Том СХУШ	Труды Всесоюзного научно-исследовательского института морского рыбного хозяйства и океанографии (ВНИРО)	1976

УДК 31:551.464.5 (262.54)

СТАТИСТИЧЕСКИЙ ПОДХОД К ЗАДАЧЕ ПОСТРОЕНИЯ ПРОГНОЗИРУЮЩИХ ФУНКЦИЙ СОЛЕНОСТИ АЗОВСКОГО МОРЯ

Л.А.Кучай

Проблема обработки наблюдений и их анализа в современных исследованиях весьма актуальна. Интерес представляет получение модели, обладающей максимальной простотой, минимальным числом параметров, и при этом адекватно описывающей наблюдения.

Цель предлагаемой работы - построение модели временного ряда наблюдений S_t , значения которого представляют собой среднегодовую величину солености Азовского моря (в $^0/oo$) за период с 1912 по 1969 г. (данные лаборатории гидрологии и гидрохимии (АзНИИРХа).

При выборе подхода для получения модели S_t были использованы статистические методы анализа временных рядов, в которых наблюдения зависимы, и характер этой зависимости интересен сам по себе.

Процесс построения модели состоит из следующих этапов /2/:

1) процесс идентификации (выбор класса модели);

2) подгонка идентифицированной модели к временному ряду (использование эмпирических данных для оценки параметров модели);

3) проверка модели для выявления неадекватности ее и выработки подходящих изменений.

Особенностью выбранного подхода анализа временных рядов является возможность исследования некоторых нестационарных процессов, а именно процессов со стационарными приращениями n -го порядка (время дискретное) /1/. Такие процессы, в частности, удовлетворяют разностной схеме:

$$W_t = \nabla^n S_t \qquad (t = 0, \pm 1, \pm 2, \ldots)$$

где S_t – исследуемый временной ряд;

∇ – разностной оператор

$(\nabla S_t = S_t - S_{t-1}, \nabla^2 S_t = \nabla S_t - \nabla S_{t-1}$ и т.д.);

n – порядок оператора ;

W_t – новый временной ряд, полученный в результате действия оператора ∇^n на исследуемый ряд.

Так как целью анализа временного ряда S_t является прогноз, существенно также и то, что подогнанная статистическая модель позволит определить дисперсию ошибок прогноза и вычислить пределы, в которых с заданной вероятностью будут лежать будущие значения ряда.

Идентификация. Методы идентификации, предложенные в работах /I/ и /2/, используют в качестве основного инструмента автокорреляционную функцию процесса. Автокорреляционная функция процесса, находящегося в статистическом равновесии, имеет тенденцию к затуханию с увеличением задержки. Отсутствие затухания выборочной автокорреляционной функции истолковывается в том смысле, что процесс ведет себя нестационарно, хотя возможно, что ряд ∇S_t , или разности более высокого порядка будут стационарными процессами.

В табл.I представлены значения выборочных автокорреляционных функций процессов. $S_t, \nabla S_t, \nabla^2 S_t$.

Таблица I

Значения выборочных автокорреляций процессов $S_t, \nabla S_t, \nabla^2 S_t$.

k	r_k		
	S_t	∇S_t	$\nabla^2 S_t$
I	0,868	0,I83	-0,45I
2	0,682	0,I2I	-0,055
3	0,453	-0,036	-0,08I
4	0,238	-0,094	-0,032
5	0,09I	-0,063	0,093
6	-0,023	-0,203	-0,0I3
7	-0,054	-0,280	-0,I60
8	-0,0I6	-0,II4	0,03I
9	0,065	0,0II	0,027
I0	0,I33	0,089	0,085
II	0,I68	0,0I7	-0,056
I2	0,I92	0,038	0,033

Медленный спад выборочных автокорреляций процесса S_t делает необходимым исследование процессов ∇S_t и $\nabla^2 S_t$ и их корреляционных функций.

Для ряда $\nabla^2 S_t$ автокорреляции малы уже для $k \geqslant 2$. Это указывает на то, что этот временной ряд может быть описан как процесс скользящего среднего первого порядка /2/, который представляет собой линейную систему, где входом служит последовательность независимых случайных величин a_t с фиксированным распределением (обычно нормальным с нулевым средним и дисперсией σ_a^2), а выходом - исследуемый ряд. В общем случае оператор системы выражает текущее значение выходного процесса как линейную комбинацию конечного числа величин a_t, т.е.

$$\nabla^2 S_t = a_t - \sum_{i=1}^{q} \theta_i a_{t-i},$$

где q - порядок процесса скользящего среднего;

θ_i - параметры системы.

