

УДК 68І.3 : 55І.579 (73)

ПРИМЕНЕНИЕ ПАРАМЕТРИЧЕСКОГО МЕТОДА ДЛЯ РАСЧЕТА БУДУЩИХ  
ЗНАЧЕНИЙ ВРЕМЕННЫХ РЯДОВ ГИДРОМЕТЕОРОЛОГИЧЕСКИХ  
ХАРАКТЕРИСТИК

С.Н.Родионов

Модели, используемые для описания процессов различной природы, можно подразделить на параметрические (для оценки небольшого числа параметров) и многопараметрические, или непараметрические (для оценки большого или бесконечного числа параметров). Конечно, это деление несколько условно.

Непараметрическим методом, применяемым при изучении физических процессов, является спектральный анализ. Этот метод нашел широкое применение (Григоркина и др., 1973). Однако он удобен лишь для описания процессов, образованных смесью синусоид и косинусоид с постоянными частотами, скрытыми в шуме. Процессы, изучаемые в океанологии, зачастую характеризуются случайными изменениями частоты, амплитуды и фазы. В этом случае выборочный спектр сильно флуктуирует и не допускает какой-либо разумной интерпретации (Дженкинс, Ваттс, 1971).

Для решения задач, связанных с временными рядами (главным образом задач прогнозирования и автоматического регулирования) лучше пользоваться параметрическим методом, разработанным Боксом и Дженкинсом (1974).

Существенная особенность этого метода состоит в том, что тренды и другие псевдоустойчивые характеристики, возможно меняющиеся во времени, рассматриваются скорее как статистические, а не детерминированные явления. Последнее особенно важно при анализе геофизических процессов.

Некоторые параметрические методы, в особенности те из них, которые учитывают нестационарность процесса, с успехом использовались в экономике (Бокс, Дженкинс, 1974). Однако в гидрометеорологической практике параметрические методы еще не нашли достаточно широкого применения.

В данной статье показана возможность использования разработанного Боксом и Дженкинсом параметрического метода для моделирования и прогноза процессов, изучаемых в гидрометеорологии.

### Краткое описание параметрического метода

Реализацию дискретного процесса, или временный ряд, для удобства рассмотрения можно представить в виде ряда отклонений:

$$X_1, X_2, \dots, X_t, \dots, X_N, \quad (I)$$

где  $X_t$  - отклонение  $t$ -го значения исходного ряда от его математического ожидания .

Будем считать, что  $X_0 = X_{-1} = X_{-2} = \dots = 0$ . Предполагается, что значения ряда (I) в любой момент  $t$  можно считать линейной комбинацией предыдущих  $\rho$  значений ряда (I):

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_\rho X_{t-\rho} + a_t \quad (2)$$

Свободный член  $a_t$  в зависимости от обстоятельств называется либо ошибкой подгонки модели (2) к ряду (I), либо добавочным импульсом в момент  $t$  .

Считается, что последовательность независимых импульсов  $a$  - это реализация случайной величины, распределенной по нормальному закону с нулевым средним значением и дисперсией  $\sigma_a^2$ . Такая последовательность в технической литературе называется белым шумом. Известно, что процесс, имеющий нормальную плотность распределения, полностью описывается своим средним значением, дисперсией и автокорреляционной функцией (Бокс, Дженкинс, 1974). Следовательно, процессы типа белого шума полностью описываются своей дисперсией.

Модель (2) называется процессом авторегрессии (АР) порядка  $\rho$  . Она содержит  $\rho + 2$  неизвестных параметра:

$$a_1, a_2, \dots, a_\rho, \mu, \sigma_a^2.$$

Значение  $X_{t-1}$  можно исключить из правой части выражения (2) подстановкой

$$X_{t-1} = \alpha_1 X_{t-2} + \alpha_2 X_{t-3} + \dots + \alpha_p X_{t-p-1} + a_{t-1}$$

Аналогичным образом можно исключить  $X_{t-2}, X_{t-3}$  и т.д., получив в результате выражение  $X_t$  в виде ряда из  $a$ . В случае, если  $X_t$  можно представить небольшим числом  $q$  предыдущих  $a_i$ , например

$$X_t = a_t - \beta_1 a_{t-1} - \beta_2 a_{t-2} - \dots - \beta_q a_{t-q} \quad (3)$$

такой процесс называется процессом скользящего среднего (СС) порядка  $q$ .

