

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ РЯДА СРЕДНЕСУТОЧНЫХ ОСАДКОВ

В.Л. Посошков

Морской гидрофизический институт
НАН Украины
Г. Севастополь, ул. Капитанская, 2
E-mail: vao@alpha.mhi.iuf.net

В рамках задачи даунскейлинга анализируется многолетний ряд ежесуточных наблюдений за осадками в Севастополе на основе применения искусственных нейронных сетей (ИНС). Глобальными переменными (предикторами) служат метеорологические поля с выхода NCEP/NCAR (National Project of Environmental for Prediction/National Centre of Atmospheric Research) реанализа.

1. Введение

Количественный прогноз осадков представляется важным звеном в различных приложениях, простирающихся от прогнозирования наводнений до долгосрочного планирования в сфере сельского хозяйства и водных ресурсов. Вместе с тем точное прогнозирование осадков все еще остается наиболее сложной и трудной задачей метеорологии. В то время как в последние годы улучшение разрешения и физики в численных моделях прогноза погоды существенно повысило точность предсказания многих метеорологических переменных, подобное же улучшение в точности прогноза суточных сумм осадков не было достигнуто как из-за сложности физики процессов осадкообразования, так и потому, что малые временные и пространственные масштабы, содержащиеся в этих процессах, все еще не разрешимы в имеющихся численных моделях. Способы компромисса между существующим разрешением модели и желаемым масштабом прогноза разрабатываются сейчас в рамках одного из важных направлений современных исследований климата, определяемого как "downscaling" [1].

В данной работе было выполнено диагностическое исследование массива ежесуточных осадков, взятого из многолетнего ряда наблюдений, проводившихся в Севастополе с 1906 г. Исследовались связи между глобальными переменными с выхода численной модели реанализа (предикторами) [2] и локальной

переменной (предиктантом) дневными суммами осадков. Оба массива данных имеют пересечение во времени, занимающее период 1982 - 1992 гг. Использовались данные, относящиеся к календарному зимнему времени (декабрь - февраль). В результате использовалась выборка данных, содержащая 900 многомерных событий. Методологически работа выполнена на основе применения искусственных нейронных сетей (ИНС).

Статья подразделяется на 5 разделов. Во втором разделе приводятся основы построения искусственных нейронных сетей, применяющихся в метеорологии. В третьем разделе дается методология выбора и сортировки данных, обучения и тестирования нейронной сети. В четвертом разделе обсуждаются результаты применения нейронной сети к воспроизведению временного ряда суточных сумм осадков и, наконец, в пятом разделе суммируются основные результаты работы.

2. Искусственные нейронные сети

Искусственные нейронные сети являются относительно новым инструментом обработки информации, имитирующим способности человеческого мозга к обучению и запоминанию, а также его чрезвычайно разветвленную архитектуру, обеспечивающую многоканальное параллельное функционирование. ...Ввиду того, что применение нейронных сетей в метеорологии еще не получило широкого распространения, остановимся вкратце на основах их функционирования.

Базовый блок сети представляет собой функционирующий элемент, являющийся грубым математическим аналогом биологического нейрона. Ядром этого элемента является дифференцируемая нелинейная функция, носящая название функции активации, или передаточной функции, аргумент которой - скалярное произведение входного вектора X и весового вектора W .

Передаточная функция имитирует нелинейную реакцию нейрона на химическое воздействие, воспринимаемое как входной сигнал. Аргумент функции активации обычно модифицируется путем добавления смещения θ , являющегося пороговой характеристикой нейрона; это означает, что нейрон возбуждается лишь когда химический сигнал, принимающийся нейроном, превышает по величине некоторый уровень. Технически в алгоритме это реализуется путем подачи на

вход дополнительного сигнала, имеющего единичную величину. Весовой вектор.

