процесса. Для прогноза элементов погоды на месяц, сезон, или год, шаг по времени может быть выбран достаточно большим, если же прогноз считается до десяти суток, шаг нужно выбирать минимально возможным, но не менее интервала, через который осуществляются измерения.

Отличительную особенность гидродинамического подхода по сравнению с эмпирическим подходом можно рассмотреть на простом примере. Предположим, что система уравнений гидродинамики сведена к одному уравнению с одним неизвестным, что при некоторых допущениях возможно. Тогда, вынося частную производную по времени от прогнозируемого

элемента в левую часть уравнения, а в правую часть – все остальные члены уравнения, от которых эта производная может зависеть, можно записать

$$\frac{\partial T}{\partial t} = A(X) ,$$

где X - некоторый вектор, а A – дифференциальный оператор, в который должны входить конвективные члены и некоторые другие комбинации производных или параметров. Представим это уравнение в разностном виде

$$\frac{T_{t+\Delta t} - T_t}{\Delta t} = \bar{A}(X),$$

где \bar{A} - конечно-разностный оператор, соответствующий дифференциальному оператору A . Тогда оставляя предсказываемое значение в левой части и перенося все остальное в правую часть, будем иметь выражение

$$T_{t+\Delta t} = T_t + \Delta t \bar{A}(X) ,$$

с помощью которого можно сделать шаг по времени вперед. Принципиальным отличием является то, что в статистическом случае функция ψ , с помощью которой делается шаг по времени, получается из опыта, а в случае гидродинамической модели соответствующий оператор получается из теоретических рассуждений. Однако любая гидродинамическая схема может быть легко синтезирована в описанную выше статистическую схему. Для этого надо лишь значение разностного оператора \bar{A} в различных точках, включить в качестве самостоятельного параметра в вектор Q . Минимизация среднего риска при этом должна показать, какой вклад вносит гидродинамическая модель в общую синтезированную модель, и насколько такое включение целесообразно

МОДЕЛЬ

На рис. 1 представлена общая схема прогнозирования поля на t шагов вперед ($1 \leq i \leq t$), в предположении, что

аппроксимирующие функции $\{\varphi_i^j\}$ уже построены. После ввода начальных данных A и матрицы функций $\{\varphi_i^j\}$ осуществляется процесс прогнозирования полей шагами по времени. При этом A преобразуется в матрицу $A(i)$, которая используется в дальнейшем как исходный материал для вычисления всех функций на i – ом шаге по времени. Далее, во внутреннем цикле, матрица $A(i)$ преобразуется в матрицу $A(i, j)$, по которой уже непосредственно вычисляются прогнозируемые метеорологические поля

$$y = \varphi(A(i, j))$$

При этом осуществляется пополнение вектора полей y^1, y^2, \dots, y^j , в результате чего, на завершающем этапе внутреннего цикла будем иметь ряд y^1, y^2, \dots, y^k , представляющий совокупность всех полей после первого шага по времени. Цикл по количеству шагов по времени завершает процесс, в результате которого на последнем шаге будем иметь совокупность полей, соответствующих заданной заблаговременности. При этом после очередного i – го шага по времени, очередная совокупность полей присоединяется к матрице $A(i + 1)$ для получения результатов на $(i + 1)$ - м шаге. Таким образом, для того чтобы дать прогноз на i – ом шаге по времени j – го элемента, надо предоставить, условно выражаясь, в качестве аргумента функции φ_i^j матрицу $A(i, j)$.

Однако если в качестве исходных данных используются не срочные наблюдения, а какие-либо осредненные характеристики элементов, то многошаговый процесс прогнозирования ставится под сомнение. В этом случае, возможно, следует сделать один единственный шаг, покрывающий всю заданную заблаговременность прогноза. То же самое может быть сказано в случае, когда исходные данные представляют не все элементы определяющие погоду, а лишь некоторую их часть. Однако окончательный ответ на этот вопрос может быть получен лишь в результате экспериментов.

