

Л.Б. Чанга, Р.П. Репинская

**ПРОГНОЗ ОСАДКОВ ДЛЯ ТАНЗАНИИ
ФИЗИКО-СТАТИСТИЧЕСКИМ МЕТОДОМ**

L.V. Changa, R.P. Repinskaya

**RAINFALL PREDICTION FOR TANZANIA USING
THE PHYSICAL-STATISTICAL METHODS**

Рассматриваются принципиальные особенности физико-статистического метода. Рассчитаны и протестированы уравнения множественной регрессионной модели для прогноза осадков в сезоне октябрь – декабрь для каждого из 4-х районов Танзании. Проведено сопоставление опытных прогнозов с исходными данными по независимой выборке за период 2000 – 2004 гг. и показано, что в целом наблюдается хорошее соответствие прогнозируемых осадков с фактическими. Высокое качество прогнозов с помощью оптимальных уравнений регрессии, полученных на зависимой выборке, и хорошее соответствие прогнозируемых и фактических осадков на независимой выборке позволяют рекомендовать данные модели для применения в оперативных условиях в Танзании.

The basic concepts of physical-statistical methods are explored. Regression models for seasonal rainfall prediction (October – December) for each of the four regions of Tanzania are constructed and tested. The comparison between the predicted and observed rainfall over the testing period (2000 – 2004) is carried out. Due to high quality of the forecast results in the training and testing period it is recommended that the models should be used for operational forecast in Tanzania.

Принципиальные особенности физико-статистического метода, использованного нами для долгосрочного прогноза осадков для Танзании, разработаны М.И. Юдиным [Юдин, 1967, 1968]. В работах [Юдин, 1967, 1968, 1972, 1974, 1975] изложены основы названного метода и способы статистической обработки и анализа исходных данных, рассматриваются также обобщенные характеристики будущей погоды, связь между ними и методика отбора статистически значимых связей. Используя эти работы, кратко изложим принципиальные особенности метода.

К физико-статистическим методам прогноза погоды [Кендалл, 1976; Блажевич, 1987; Малинин, 2002, 2003] относятся методы, опирающиеся на определенные данные о физических процессах в атмосфере, используемые в основном для выбора предикторов. Существенность связей между предикторами и предиктантами определяется путем статистической обработки эмпирических данных. При прогнозе большой заблаговременности необходимо учесть информацию об атмосферных процессах большой протяженности в пространстве и во времени [Юдин, 1969]. Эмпирический путь установления асинхронных статистических связей дает возможность получить информацию о характере ожидаемой погоды. Весьма часто установленные на материале прошлых лет статистические зависимости между рассматриваемыми величинами теряют свое зна-

чение при переходе к текущим материалам сразу или через несколько лет [Юдин, 1974]. В результате основанные на использовании таких зависимостей статистические прогнозы оказываются не эффективными. Опасность этого рода исключается в тех случаях, когда физическая природа зависимости между явлениями ясна и статистический анализ нужен лишь для числового выражения соответствующей связи. Естественно, что физические величины, аномалии которых имеют одинаковый знак на больших пространствах и сохраняются в течение длительного промежутка времени, представляют собой интерес для целей долгосрочного прогноза погоды (ДПП). Действительно, значение такой величины в начальный момент характеризует особенности переноса тепла, влаги и напряженности циркуляционного поля в обширных районах в течение последующего достаточно большого интервала времени. Ход будущей погоды определяется не только такими длительными постоянными воздействиями, но и всевозможными возмущениями меньших пространственных и временных масштабов (т.е. региональных и локальных масштабов).

Консервативным предиктором называют такой, значение которого в начальный период определяет его поведение в течение более или менее длительного промежутка времени, примыкающего к начальному. Для повышения успешности статистического прогноза целесообразно отфильтровать флуктуации и уменьшить влияние микроклиматической неоднородности данных, погрешности измерений и т.д. Как известно, основным способом решения этой задачи является усреднение значений случайных величин. Сглаживание может быть пространственным, временным или пространственно-временным.