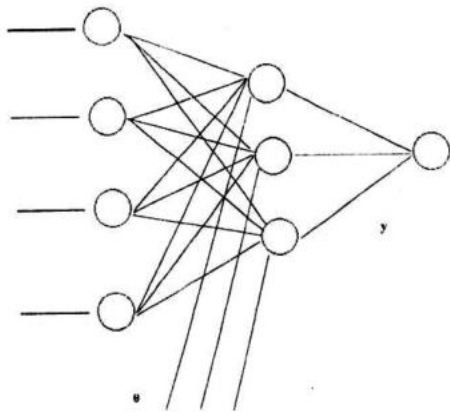


Рис.1 Структурная схема ИНС

образующий синаптические связи между нейронами, носит в себе функции контроля количества информации, получаемой нейроном. В процессе обучения сами компоненты этого вектора приобретают некоторые оптимальные значения, которые меняются по мере поступления новой информации. В целом весовой вектор образует пространство, где хранится приобретенное знание, т.е. то, что мы называем памятью. В практической реализации нейронной сети нейроны организуются в так называемые слои, и информация последовательно преобразуется по мере послонного продвижения от входа к выходу.

За последние десятилетия разработано множество искусственных нейронных сетей [3], различающихся своей топологией, т.е. количеством входов и выходов, числом скрытых уровней (слоев между входными и выходными слоями), числом нейронов на каждом уровне. Для целей настоящей работы выбрана прямая цепь, или цепь прямого распространения, состоящая из трех слоев (трехслойный перцептрон): входной слой (рис.1), где сеть получает информацию, скрытый слой, где происходит ее обработка и выходной слой, представляющий собой реакцию на входную информацию. В нашем частном случае выходной слой представлен единичным нейроном.

3. Методология

Топология сети. Подготовка нейронной сети включала определение ее топологии, выбор данных для обучения, обучение и тестирование сети. Суть сети прямого распространения состоит в том, что во-первых отсутствуют обратные связи между входом и выходом и во-вторых, - связаны

между собой только нейроны различных слоев и отсутствуют боковые связи между нейронами в пределах одного и того же слоя.

ИНС прямого распространения – это, по существу функция, отображающая ряд n переменных, x_i ($i=1, \dots, n$), называемых входными нейронами, в ряд переменных y_k ($k=1$ в данной статье), называемых выходными узлами. Для ИНС с одним скрытым слоем функция параметризуется как

$$y = \tilde{f} \left[\sum_{i=1}^m \tilde{W}_{i1} f \left(\sum_{j=1}^n W_{ij} x_j - \theta_j \right) - \tilde{\theta} \right], \quad (1)$$

где $W_{ij}, \tilde{W}_{i1}, \theta_j, \tilde{\theta}$ являются параметрами (весами), которые требуется определить из данных, называемых обучающей выборкой. m – число скрытых нейронов в промежуточном (скрытом) слое. Функции f и \tilde{f} – функции активации. Определение оптимальных величин весовых коэффициентов и смещений осуществлялось на основе минимизации целевой функции (функционала ошибки) методом обратного распространения ошибки. [4]

В данной работе существенным оказался выбор функции активации. Наилучший результат при настройке сети достигался, когда использовались различные фильтрующие функции в разных слоях сети. Если на входе использовалась сигмоидная функция вида

$$f(y) = \frac{1}{1 + \exp(y - y_c)}, \quad (2)$$

где порог y_c – величина входа, при достижении которой нейрон становится активным, то на выходе функция активации задавалась в виде экспоненты [5]

$$f(y) = \begin{cases} 0, & y \leq y_c, \\ e^{\gamma(y - y_c)} - 1, & y > y_c, \end{cases} \quad (3)$$

где γ – произвольный параметр порядка единицы. В результате сигмоидная функция лучше распознает поступающие образы, имеющие знакопеременный вид, а экспоненциальная функция активации на следующем этапе придает этим образам существенно несимметричный вид, подавляя до нуля слабые сигналы и растягивая большие.