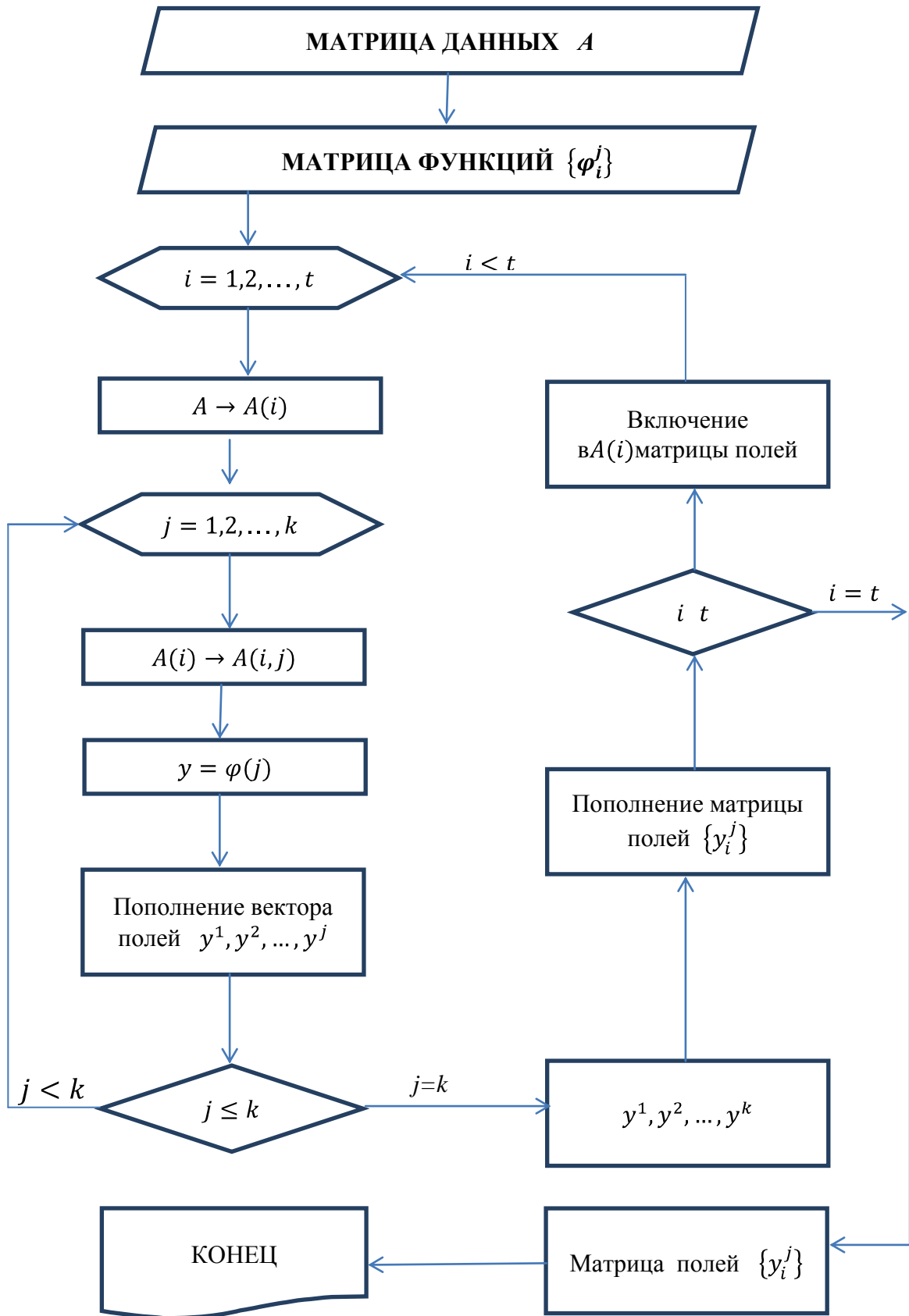


Рис. 1 Схема прогнозирования полей nat шагов вперед
 (i – индекс числа шагов, j – индекс числа функций)

Заметим, что целесообразность подробных шагов по времени имеет место, прежде всего, при краткосрочных и сверхкраткосрочных прогнозах. Об этом свидетельствуют как сама конструкция уравнений гидродинамики, так и весь опыт гидродинамического моделирования атмосферы в целом. Действительно, шаг по времени в разностных схемах возникает из частных производных по времени, которые предполагают скорость изменения функции в какой-то определенный момент. Поэтому, всякое непропорциональное увеличение шага по времени может привести не только к потере важной информации, но и к другим нежелательным эффектам, отрицательно влияющим на результат.