Применяются также способы выделения первых членов разложений по тригонометрическим функциям, по полиномам Чебышева и т.д. Одним из наиболее эффективных способов отсекающего мелкомасштабных компонентов в поле метеорологического параметра является выделение первых членов ряда в разложении по эмпирическим ортогональным составляющим (ЭОС). Так, было показано [Юдин, 1975; Блажевич, 1986], что разложение метеорологических полей в ряды по ЭОС позволяет выделить наиболее интенсивные и крупномасштабные элементарные колебания типа стоячих волн и отфильтровать мелкомасштабные возмущения.

Физико-статистический метод сводится к следующим этапам:

- определение системы характеристик исходного состояния атмосферы, среди которых отыскиваются наиболее информативные предикторы;
- предварительная обработка совокупности полей предиктантов и предикторов с целью получения их более репрезентативных характеристик;
- вместо данных по отдельным пунктам рассмотрение их средних значений для отдельных районов, представляющих собой достаточно однородную в физико-географическом отношении информацию;
- расчет коэффициентов линейной корреляции между предикторами и предиктантами и выделение связей, удовлетворяющих жестким статистическим критериям правдоподобия;

– дополнительный контроль установленных статистических соотношений с целью получения надежного прогностического вывода в заданной области значений предиктора;

– контроль предикторов с точки зрения объема содержащейся в них информации, установление последовательности предикторов, используемых для прогноза каждого предиктанта, и их ортогонализация по Чебышеву;

– построение уравнений регрессии и вычисление прогностических значений предиктантов;

– оценка успешности выполненного прогноза с помощью ряда статистических критериев, используемых в метеорологической практике.

Метод множественной линейной регрессии. Наиболее адекватным аппаратом построения прогностической модели между физически установленными предикторами и предиктантом является метод множественной регрессии. Модели множественной линейной регрессии (МЛР) соответствует уравнение

$$Y_j = b_0 + \sum_{j=1}^m b_j X_{ij} + \varepsilon_i, \quad (1)$$

где b – коэффициенты уравнения регрессии; b_0 – градиент; ε_i – вектор остатков (ошибок), не описываемых уравнением регрессии.

Коэффициенты регрессии определяются с помощью метода наименьших квадратов (МНК), в соответствии с которым требуется минимизировать квадрат суммы разности фактических и вычисленных по уравнению (1) значений функции отклика, т.е.

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \tilde{Y}_i)^2 = \sum_{i=1}^n [Y_i - (b_0 + b_1 X_{i1} + \dots + b_m X_{im})]^2 = \min,$$

где \tilde{Y}_i – вычисленные по уравнению МЛР значения функции отклика.

Основные критерии качества модели МЛР. 1. Множественный коэффициент линейной корреляции ($0 \leq R \leq 1$) представляет собой аналог обычного парного коэффициента корреляции и характеризует меру линейной связи между фактическими и вычисленными по уравнению МЛР значениями функции отклика, т.е.

$$R = \frac{1}{n \sigma_y \sigma_{y(x)}} \sum_{i=1}^n (Y_i - \bar{Y})(\tilde{Y}_i - \bar{Y}), \quad (2)$$

где \tilde{Y}_i – вычисленные по модели МЛР значения функции отклика; $\sigma_{y(x)}$ – стандартное отклонение значений \tilde{Y}_i .

2. Линейный коэффициент детерминации представляет собой квадрат множественного коэффициента корреляции $R^2 = D_{y(x)}/D_y$, где $D_{y(x)}$ – дисперсия, вычисленная по уравнению регрессии значений функции отклика. Коэффициент детерминации показывает часть объясненной дисперсии функции отклика.

3. Среднеквадратическое отклонение модели $\sigma_{y(x)} = \sqrt{\sum (Y_i - \tilde{Y}_i)^2 / (n - m - 1)}$.