Модель (3) содержит  $q+2$  неизвестных параметра:

$$\beta_1, \beta_2, \dots, \beta_q, \mu, \sigma_a^2.$$

Часто для достижения большей гибкости в подгонке моделей к наблюдаемым временным рядам целесообразно использовать комбинированную модель авторегрессии - скользящего среднего (АРСС) порядка  $(p, q)$ :

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + a_t - \beta_1 a_{t-1} - \dots - \beta_q a_{t-q}, \quad (4)$$

в которой имеется  $p+q+2$  неизвестных параметра:

$$\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q, \mu, \sigma_a^2.$$

Рассмотренные выше три модели достаточно хорошо описывают стационарные процессы. Однако на практике гидрометеорологические процессы часто имеют нестационарный характер, что проявляется в различного рода трендах и в отсутствии фиксированного среднего значения. Правда, за исключением изменения уровня и наклона, поведение океанологических рядов оказывается во многом сходным. Для описания таких рядов Бокс и Дженкинс (1974) предлагают предварительно взять разность

$$Z_t = X_t - X_{t-1} = \Delta X_t.$$

Дженкинс и Ватс (1971) считают, что эта операция играет роль высокочастотного фильтра, поэтому после взятия определенного числа разностей ряд становится стационарным.

$$\text{Модель } Z_t = \alpha_1 Z_{t-1} + \dots + \alpha_p Z_{t-p} + a_t - \beta_1 a_{t-1} - \dots - \beta_q a_{t-q}, \quad (5)$$

где  $Z_t = \Delta^d X_t$  - значения ряда после взятия  $d$ -й разности, называется моделью авторегрессии - проинтегрированного скользящего среднего (АРПСС) порядка  $(p, d, q)$ . На практике значения  $p, d, q$  часто равны или меньше 3.

По Боксу и Дженкинсу (1974), взятие первых разностей позволяет учесть только уровень, а взятие вторых разностей — как уровень, так и наклон.

Параметрические модели строят в три этапа. На первом этапе находят порядок модели, т.е. значения  $p, d,$  и  $q$ . Этот этап носит название идентификации модели. Так как методы идентификации довольно грубы, часто выбирают две, а то и три пробные модели. На втором этапе определяют параметры  $\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q$  пробной модели. На третьем этапе, используя остаточные ошибки, проверяют, адекватна ли модель. Выбирают наилучшую из пробных моделей. Если обнаружены дефекты, пытаются их устранить методами, описанными ниже. Если же модель удовлетворяет критерию согласия, она готова к использованию для прогноза.

### Пример прогноза ряда температуры

Рассмотрим теперь применение этого подхода к ряду среднемесячных значений температуры с 1860 по 1972 г. в г. Сан-Диего (США). Модель подгоняется по значениям температуры с 1860 г по 1960 г.

Для определения порядка  $p$  и  $q$  модели воспользуемся автокорреляционной и частной автокорреляционной функциями. Понятие частной автокорреляционной функции состоит в следующем. Умножим все члены уравнения (2) на  $X_{t-k}$ . Тогда при условии, что ряд достаточно длинный, получим

$$\gamma_k = \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \dots + \alpha_p \gamma_{k-p} \quad \text{при } k > 0, \quad (6)$$

где  $\gamma_k$  — оценка автокорреляционной функции для задержки которая вычисляется по формуле

$$\gamma_k = \frac{1}{N} \sum_{t=1}^{N-k} X_t \cdot X_{t+k}.$$

Обозначим  $K$ -й коэффициент процесса авторегрессии порядка  $p$  через  $\alpha_{pk}$ , так что последним коэффициентом будет  $\alpha_{pp}$ . Тогда (6) приобретает вид

$$\gamma_k = \alpha_{p1} \gamma_{k-1} + \alpha_{p2} \gamma_{k-2} + \dots + \alpha_{pp} \gamma_{k-p}. \quad (7)$$

Величина  $\alpha_{pp}$  рассматриваемая как функция аргумента  $p$ , называется функцией частной автокорреляции.

Характерные особенности поведения автокорреляционной и частной автокорреляционной функций состоят в следующем.

1. В то время как автокорреляционная функция процесса АР ( $\rho$ ) спадает плавно, ее частная автокорреляционная функция имеет обрыв после задержки  $\rho$ .
2. Так как процессы АР и СС взаимно обратимы, автокорреляционная функция процесса СС ( $q$ ) обрывается после задержки  $q$ , в то время как ее частная автокорреляционная функция плавно спадает с ростом задержки.
3. Автокорреляционная функция смешанного процесса АПСС ( $\rho, q$ ) после первых  $q$ - $\rho$  задержек представляется в виде суммы экспонент и затухающих синусоид. Обратно частная автокорреляционная функция представляется суммой экспонент и затухающих синусоид после  $\rho$ - $q$  задержек.