ЭКСПЕРИМЕНТЫ

Для моделирования крупномасштабных процессов в атмосфере в качестве исходной информации могут быть использованы как данные измерений на станциях, так и результаты объективного анализа, представляющие собой значения в узлах регулярной сетки. В режиме счета шагами по времени с помощью рассмотренной модели оба эти варианта информации имеют как свои преимущества, так и свои недостатки. Наибольший интерес представляет использование первичной информации, поскольку при этом удастся избежать дополнительной обработки данных, связанных с объективным анализом. Однако доступ к первичной информации крайне затруднителен, поэтому ее использование для целей моделирования на данном этапе не представляется возможным. В отличие от первичной информации, данные объективного анализа полей вполне доступны и могут использоваться как для текущего прогнозирования, так и для построения моделей (см., например, [7])

В настоящей работе для проверки эмпирического подхода использовались полученные с сайта [7] файлы NCEP/NCAR, представляющие собой

срочные данные РЕАНАЛИЗА, снятые через каждые 6 часов. Эксперименты проводились по данным за июль и январь месяцы с целью получения прогностических полей температуры на один шестичасовой шаг вперед. На рис. 2 изображены кривые характеризующие поведение оценок среднего риска в зависимости от размерности вектора параметров, построенные для одношагового прогноза температуры в июле месяце. При этом, верхняя кривая построена при участии координат станций, нижняя, - без их участия. Как видно из рисунков, оценки среднего риска при восстановлении температуры без координат, существенно превосходят соответствующие оценки, полученные с участием широты и долготы. И это вопреки тому, что сами по себе координаты, взятые как отдельные параметры, обладают достаточно высокой информативностью. Такой результат, изначально казавшийся парадоксальным, объясняется достаточно просто. Широта долгота, в точках измерений участвуют при восстановлении зависимостей как два свободных параметра, которые непременно должны присутствовать в итоговом наборе параметров. Однако конструкция данных при восстановлении такова, что позволяет вообще не использовать координаты станций, поскольку положения точек измерений косвенным образом уже зафиксированы в данных, и этот порядок должен сохраняться на всех шагах по времени. Фактически в этом случае, координаты станции (или координаты ячеек) заменяются некоторым порядковым номером (или символом), который при восстановлении функций не используется, но при текущем прогнозировании позволяет при надобности идентифицировать очередную ситуацию. Поэтому все дальнейшие эксперименты по восстановлению полей температуры проводились без участия широты и долготы.

Заметим, что на этом рисунке, как и на всех последующих, по оси ординат отложена не сама квадратичная ошибка, а величина пропорциональная квадратичной ошибке. Однако, об эффективности аппроксимации, прежде всего, говорит не сама квадратичная, а скорость ее

убывания. Дополнительным ориентиром, при этом, может служить ошибка на одном первом параметре, которая соответствует средне климатическому прогнозу

На рис. 3 изображена кривая поведения ошибок прогноза на 6 часов вперед, которые рассчитывались для 2 января, начиная от 00 часов. При этом максимальный интервал предыстории процесса составлял 18 часов. Таким образом, при восстановлении асинхронной зависимости принимало участие 8 параметров, из которых 4 представляли собой высоту изобарической поверхности на уровне 500 mb., а остальные - температуру на том же уровне.