4. Критерий Фишера используется для оценки адекватности (значимости) модели МЛР. С этой целью проверяется нулевая гипотеза вида $H_0: D_{y(x)} = D_\varepsilon$, т.е. дисперсия вычисленных по уравнению МЛР значений функции отклика равна дисперсии ошибки. Нулевая гипотеза проверяется с помощью критерия Фишера: $F = D_{y(x)}(n - m - 1) / (D_\varepsilon m)$. Вычисленное значение критерия F сравнивается с критическим значением $F_{кр}(\alpha, v_1, v_2)$ при заданном уровне значимости α и степенях свободы $v_1 = m$, $v_2 = n$. Если $F > F_{кр}$, то нулевая гипотеза отвергается и делается предположение, что дисперсия, описываемая моделью МЛР, неслучайным образом отличается от дисперсии ошибок. Это означает, что рассматриваемая модель является адекватной исходным данным функции отклика.

Выбор системы эффективных предикторов непосредственно связан с проблемой построения оптимальной модели МЛР и в большинстве современных пакетов прикладных статистических программ (ППСП) выполняется одновременно. Процедура отбора наиболее существенных переменных может рассматриваться как процесс выбора размерности линейной модели. В настоящее время наиболее эффективным методом решения данной задачи считаются пошаговые процедуры, которые включают в себя несколько различных алгоритмов: метод включения переменных; метод исключения переменных; метод включения–исключения.

Суть метода включения переменных заключается в том, что на первом шаге выбирается наиболее коррелированный с функцией отклика предиктор и рассчитываются все параметры модели парной регрессии. После этого вычисляются, например, частные коэффициенты корреляции для оставшихся $m - 1$ предикторов, которые показывают «чистый» вклад каждой переменной в дисперсию функции отклика. Таким образом, выбирается вторая переменная, имеющая максимальный частный коэффициент корреляции, и строится новая модель $Y = f(X_1, X_2)$. Данная процедура может повторяться до тех пор, пока не будут построены все m моделей. Наиболее важным ее моментом является выбор оптимальной модели. В ППСП этот вопрос решается с помощью *частного F-критерия* [Юдин, 1972], который представляет собой обычный *F-критерий* для каждой переменной при условии, что она оказывается последней переменной, включенной в регрессию.

На каждом шаге происходит проверка адекватности модели и сравнение с некоторым критическим значением $F_{кр}$. Величина $F_{кр}$ задается априори. Как только величина F становится меньше $F_{кр}$, последний шаг принимается за оптимальную модель регрессии.

Метод исключения переменных реализует обратную процедуру. Вначале строится полная (из m переменных) модель МЛР. Затем из нее исключается наименее значимый (по частному *F-критерию*) фактор. После этого из модели исключается следующий по значимости фактор. Так может продолжаться до тех пор, пока не останется самый значимый фактор. Выбор оптимальной

модели также осуществляется по частному *F-критерию*, который на каждом шаге сравнивается с $F_{кр}$. При достижении условия $F > F_{кр}$ полученное уравнение МЛР считается оптимальным.

Метод включения–исключения в определенной степени является синтезом первых двух алгоритмов и заключается в том, что на каждом шаге одновременно с проверкой модели на адекватность по *F-критерию* осуществляется проверка всех коэффициентов регрессии на значимость по критерию Стьюдента. В результате из модели может быть исключена та переменная, которая имеет уровень значимости (*p-level*), превышающий заданное пороговое значение. Причем такое исключение осуществляется вне зависимости от того, когда переменная была включена в модель. Причина исключения может заключаться в тех связях, которые существуют между этой и другими переменными, содержащимися в модели. Такая проверка по *F-критерию* и *p-level* осуществляется до тех пор, пока не удастся исключить ни одну из переменных, содержащихся в текущей модели, а те переменные, которые находятся в резерве, не могут занять место в модели.