В рассматриваемом случае q = I и, следовательно, ряд $\nabla^2 S_t$ идентифицируется моделью

$$\nabla^2 S_t = a_t - \theta a_{t-1}. \qquad (I)$$

<u>Оценка параметра модели</u>. Раскрывая оператор ∇^2 модели (I), получим

$$S_t - 2S_{t-1} + S_{t-2} = a_t - \theta a_{t-1},$$

т.е.

$$S_t = 2S_{t-1} - S_{t-2} + a_t - \theta a_{t-1}. \qquad (2)$$

Если ищется прогноз значения $S_{t+\ell}$, $\ell \geqslant$ I, говорят, что прогноз делается в момент t с упреждением ℓ. Из формулы (2) следует, что

$$S_{t+\ell} = 2S_{t+\ell-1} - S_{t+\ell-2} + a_{t+\ell} - \theta a_{t+\ell-1}. \qquad (3)$$

Сделанный в работе /2/ вывод о том, что прогноз с минимальной среднеквадратической ошибкой в момент t с упреждением ℓ есть условное математическое ожидание $S_{t+\ell}$ в момент t, дает возможность использовать модель в качестве рекуррентной формулы для прогнозирования будущих значений ряда.

Если рассматривать $S_{t+\ell}$ как функцию ℓ при фиксированном t, то $\hat{S}_t(\ell)$ - прогнозирующая функция для момента t:

$$\hat{S}_t(1) = 2S_t - S_{t-1} - \theta a_t \qquad \text{для } \ell = 1, \qquad (4)$$

$$\hat{S}_t(\ell) = 2\hat{S}_t(\ell-1) - \hat{S}_t(\ell-2) \qquad \text{для } \ell \geqslant 2. \qquad (5)$$

При этом последовательность a_t, которая генерирует рассматриваемый процесс S_t, является рядом остаточных ошибок при прогнозировании на шаг вперед /2/, т.е.

$$a_t = S_t - \hat{S}_{t-1}. \qquad (1).$$

Этот важный вывод используется для оценки параметра θ выбранной модели, так как совершенно очевидно, что для различных значений θ модель (I) аппроксимирует рассматриваемый ряд наблюдений S_t с различной точностью, и оптимальным будет то значение параметра θ, при которм величина Р (θ) станет минимальной

$$P(\theta) = \sum_{t=1}^{m} a_t^2(\theta), \qquad (6)$$

где N – количество членов рассматриваемого ряда;

n – порядок разностного оператора модели (I) (в рассматриваемом случае N = 20, n = 2).

Таблица 2

θ	$P(\theta)$
0,3	I4,26
0,35	I3,86
0,4	I3,50
0,45	I3,I9
0,5	I2,9I
0,55	I2,67
0,6	I2,49
0,65	I2,44
0,7	I2,70
0,75	I3,54
0,8	I5,24
0,85	I7,55

В табл.2 представлены значения величины $P(\theta)$ в зависимости от значения θ. Минимальное значение $P(\theta)$ достигается для θ = 0,65

Проверка модели в данной работе основана на анализе остаточных ошибок a_t. Для адекватной модели по мере увеличения длины ряда процесс a_t становится все ближе к белому шуму. Следовательно, последовательные значения a_t должны быть некоррелированы. При изучении автокорреляционной функции $r_k(a)$ воспользуемся общим критерием согласия /2/, который заключается в том, что если подгоняемая модель удовлетворительна, то величина

$$Q = m\sum_{k=1}^{R} r_k^2(a)$$

распределена приближенно как $\chi^2(R-1)$. В табл.3 приведены значения $r_k(a)$ для параметров θ = 65 и θ = 0,9, причем последний приведен для сравнения результатов подгонки модели.

Таблица 3

Выборочные автокорреляции $a_t(\theta)$

k	$r_k(a)$	
	θ = 0,65	θ = 0,9
I	0,0II	0,I7
2	-0,06	0,039
3	-0,I34	-0,076
4	0,II3	-0,II
5	-0,05	-0,057
6	-0,II	-0,I69
7	-0,3I	-0,339
8	-0,029	-0,083
9	0,037	0,004
I0	0,097	0,09
II	-0,02	0,029
I2	0,093	0,I36
I3	-0,03I	0,042
I4	0,I23	0,I69
I5	0,224	0,253
I6	0,20I	0,I89
I7	-0,044	-0,057
I8	-0,258	0,273
I9	-0,I32	-0,I92
20	-0,076	-0,I55
Q	20,32	27,8

Величины Q сравниваем с таблицей χ^2 с I9 степенями свободы. 90%- и 95%-ные квантили для χ^2 равны соответственно 27,2 и 30,I.

Несмотря на сильное различие параметров θ = 0,65 и θ = 0,9, последний почти укладывается в норму при проверке адекватности модели. Это вызывает сомнения и требует дополнительной проверки.

Если бы процесс a_t был белым шумом, то график его проинтегрированного спектра имел бы разброс относительно прямой, соединяющей точки (0;0) и (0,5; I). Из-за неадекватности модели ряд a_t становится неслучайным и его проинтегрированный спектр систематически отклоняется от ожидаемой прямой.