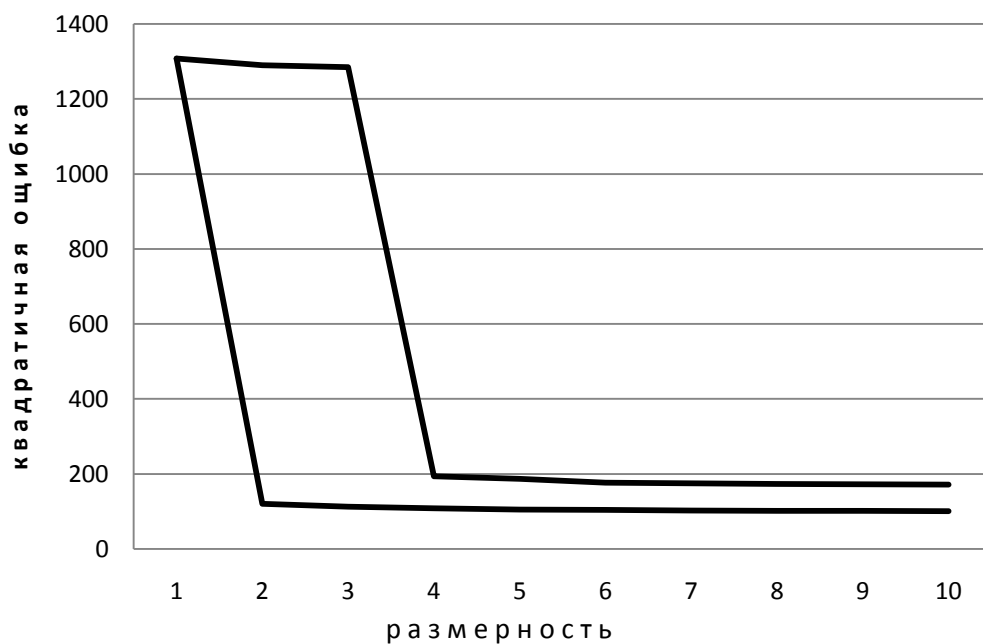


Рис. 2. Кривые оценок среднего риска, построенные по одному и тому же материалу. Верхняя кривая – с участием широты и долготы, нижняя кривая – без их участия.

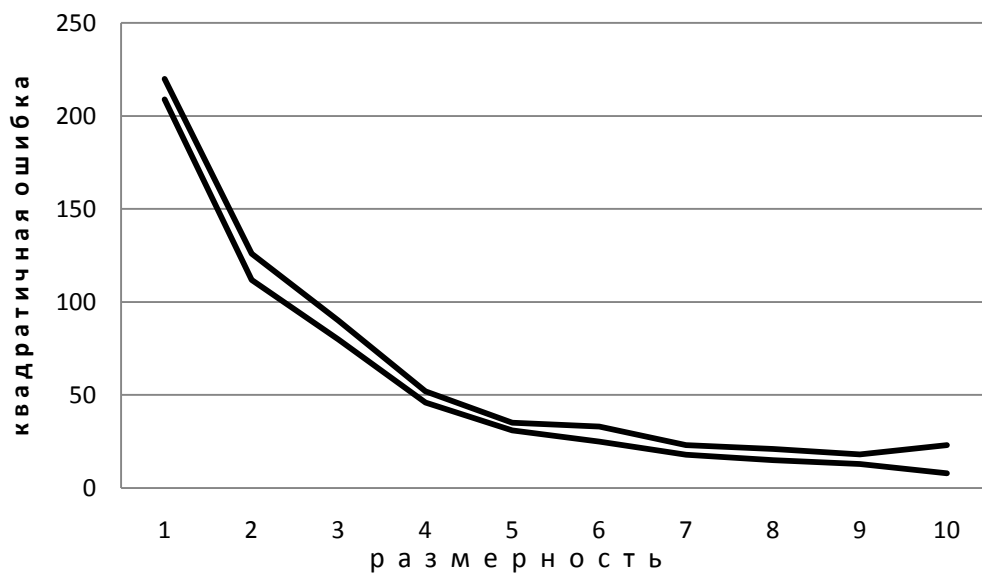


Рис. 3. Кривые поведения ошибок. Нижняя кривая соответствует средне-квадратичным ошибкам, верхняя кривая – ошибкам скользящего контроля

Как видно из рисунка, кривые располагаются очень близко друг к другу, что естественным образом объясняется использованием большого числа ситуаций, при относительно небольшом числе потенциально возможных аргументов.