Метод проб и ошибок остается единственным приемлемым способом определения величин, которые наилучшим образом формируют топологию сети, желаемый уровень точности и количество времени обучения. В то время как число

входных и выходных нейронов определяется выбором соответствующих входных и выходных переменных, диктуемым самой задачей, выбор числа

скрытых нейронов далеко не очевиден. Если берется слишком малое число узлов, ИНС не способна решить задачу на обучающей стадии. Если их слишком много, замедляется сходимость процесса алгоритма обучения. Кроме того, благодаря большому числу параметров (т.е. θ_j и w_{ij}) существует риск переопределения параметрами. В этом случае: ИНС может слишком точно описывать обучающую выборку данных, и "теряться" в новых ситуациях при поступлении новых данных [6]. Тестирование показало, что оптимальное число скрытых нейронов для данной задачи составляет 7-9. Это

укладывается в пределы $2n^{1/2} + m$ $2n + 1$ [7], где n – число входных, а m – число выходных нейронов, рекомендованные для идеального числа скрытых нейронов. В этой области число скрытых нейронов мало влияло на результат, поэтому было выбрано наименьшее число. Минимальная топология ИНС составила 11-7-1 (число нейронов соответственно в трех слоях сети).

Обучение. При определении данных для обучения формировались обучающая и контрольная выборки. В процессе обучения и по его окончании выполнялось тестирование сети на оставшихся случаях.

По существу, при настройке сети требуются три опорных ряда наблюдений, или три выборки данных [8]. Это – обучающая выборка, на которой происходит обучение, тестирующая выборка, на которой проводится испытание на способность ИНС к обобщению (при перекрестной методике обучения тестирующая и обучающая выборки могут выбираться из одного ряда в разных вариантах [9, 10] и, наконец, контрольная выборка, включающая вновь поступающие данные, по которой окончательно оценивается эффективность нейронной сети. Иногда контрольная и тестирующая выборки объединяются, что, в частности, сделано и в данной работе. Основные требования к обучающей выборке состоят в том, чтобы она включала данные, охватывающие основные циклы физических процессов. Общее число наблюдений составило, как уже указывалось, 900.

Для настройки ИНС важно правильно выбрать обучающую выборку. Во-первых, эта выборка должна содержать качественные данные, т.е. число ошибочных данных должно быть сведено к минимуму, и во-вторых, – она должна заключать в себе достаточный

диапазон временных масштабов, который охватывал бы как можно большее число физических процессов, ответственных за осадкообразование. К сожалению, сравнительно небольшая длина ряда (10 лет) не позволяет надежно заложить в нейронную сеть межгодовые закономерности. Это накладывает ограничения на способность сети к обобщению на независимой выборке, но поскольку изменчивость суточных сумм осадков в основном проявляется на временных интервалах синоптического масштаба, это позволяет в поисках оптимальной обучающей выборки варьировать ее размеры в достаточно широких пределах.

Заметим, что в принципе для удовлетворительного функционирования сети оказалось достаточным информации, содержащейся всего в одном зимнем периоде (90 суток). Дальнейшее увеличение обучающей выборки не дает принципиальных подвижек в качестве функционирования сети. Все же, чтобы избежать влияния отдельных случайных выбросов, не имеющих под собой физической основы, весь ряд наблюдений был разделен в отношении 1:2, т.е. ИНС настраивалась на одной трети ряда, а остальные две трети использовались для контроля качества сети.

Сортировка данных и выбор предикторов.

Принято считать, что эффект «сверхподгонки» модели к данным (overfitting) возникает от чрезмерно большого числа итераций (циклов обучения), когда модель подгоняется не только под флуктуации, имеющие под собой физическую основу, но и начинает описывать шум. На самом деле это так, но есть и другие причины переопределения параметрами. Использование на входе незначимых предикторов также может привести к перенасыщению параметрами [11]. ИНС может находить связи между несущественным предиктором и предиктантом, которые не имеют физической основы и становятся ложными вне пределов обучающей выборки. Следовательно, переменные-предикторы следует подвергать предварительному отбору (сортировке). Поэтому применение ИНС происходит в два шага. Сортировка данных – определение, какие входные данные ответственны в данной ситуации и затем – обучение. Критерием при отборе предикторов служила не только степень индивидуальной связи между дневными суммами осадков и соответствующим предиктором, но и точность, с которой моделируется в глобальной численной модели соответствующий предиктор.