На рис. I. показаны графики автокорреляционной и частной автокорреляционной функций исследуемого ряда температуры.

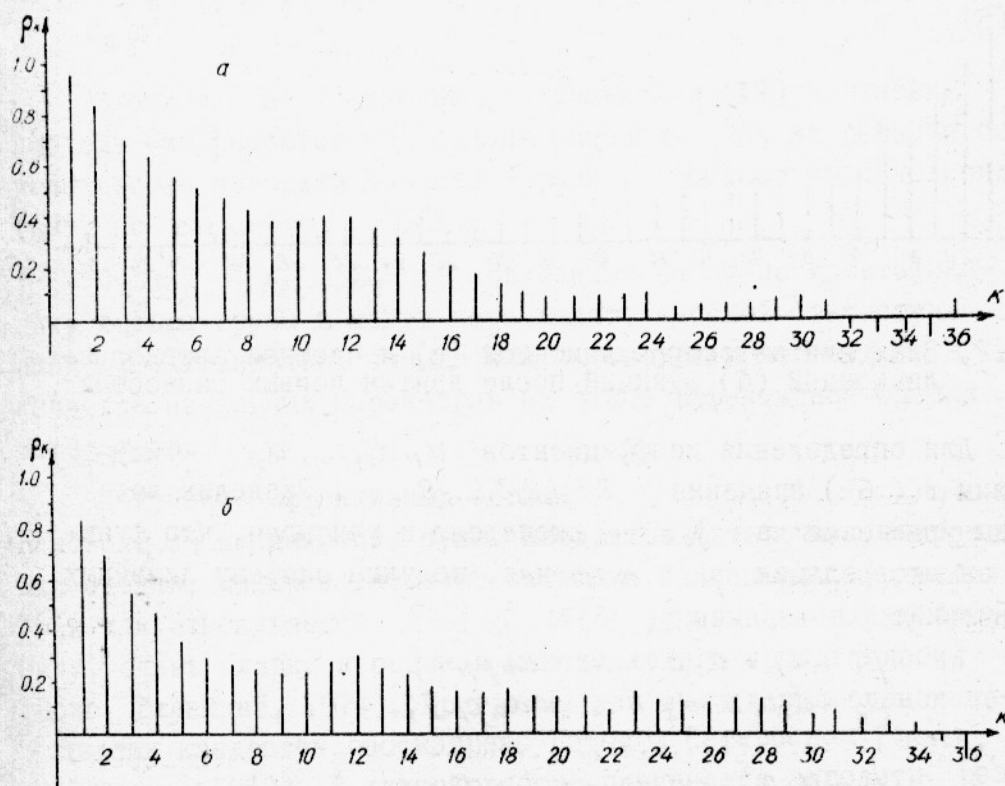


Рис. I. Значения автокорреляционной (а) и частной автокорреляционной (б) функций для исходного ряда температуры

Очень медленное затухание функции автокорреляции свидетельствует о наличии нестационарности процесса (Бокс, Дженкинс, 1974). На рис.2 эти же функции показаны после взятия первых разностей. Как следует из особенностей поведения этих функций, исследуемый процесс можно описать моделью АРПСС (0, I, I).