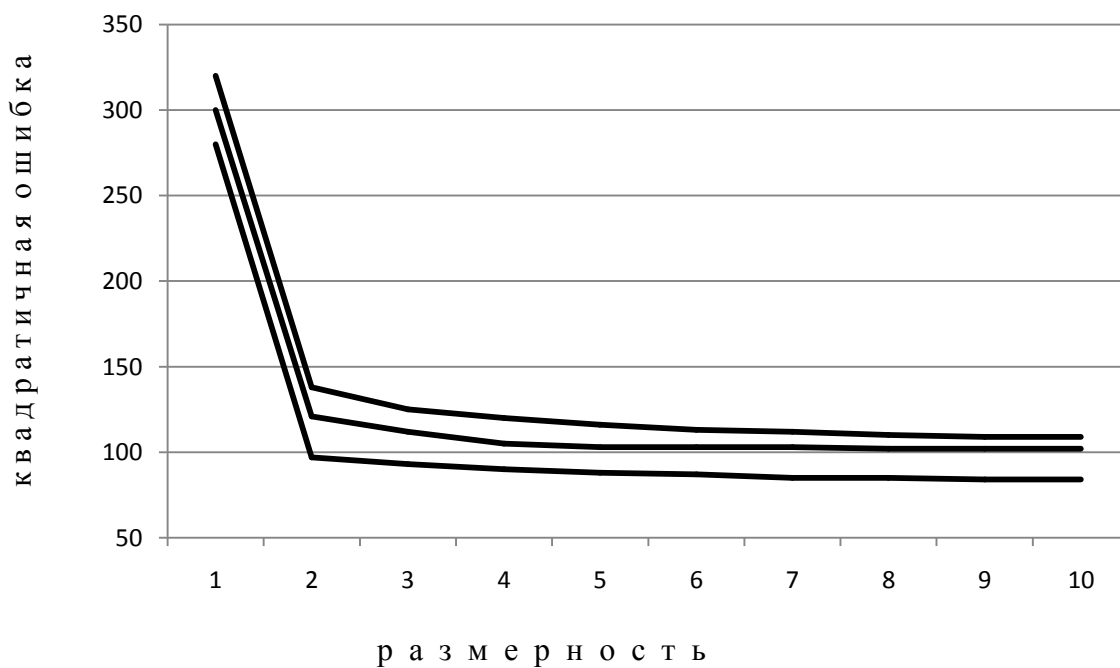


Рис.4. Кривые оценок среднего риска для сферы (верхняя кривая), полусферы (средняя кривая), южного полушария (нижняя кривая)

На рис. 4 отображены аналогичные кривые, полученные при прогнозировании поля температуры, как для всей сферы, так и для северного и южного полушария отдельно. Как видно из рисунка, ошибки в глобальном случае почти такие же, как и ошибки прогнозов для северного полушария. Что же касается южного полушария, то в этом случае ошибки прогнозов стабильно меньше ошибок, как на северном полушарии, так и на всей сфере. Причины тому могут быть разными; одна из причин, – это смена сезонов: в момент прогноза южном полушарии, в отличие от северного полушария, должно наблюдаться лето. Более того, получаемые при этом прогностические формулы практически не отличаются друг от друга по виду функциональной зависимости. При этом, для всей сферы мы имеем формулу

$$T(t + \Delta t) = \alpha_0 + \alpha_1 T(t - 3\Delta t) + \alpha_2 T(t - 2\Delta t) + \alpha_3 H(t - 3\Delta t) + \alpha_4 H(t - 2\Delta t) + \alpha_5 T(t),$$

тогда как для полусферы мы имеем аналогичную формулу

$$T(t + \Delta t) = \alpha_0 + \alpha_1 T(t - 3\Delta t) + \alpha_2 T(t - 2\Delta t) + \alpha_3 H(t - 3\Delta t) + \alpha_4 H(t - 2\Delta t) + \alpha_5 T(t) + \alpha_6 T^3(t) H(t - 2\Delta t),$$

отличающуюся от предшествующей формулы лишь последним членом.