При выборе предикторов выяснилось, что результат не улучшается при увеличении их

числа свыше десяти. Видимо, это связано со значительной коррелированностью между собой метеорологических переменных [11].

Логично предположить, что осадкообразование определяется не только самим геопотенциалом, но и его значением в окружающих 8 узлах сетки. Это значительно увеличивает число предикторов, что согласно высказанным соображениям нецелесообразно. Поэтому предварительно проводилось сжатие полей геопотенциала. По значениям геопотенциала в 9 узлах сетки, характеризующим локальное барическое поле, формировался новый параметр, характеризующий двумерное искривление (геометрию) барической поверхности в центральном узле сетки. Косвенной мерой такого искривления в точке (i,j) может служить дискретный аналог оператора Лапласа в данной точке, построенный по следующему алгоритму

$$l_{i,j} = \frac{\begin{bmatrix} (h_{i-1,j} + h_{i+1,j} + h_{i,j-1} + \\ h_{i,j+1} - 4h_{i,j}) + \\ 2(h_{i-1,j-1} + h_{i+1,j+1} + \\ h_{i-1,j+1} + h_{i+1,j-1} - 4h_{i,j}) \end{bmatrix}}{4\Delta h^2} \quad (4)$$

где Δh - расстояние между узлами сетки, равное 2.5 град. широты. Таким образом, вместо 8 предикторов используется лишь один: $l_{i,j}$.

В результате сортировки данных были выбраны геопотенциал на трех уровнях (1000, 700 и 500 гПа), лапласианы (формула (4)) геопотенциала на тех же уровнях, компоненты приземного ветра, влагосодержание воздушного столба и глобальная переменная - индекс североатлантического колебания (NAO), который определялся по разности между значениями геопотенциальной поверхности h_{1000} между точкам с координатами 65.7 град с.ш., 343.1 град в.д. (Исландия) и 39.0 град с.ш., 331.9 град в.д. (Азорские острова).

4. Результаты и обсуждение

При настройке сети решающим оказался выбор функции активации. На рис. 2 представлены образцы сравнения выхода ИНС с наблюдениями при различных вариантах выбора передаточных функций. Если через f_1 обозначить сигмоидную функцию, а через f_2 экспоненту, то варианты (f_1, f_1) , (f_1, f_2) , (f_2, f_1) и (f_2, f_2) отображены

соответственно на рисунках а) - г), где комбинация (f_1, f_2) означает, что на входе используется активирующая функция f_1 , а на выходе - f_2 . Как видно из рисунка, лучший результат достигается, когда во входном слое используется функция типа Ферми, а в промежуточном - экспоненциальная функция.

Сеть довольно быстро поддается обучению. Уже 10 циклов достаточно (рис.4а), чтобы сеть достаточно хорошо распознавала события, а в большинстве случаев хороший результат достигается менее, чем за 100 циклов обучения. Дальнейшее увеличение числа итераций приводит к переопределению. Это наглядно видно на рис. 3. Сеть удовлетворительно воспроизводит осадки на контрольной выборке при числе итераций $n=100$; (рис. 2), а при $n=500$ (рис. 4д) очень точно моделирует осадки на обучающей выборке и теряет эти свойства за ее пределами. В большинстве расчетов использовалось от 80 до 120 циклов обучения. При сравнении с наблюдаемыми осадками заметно смещение в «сухую» сторону, т.е. возникает такое ощущение, что сами события-дожди прогнозируются лучше, чем их амплитуды; при этом труднее воспроизводятся экстремальные осадки. Вообще, традиционный коэффициент корреляции невысок и имеет порядок 0.5-0.6. Поэтому в метеорологии для оценки качества прогноза столь трудно прогнозируемых процессов, как ежедневные осадки давно применяются другие количественные критерии. Обратимся и мы к некоторым из них.