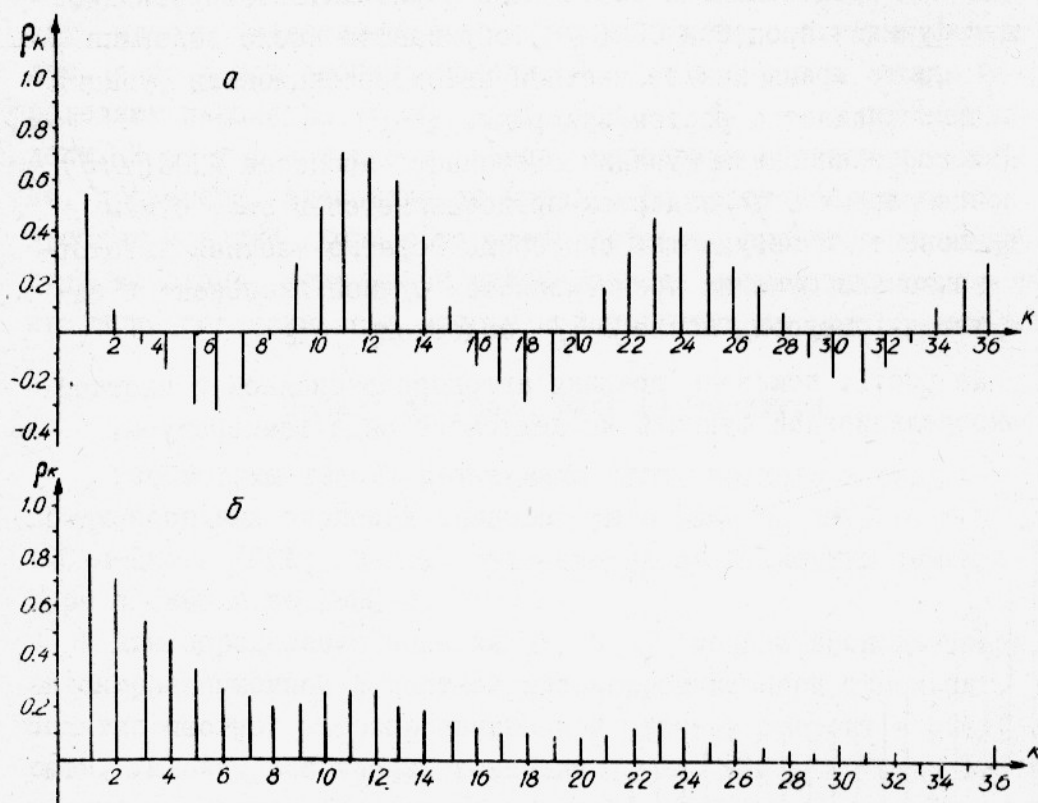


Рис.2. Значения автокорреляционной (а) и частной автокорреляционной (б) функций после взятия первых разностей

Для определения коэффициентов  $\alpha_1, \alpha_2, \dots, \alpha_p$  подставим в (6) значения  $k = 1, 2, 3, \dots, p$ . Разделив все члены уравнения на  $\gamma_0$  - дисперсию и учитывая, что функция автокорреляции  $\rho$  - четная, получим систему линейных уравнений:

$$\begin{aligned}
 \rho_1 &= \alpha_1 + \alpha_2\rho_1 + \dots + \alpha_p\rho_{p-1}; \\
 \rho_2 &= \alpha_1\rho_1 + \alpha_2 + \dots + \alpha_p\rho_{p-2}; \\
 &\dots \\
 \rho_p &= \alpha_1\rho_{p-1} + \alpha_2\rho_{p-2} + \dots + \alpha_p.
 \end{aligned}$$

(8)

Решив эту систему, найдем коэффициенты  $\alpha_1, \alpha_2, \dots, \alpha_p$ .

Аналогично находим коэффициенты  $\beta_1, \beta_2, \dots, \beta_q$ .

Автокорреляционная функция процесса  $CC(q)$  будет равна

$$\begin{aligned} \gamma_k = & \frac{1}{N} \sum_{t=1}^{N-k} (a_t - \beta_1 a_{t-1} - \dots - \beta_q a_{t-q}) \times \\ & \times (a_{t+k} - \beta_1 a_{t+k-1} - \dots - \beta_q a_{t+k-q}). \end{aligned} \quad (9)$$

Следовательно, дисперсия процесса равна

$$\gamma_0 = (1 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2) \sigma_a^2 \quad (10)$$

и

$$\gamma_k = \begin{cases} -\beta_k + \beta_1 \beta_{k+1} + \beta_2 \beta_{k+2} + \dots + \beta_{q-k} \beta_q, \\ \text{при } k = 1, 2, 3, \dots, q \\ 0, \text{ при } k > q. \end{cases} \quad (11)$$

Отсюда автокорреляционная функция имеет вид

$$\rho_k = \begin{cases} \frac{-\beta_k + \beta_1 \beta_{k+1} + \beta_2 \beta_{k+2} + \dots + \beta_{q-k} \beta_q}{1 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2}, \text{ при } k = 1, 2, 3, \dots, q \\ 0, \text{ при } k > 0. \end{cases} \quad (12)$$

В отличие от уравнений (8) уравнения (12) нелинейны, поэтому они решаются численными методами. Для их решения был использован алгоритм Ньютона-Рафсона, предложенный Вильсоном (цит. по Marquardt, 1963).

Оценки  $\beta_1, \beta_2, \dots, \beta_q$ , полученные на этапе идентификации, весьма грубы и могут не обладать высокой статистической эффективностью, однако они с успехом используются в качестве начальных параметров на этапе эффективной оценки параметров.