Формула для южного полушария отличается от первой формулы лишь двумя последними членами

$$T(t + \Delta t) = \alpha_0 + \alpha_1 T(t - 3\Delta t) + \alpha_2 T(t - 2\Delta t) + \alpha_3 H(t - 3\Delta t) + \alpha_4 H(t - 2\Delta t) + \alpha_5 T^3(t) + \alpha_6 T^3(t) H(t - \Delta t),$$

которые на самом деле вносят ничтожный вклад по сравнению с предшествующими членами. Такой результат свидетельствует о том, что выбираемые распределения ситуаций являются достаточно однородными, а получаемые по этим распределениям формулы являются достаточно устойчивыми. Примечательно, что во всех трех формулах преобладают одночлены первой степени. Однако, это никак не свидетельствует о простоте моделируемых процессов. Выбранный эксперимент соответствует лишь

одноуровневой модели при двух исходных параметрах, и поэтому, результаты описания моделируемых процессов никак не могут претендовать на их полноту. Если включить в эксперимент дополнительные параметры, а также дополнительные уровни, на которых эти параметры измеряются, то полиномиальное представление полей будет соответствовать более адекватному описанию реальных процессов, и в этом случае преобладание одночленов более высоких степеней будет вполне вероятным.

ЗАКЛЮЧЕНИЕ

В заключении сформулируем основные отличия и основные преимущества нового подхода по сравнению с традиционными методами глобального прогнозирования. Сравнения с другими статистическими методами глобального прогнозирования едва ли в данном случае уместны, ввиду того, что глобальный прогноз до сих пор был связан лишь с гидродинамическими методами. Что же касается гидродинамического подхода, то главное преимущество по сравнению с ним состоит в том, что нет необходимости вычислять производные, что связано с идеализацией действительности и необходимостью решения некорректных задач. Кроме того, нет необходимости оперировать системой разностных уравнений, что связано с решением проблемы устойчивости разностных схем. Вопрос существования и единственности решения в данном случае не стоит так остро, как это имеет место при гидродинамическом моделировании, поскольку существование решения, и единственность определяется в процессе построения с помощью эффективных статистических критериев.

Выбор шага по времени и пространству при решении систем дифференциальных уравнений приходится делать из соображений устойчивости разностных схем, что напрямую не связано с целями построения. В описанной же конструкции проблема выбора шага по времени и пространству вообще не стоит, поскольку этот выбор полностью

диктуется структурой поступающей гидрометеорологической информации. В случае сверхдолгосрочных прогнозов, когда прогнозируются осредненные характеристики, шаг по времени может быть значительно увеличен, однако он по-прежнему должен оставаться кратным временному интервалу измерения параметров.

Отсутствие дополнительных этапов в построениях, которые, как правило, не вписываются в единый контролируемый вычислительный процесс, такие как согласование, полей, или объективный анализ, является несомненным достоинством описанного подхода по сравнению с гидродинамическими методами. Основное же достоинства подхода состоит в том, что весь процесс построения модели контролируется с помощью единого критерия (критерий среднего риска), и этот критерий напрямую связан с конечной целью построений. Все это указывает на перспективы успешного развития и применения эмпирического подхода с целью глобального моделирования.

ЛИТЕРАТУРА

1. Курант Р., Гильберт Д., Методы математической физики, т. 1, Гостехиздат, М.-Л., 1951
2. **Vapnik.V (1998)** Statistical Learning Theory, John Wiley, 1998, NY, p.732.
3. Поляков Г.Г., Романов Л.Н. Скользящий контроль и линейная регрессия. Метеорология и Гидрология, 1988, N 9.
4. Романов Л.Н. Минимизация риска и восстановление пропусков в атмосферных данных. Сиб. журн. вычисл. математики, РАН. Сиб. отд-ние, 2009, Т. 12, № 2.
5. Поляков Г.Г., Романов Л.Н. Об аппроксимации зависимостей с помощью линейных функций. Труды ЗапСибНИГМИ, вып. 83, 1988.
6. Quenouille M.H. Approximate tests of correlation in timeseries. J.R. Statist. Soc., 1949, B 11, p. 68-84
7. www.tsrl.noa.gov/psd/data/composites/hour