Числовые меры успешности прогноза. Хотя наше исследование является диагностическим, в целях дальнейшего анализа результатов с позиций оценки качества моделирования физической переменной нейронной сетью в сравнении с фактическими (наблюдаемыми) значениями воспользуемся методикой оценивания, а также некоторыми числовыми методами оценки, разработанные в прогностических моделях метеорологии. Так, будем считать модельные величины с выхода ИНС прогностическими, а из независимой выборки данных, на которой производится тестирование - фактические величины-наблюдениями. Одна из возможных числовых мер качества прогноза - это взаимная корреляционная функция. Применимость ее к гауссовским процессам не вызывает сомнений, но в приложении к такому существенно негауссовскому процессу как ежесуточные осадки корреляционная функция представляется недостаточно надежной мерой. С прогностической точки зрения

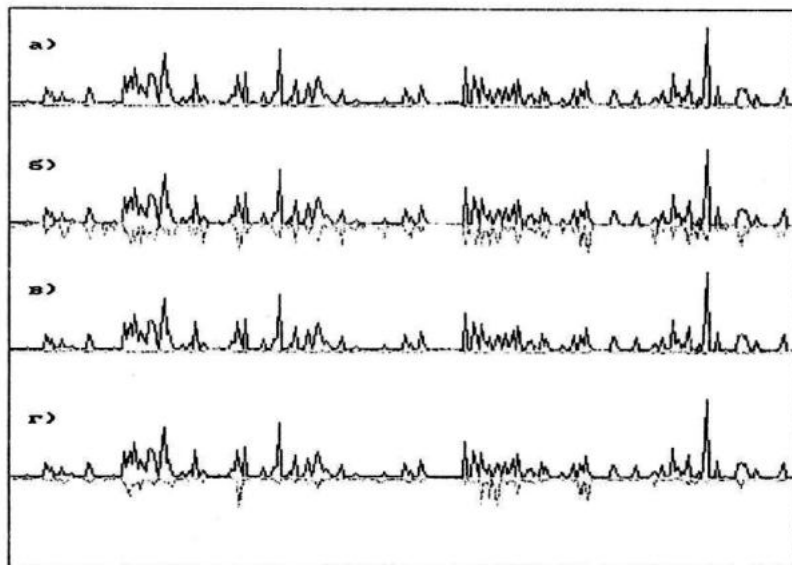


Рис.2. Образцы выхода ИНС при различных сочетаниях функций активации во входном и промежуточных слоях сети (нижние кривые) в сравнении с фактическими наблюдениями (верхние).

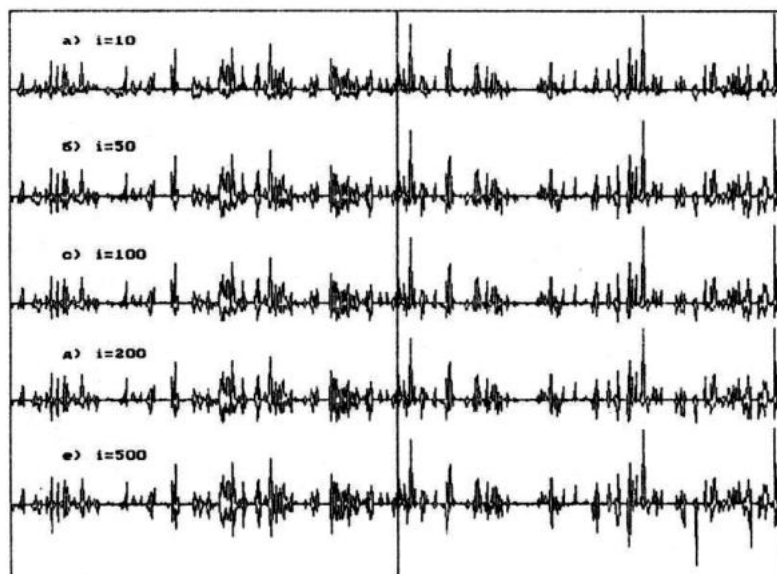


Рис. 3 Фрагменты временных рядов ежесуточных осадков, взятых из наблюдений (верхние кривые) и с выхода нейронной сети (нижние кривые) при различном числе итераций i (а – д). Для удобства сравнения модельные кривые на рисунке приводятся с обратным знаком. Вертикальная прямая обозначает границу между обучающей (слева) и контрольной (справа) выборками данных