Достоинство пошаговых процедур состоит в простоте алгоритмов и возможности построения оптимального уравнения из очень большого числа потенциальных предикторов, а недостаток – в раздельном анализе переменных.

Анализ результатов прогноза осадков для Танзании и его уточнение.

Весь архив сезонных осадков был разбит нами на две выборки: зависимую (1980 – 1999 гг.), по которой рассчитывались и тестировались модели, и независимую выборку (2000 – 2004 гг.), используемую для оценки оправдываемости опытных прогнозов. Для построения прогностических моделей использовался аппарат множественного регрессионного анализа. Рассчитаны и протестированы уравнения множественной регрессионной модели для долгосрочного прогноза осадков в сезонах октябрь – декабрь (OND) и март – май (МAM) для каждого из 10 районов Танзании [Чанга, 2005, 2005a] по зависимой выборке на основе процедуры «исключения». Все модели содержат не более 5 предикторов с заблаговременностью от 1 до 11 месяцев и описывают подавляющую часть дисперсии осадков (от 58 до 87 % дисперсии). Основные параметры этих моделей приведены в табл. 1. Отметим, что все модели имеют высокую значимость по всем параметрам и являются адекватными по критерию Фишера (табл. 1). Так как отношение $\sigma_{Y(X)}/\sigma_Y$ существенно меньше 1, 0, то естественно ожидать успешных результатов при оценке этих моделей на независимых данных.

Таблица 1

Оценки параметров регрессионных моделей по зависимой выборке (1980 – 1999 гг.) для районов Танзании в сезоне октябрь – декабрь

Район	Параметр прогностической модели				
	R^2	F	$\sigma_{Y(X)}, \text{мм}$	$\sigma_Y, \text{мм}$	$\delta = \sigma_{Y(X)}/\sigma_Y$
1	0,79	19,9	61,8	123,3	0,50
6	0,72	13,8	139,6	242,7	0,58
7	0,83	13,9	50,5	105,9	0,48
9	0,85	20,8	45,8	104,1	0,44

Сопоставление опытных прогнозов осадков в сезоне OND с исходными данными по независимой выборке представлено на рис. 1 и в табл. 2. Видно, что в целом наблюдается хорошее соответствие прогнозируемых осадков с фактическими.

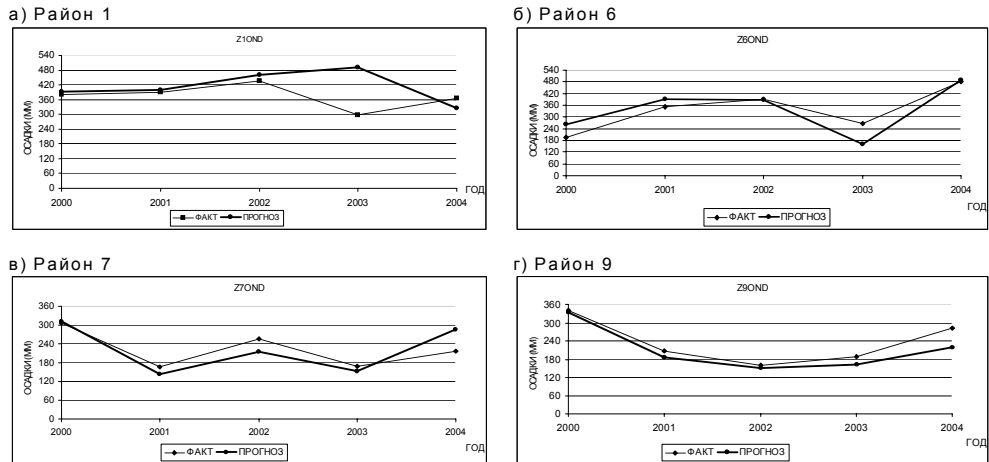


Рис. 1. Фактические и вычисленные по моделям значения осадков в сезоне октябрь – декабрь в районах 1 (а), 6 (б), 7 (в) и 9 (г).