Несмещенная оценка проинтегрированного спектра есть /2/:

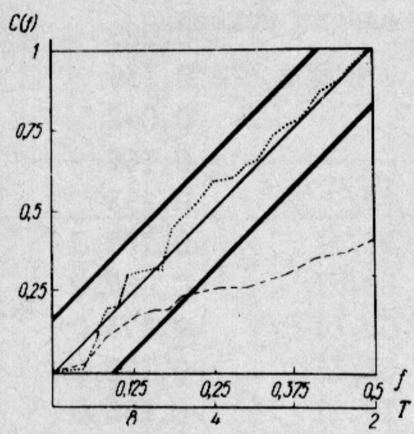
$$C(f_j) = \sum_{i=1}^{j} J(f_i)/m\sigma_a^2 ,$$

где σ_a^2 - дисперсия процесса a;

$$J(f_i) = \left[\left(\sum_{t=1}^{m} a_t \cos 2\pi f_i t\right)^2 + \left(\sum_{t=1}^{m} a_t \sin 2\pi f_i t\right)^2\right];$$

$f_i = \frac{i}{m}$ - частота.

Отклонение полученного спектра от ожидаемого для белого шума оценивается при помощи критерия Колмогорова. На рисунке приведены проинтегрированные спектры процессов a_t для $\theta = 0,65$ (точечная линия) и $\theta = 0,9$ (пунктирная линия). Эта дополнительная проверка показывает, что процесс a_t для $\theta = 0,9$ не обладает свойствами белого шума, и что оценка параметра θ была произведена верно.



Проинтегрированный спектр остаточных ошибок для подогнанной модели ряда

<u>Прогнозирование</u>. Наиболее удобной формой модели, используемой при прогнозировании, являются рекурентные формулы (4)-(5). Табл.4 представляет собой наглядную схему вычислений прогнозов по этим формулам. Доверительные интервалы, приведенные в схеме, строятся исходя из предположения, что a_t подчиняется нормальному закону распределения /2/.

Существует еще одна форма модели (I), которая может быть использована при прогнозировании, так как дает возможность выразить значение $S_{t+\ell}$ как бесконечно взвешенную сумму предыдущих наблюдений и импульса $a_{t+\ell}$ т.е.

$$S_{t+\ell} = \sum_{j=1}^{\infty} \alpha_j S_{t-j+\ell} + a_{t+\ell}. \qquad (7)$$

Веса α_j можно получить из обращенной формы модели (I) /2/.

В табл.5 представлены веса α_j для $\theta = 0,65$.

Т а б л и ц а 4

Схема вычисления прогнозов

t	Момент начала прогноза	S_t	a_t	Время упреждения ℓ					
				1	2	3	4	5	6
				Доверительные интервалы					
				50%					
				±0,57	±0,82	±1,08	±1,37	±1,67	±1,99
				95%					
				±1,66	±2,38	±3,15	±3,98	±4,86	±5,79
1946		9,9							
1947	Н	10,4							
1948		10,5	+0,4	10,9					
1949		11,3			11,2				
1950		11,6				11,5			
1951		11,9					11,8		
1952		12,1						12,8	
1953	Н	12,4							12,4
1954		11,8	+0,9	12,7					
1955		12,3			12,43				
1956		11,9				12,1			
1957		11,1					11,8		

Т а б л и ц а 5

Веса α_j для модели (7)

j	1	2	3	4	5
α_j	1,37	-0,137	-0,086	-0,054	-0,034
j	6	7	8	9	
α_i	-0,022	-0,140	-0,008	-0,005	-0,0

В табл.6 приведены значения прогнозов ряда S_t, просчитанные по весам α_j с 1971 по 1974 г. Здесь же для сравнения приведены значения S_t.

Таблица 6

Прогнозы S_t, рассчитанные с помощью весов α_j

Год	Прогноз S_t	Измерение S_t
1971	11,83	11,8
1972	12,20	12,3
1973	12,57	12,6
1974	12,87	12,9

Заключение

Приведенная модель является лишь начальным этапом в использовании статистического подхода для прогнозирования процесса S_t, однако она, безусловно, полезна для дальнейших исследований в этом направлении.

Список использованной литературы

1. Г.Дженкис, Д.Ваттс. Спектральный анализ и его приложение. М., "Мир", 1971, с.
2. Дж.Бокс, Г.Дженкис. Анализ временных рядов, прогноз и управление. М., "Мир", 1974, с.

A statistical approach to the problem of constructing forecasting functions of salinity in the Azov Sea.

L.A.Kuchai

Summary

The problem of modelling a time series is discussed. A parametric model with a single parameter representing the mean annual salinity in the Azov Sea is suggested. It may forecast the basic time series one or two years in advance. The adequacy of the model is checked up using the data of actual observations.