представляется более важным прогноз событий (дождь/нет дождя). В самом деле, предварительный анализ результатов показал, что ИНС недостаточно хорошо воспроизводит амплитуды ежесуточных осадков, но похоже, лучше предсказывает осадки как события. Поэтому мы перейдем к методам категориального, или событийного прогноза, т.е. к методам оценивания достоверности процесса, состоящего из событий – в данном случае двух взаимоисключающих событий «дождь/сухо», или «да/нет», или 1 и 0 в числовом выражении. Для каждой пороговой величины осадков (или просто порога) ряд непрерывных функций ежедневных норм осадков, как наблюдаемых, так и модельных. Пусть P_1 и P_0 соответственно число модельных дождливых и сухих дней, а N_1 и N_0 – те же величины, относящиеся к наблюдениям. Очевидно, $N_1 + N_0 = P_1 + P_0 = N$, где N – длина выборки данных. Совместное множество взаимоисключающих событий можно описать в терминах следующей таблицы условных вероятностей, или С-таблицы [8].

Таблица 1. Условные вероятности возможных событий

Прогноз	Наблюдения	
	нет	да
нет	a	b
да	c	d

Если обозначить элементы С-матрицы через p_{ij} , (i – строка, j – столбец), то индексы означают, что происходит событие j , предсказанное как событие i . Иными словами, a – число событий «нет», предсказанных как «нет», b – число событий «да», предсказанных как «нет» (misses), c – число сухих дней, предсказанных как дожди (false alarms) и, наконец d – число событий, предсказанных как события (hits), т.е. правильный прогноз.

На основании табл.1 было введено множество числовых мер, характеризующих качество прогноза [8]. Большинство из них хорошо работают, когда климатическое отношение N_0/N_1 , т.е. отношение числа сухих дней к числу дней с осадками равно половине. Если же такая статистика нарушается в сторону преобладания одних событий в ущерб другим, то эти числовые критерии дают завышенные величины, если прогноз осуществляется в пользу климатологических величин. Поэтому мы отдаем предпочтение

тому критерию, который дает минимальное качество прогноза (в числовом выражении) как в случае прогноза, основанного на случайном выборе, так и в случае постоянного прогноза, основанного на климатологии. Такое число, введенное Хансеном и Купером [12], записывается следующим образом

$$TSS = \frac{ad - bc}{N_1 N_0}, \quad (5)$$

Нетрудно убедиться, что эта величина изменяется в пределах $(-1,1)$. Она равна единице в случае идеального прогноза ($a=N_0$, $d=N_1$, $b=c=0$), равна нулю как в случае климатологического, так и случайного прогноза и равна -1 в случае абсолютно неправильного прогноза ($b=N_1$, $c=N_0$, $a=d=0$). Заметим, что последний случай эквивалентен идеальному прогнозу.

Остается вопрос определения порога для отделения осадков от «неосадков». Кумар и др.[13] выбирали порог 0.5 мм, но само определение события «осадки» в достаточной степени субъективно и всецело зависит от уровня отсечки, или порога – суточной суммы осадков, принимаемой как несущественная величина.