Таблица 2

Сопоставление опытных прогнозов осадков (\tilde{Y}) с их фактическими значениями (Y) в сезоне октябрь – декабрь на независимой выборке

Год	Район 1			Район 6		
	Y	\tilde{Y}	$ Y - \tilde{Y} $	Y	\tilde{Y}	$ Y - \tilde{Y} $
2000	381,4	394,0	12,6	197,3	261,4	64,1
2001	388,7	399,0	10,3	353,8	393,2	39,4
2002	436,4	459,0	22,6	392,9	388,8	4,1
2003	296,7	491,0	194,3	267,3	161,5	105,8
2004	366,2	325,0	41,2	482,9	490,9	8
			δ 56,2	δ 44,3		
Год	Район 7			Район 9		
	Y	\tilde{Y}	$ Y - \tilde{Y} $	Y	\tilde{Y}	$ Y - \tilde{Y} $
2000	306,4	311,0	4,6	340,3	333,0	7,3
2001	166,4	142,1	24,3	206,9	185,2	21,7
2002	257,3	214,6	42,7	161,1	151,7	9,4
2003	167,9	152,0	15,9	188,8	162,3	26,5
2004	216,0	285,7	69,7	283,3	218,4	64,9
			δ 31,5	δ 26,0		

Примечание: $\delta = 1/N \sum_{i=1}^N |Y_i - \tilde{Y}_i|$ – средняя абсолютная ошибка прогноза [Хандожко, 2005].

Выводы

Впервые осуществлена разработка для Танзании физико-статистического метода прогноза сезонных сумм осадков большой заблаговременности. Полученные результаты дают основание рекомендовать разработанную методику при составлении оперативных прогнозов осадков и других элементов погоды для территории Танзании и Восточной Африки.

В процессе разработки физико-статистического метода долгосрочного прогноза сезонных осадков:

1. Выявлены наиболее информативные тест-предикторы для прогноза в сезоне OND, включающие индекс пассатного ветра на уровне 850 гПа в восточной, центральной и западной частях тихоокеанской области тропиков и первые четыре главных компонента ТПО Индийского океана в районе INDO3, INDO6 и INDO11 [Чанга, 2005, 2005а].

2. Проведено сопоставление опытных прогнозов осадков в OND с исходными данными по независимой выборке и показано, что в целом наблюдается хорошее соответствие прогнозных осадков с фактическими.

3. Высокое качество прогнозов с помощью оптимальных уравнений регрессии, полученных на зависимой выборке, и хорошее соответствие прогнозных и фактических осадков на независимой выборке позволяет рекомендовать данные модели для применения в оперативных условиях в Танзании.

В разработанных моделях мы использовали массивы всех более или менее доступных в сети *Internet* виртуальных предикторов. Представляется, что в перспективе целесообразно рассмотреть также: широтно-долготную миграцию и интенсивность антициклонических климатических центров действия в районах Маскаренских островов и острова Святой Елены; первые естественные колебания в полях движения, массы и температуры воздуха над районом действия глобального климатообразующего процесса ЭНЮК, а также в ТПО над всей тропической зоной Тихого океана. Безусловно, важно учесть: интенсивность и местоположение ВЗК; мощности юго-восточного муссона в районе Гвинейского залива и Северной Африки; ТПО и испаряемости в районе Аравийского моря; интенсивности субтропической зоны высокого давления над Южной Азией и многоцентральной депрессии над Мексикой, Центральной Америкой и Колумбией и тропосферных струйных течений над Южным полушарием (субтропического, полярного и антарктического), а также мощности летнего меридионального течения южного направления вдоль восточного побережья Африки и над западными акваториями Индийского океана на уровне 850 гПа (течение Финлейтера, $V > 12$ м/с), свидетельствующего о перетекании юго-восточного пассата в Северное полушарие и его трансформировании в юго-западный муссонный поток (в результате чего над этим регионом глобальный пояс облачности ВЗК оказывается разорванным на протяжении более 1000 км). Думается, необходимо рассматривать и влагосодержание различных масс воздуха, особенно формирующихся над океаниче-

скими акваториями, и ряд других крупномасштабных процессов, ответственных за формирование полей облачности и осадков к югу от экватора. Несомненно, определенную роль в этих процессах играют и такие факторы, как испаряемость, влажность почвы, количество осадков и др. Ясно, что подготовка данных по такому многомерному виртуальному вектору-предиктору требует серьезных усилий большого научного коллектива, современных быстроедействующих ЭВМ, реализации новейших алгоритмов статистической обработки экспериментальных данных и, возможно, результатов гидродинамического реанализа, а также иерархического комплекса решающих правил, используемых для отбора носителей долговременной памяти атмосферы.

Литература

1. Блажевич В.Г., Белянкина И.Г. Физико-статистический прогноз температуры и осадков на весенне-летний период по предикторам большой заблаговременности. // Труды ГГО, 1986, вып. 505, с. 150–154.
2. Блажевич В.Г., Меццарская А.В., Юдин М.И. Статистическая оценка информативности предикторов в зависимости от длины исходных рядов // Тр. V Всесоюз. совещания: Применение статистических методов в метеорологии. Л.: Гидрометеиздат 1987, с. 44 – 48.
3. Кендалл М.Д., Стьюарт А. Многомерный статистический анализ и временные ряды. – М.: Наука, 1976. – 736 с.
4. Малинин В.Н., Гордеева С.М. Канарский апвеллинг: крупномасштабная изменчивость и прогноз температуры воды. – СПб.: Гидрометеиздат, 2002. – 153 с.
5. Малинин В.Н., Гордеева С.М. Физико-статистический метод прогноза океанологических характеристик.– Мурманск: изд. ПИНРО, 2003. – 162 с.
6. Хандождо Л.А. Экономическая метеорология.– СПб.: Гидрометеиздат, 2005. – 489 с.
7. Чанга Л.Б. Климатическое районирование территории Танзании по данным за 1972 – 2002 гг. // Международная школа-конференция. Изменение климата и окружающая среда: СПб., РГГМУ, 6 – 9 декабря 2005 г. – СПб.: изд. РГГМУ, 2005, с. 100–104.
8. Чанга Л.Б. О потенциальных предикторах для долгосрочного прогноза осадков в Танзании // Международная школа-конференция. Изменение климата и окружающая среда: СПб., РГГМУ, 6 – 9 декабря 2005 г. – СПб.: изд. РГГМУ, 2005, с. 96–100.
9. Юдин М.И. Физико-статистические методы прогнозов погоды и возможность их внедрения. // Метеорология и гидрология, 1967, № 11, с. 39–49.
10. Юдин М.И. Физико-статистический метод долгосрочных прогнозов погоды. – Л.: Гидрометеиздат, 1968. – 28 с.
11. Юдин М.И., Меццарская А.В., Блажевич В.Г. Характеристика гидрометеорологической информации, использованной при долгосрочном физико-статистическом прогнозе температуры и осадков для районов недостаточного увлажнения // Труды ГГО, 1969, вып. 236, с. 45–63.
12. Юдин М.И., Меццарская А.В. Результаты применения физико-статистического метода прогноза осадков и температуры с большой заблаговременностью. // В кн.: Тр. V Всесоюз. метеорол. съезда. Т. 2. Л.: Гидрометеиздат, 1972, с. 83–94.
13. Юдин М.И., Ретинская Р.П. Прогноз внутримесячного хода давления физико-статистическим методом. // Метеорология и гидрология, 1974, № 1, с. 24–35.
14. Юдин М.И., Блажевич В.Г. Оценка значимости предикторов на основе комплексного статистического испытания. // Труды ГГО, 1975, вып. 329, с. 41–53.