Заметим, что в качестве предикторов использовались двухсрочные данные реанализа, а выход ИНС представлял собой осредненные за сутки суммы осадков. Тестирование показало, что лучший результат достигается с дневными предикторами. Использование среднесуточных предикторов вместо двухсрочных также дает худший результат по сравнению с использованием одних лишь неосредненных данных за дневное время. Сдвиг предикторов по времени на один срок (12час) в ту или другую сторону ухудшает настройку ИНС, так что дневные предикторы за тот же день являются наиболее значимыми. Вообще, синхронность предикторов и смоделированного сетью временного ряда, оказалась очень существенной. Это еще раз говорит о том, что нейронная сеть настраивается не на произвольный временной ряд, а только на тот, который физически связан с процессами, представленными входным слоем нейронов.

5. Заключение

Нейронная сеть позволила смоделировать временной ряд дневных осадков, достаточно хорошо согласующийся с рядом наблюдений как по событиям (дождливым дням), так и по величине выпавших осадков. Ряд оценочных критериев, характеризующих качество

моделирования осадков, применялся одновременно к выходам ИНС и NSEP-реанализа., что позволяет судить о репрезентативности этих двух типов

модельных данных: Суммируем некоторые числовые результаты сравнения в виде таблицы.

Таблица 2

Числовые меры согласования между модельными и наблюдаемыми ежесуточными рядами осадков

	Средне-суточная норма, мм	Коэффициент корреляции	TSS	Вероятность выпадения дождливых дней
Наблюдения	1,18	-	-	-
Нейронная сеть	1,11	0,55	0,575	0,780
Реанализ	0,40	0,36	0,397	0,492

Основным результатом является то, что искусственная нейронная сеть, учитывающая косвенно региональные особенности данного района, путем постобработки выхода модели NSEP-реанализа позволяет воспроизводить ряд ежедневных зимних осадков с большей достоверностью, нежели это делается в самой модели.

ЛИТЕРАТУРА

1. CLIVAR – A Research Programme on Climate Variability and Predictability for the 21st Century.// (WMO/TD-1997.-No. 853)
2. Kalnay, E., and Coauthors. The NCEP/NCAR 40-Year Reanalysis Project // Bull. Amer. Meteor. Sci.- 1996.-77.- 437-471.
3. А.Б. Забавин. Использование искусственных нейронных сетей в задачах изучения Земли из космоса // Исследование Земли из космоса – 2000.- № 6.- С.79-93.
4. Le Cun, Y. A learning scheme for asymmetric threshold networks // Proc. Cognitiva Conf. 85, Paris, France, CESTA-1985- 599-604.
5. Zorita, E., and H. Von Storch. The Analog Method as a simple Statistical Downscaling Technique: Comparison with More Complicated Methods // J. Climate-1999.- 12.- 2474-2489.
6. Aviolet, F., T.Cornu, and D. Kattani. Automatic Clouds Observation Improved by

Artificial Neural Network. // J. of Atmospheric and Oceanic Technology-1998.- 15.- 114-126.

7. Fletcher, D., and E. Goss. Forecasting with neural networks: An application using bankruptcy data // Inf.Manage.-1993.-159-167.

8. Marzban C. and Stumpf G.G. A Neural Network for Damaging Wind Prediction // Wea. Forecasting.-1998.-13.-151-163.

9. Elsner, J.B., and C.P. Schertmann. Assessing forecast skill through cross validation // Wea. Forecasting.-1994.- 619-624.

10. Kuligowski, R.G. and A.P.Barros. Experiments in short-term precipitation forecasting using artificial neural networks // Mon. Wea. Rev.-1998.- 126- 470-482.

11. Kuligowski, R.G. and A.P.Barros. Localized Precipitation Forecasts from Numerical Weather Prediction model Using Artificial Neural Networks // Wea. Forecasting.-1998.- 13- 1194-1203.

12. Hansen, A.W., and W.J.A. Kuipers. On the relationship between the frequency of rain and various meteorological parameters. Konincliijk // Nederlands Meteorologisch Institut. Meded.Verhand.-1965.- 81.- 2-15.

13. Kumar A., Parvinder M., and V. Singh. An Operational Model for Forecasting Probability and Yes/No Forecast // Wea. Forecasting.-1999.- 14.- 38-48.