Получение эффективных оценок  $\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q$  основано на минимизации суммы квадратов остаточных ошибок. Эти оценки, которые называют оценками наименьших квадратов (так как отыскивается  $S = \sum \alpha_t^2 - \min$ ) оказываются, как правило, очень близки к оценкам максимального правдоподобия (Бокс, Дженкинс, 1974). Программы для вычисления оценок наименьших квадратов рассмотрены Ботом и Петерсоном (Both, Peterson, 1959). В данной работе применялся алгоритм, разработанный Марквардом (Marquardt, 1963).

В нашем примере параметр  $\beta_1$  получился равным 0,65, следовательно модель ряда температуры будет иметь вид

$$Z_t = a_t - 0,65a_{t-1}, \text{ где } Z_t = X_t - X_{t-1}. \quad (13)$$

На рис.2 хорошо заметны пики автокорреляционной функции на задержках 12, 24, 36, что свидетельствует о сезонной периодичности. Для учета периодичности Бокс и Дженкинс (1974) предлагают операцию взятия разности повторять с интервалом, равным периоду  $S$  :

$$X'_t = a_t - a_{t-s}, \quad (14)$$

где  $t = S+1, S+2, \dots, N-d$ .

В нашем примере период  $S = 12$ .

Для ряда  $X'_t$  заново выполняются этапы идентификации и находятся параметры  $\psi_1, \psi_2, \dots, \psi_p, \varphi_1, \varphi_2, \dots, \varphi_q$  модели

$$X'_t = \psi_1 X'_{t-S} + \psi_2 X'_{t-2S} + \dots + \psi_p X'_{t-pS} + \lambda_t - \varphi_1 \lambda_{t-S} - \dots - \varphi_q \lambda_{t-qS}. \quad (15)$$

В нашем примере  $P = 1, Q = 1, \psi = 0,13, \varphi = 0,71$ , т.е.

$$X'_t = 0,13 X'_{t-12} + \lambda_t - 0,71 \lambda_{t-12}. \quad (16)$$

Комбинируя уравнения (13) и (16), получим следующую модель временного ряда температуры:

$$X_t = X_{t-1} + 1,13 X_{t-12} - 1,13 X_{t-13} - 0,13 X_{t-24} + \\ + 0,13 X_{t-25} + \lambda_t - 0,65 \lambda_{t-1} - 0,7 \lambda_{t-12} + 0,46 \lambda_{t-13}, \quad (17)$$

где  $t = n, n+1, \dots, N$ ;  
 $n = p + P \times S + D \times S + d$ ;

$D$  - количество взятия разностей  $X'$ .

После того как модель идентифицирована и параметры оценены, подгоняемая модель подвергается диагностической проверке. Методы диагностической проверки описаны в работах Анскомба (Anscombe, 1961), Анскомба и Тьюки (Anscombe, Tukey, 1963), Даниеля (Daniel, 1959), Бокса и Дженкинса (1974). На этапе диагностической проверки решается, адекватна ли модель, и если нет, то что именно неадекватно в модели и как ее изменить. Для этого из уравнения (17), описывающего процесс, на-



ходят ошибки  $\lambda_t$ . Первоначально значения  $\lambda_t, \lambda_{t-12}, \lambda_{t-13}$  приравняются нулю. Вычисляется совокупный критерий согласия  $Q'$ , распределенный приближенно как  $\chi^2$  (Бокс, Дженкинс, 1974).

$$Q' = n \sum_{k=1}^k \rho_k^2(\lambda)$$

Сравнивая значения  $Q'$  с табличным значением  $\chi^2$  при  $\nu = k - p - q - P - Q$  степенях свободы, определяют, адекватна ли модель. Если  $Q' < \chi^2$ , то модель достаточно хорошо описывает временной ряд. В противном случае снова повторяются процедуры идентификации и определения параметров, но уже для значений  $\lambda_t$ . В нашем примере  $Q' = 35,7$ ,  $\chi_{30}^2 = 43,8$  для уровня значимости  $\varepsilon = 0,05$ , т.е. подобранная модель удовлетворительно описывает ряд температуры.

Будущие значения ряда легко прогнозировать по формуле (I7). Обозначим прогноз в момент  $t$  с упреждением  $l$  через  $\hat{x}_t(l)$ . Тогда для следующего момента

$$\hat{x}_t(1) = x_t + 1,13x_{t-11} - 1,13x_{t-12} - 0,13x_{t-23} + 0,13x_{t-24}.$$

Так как будущие значения  $\lambda_{t+l}$  неизвестны, они принимаются равными нулю:

$$\hat{x}_t(2) = \hat{x}_t(1) + 1,13x_{t-10} - 1,13x_{t-11} - 0,13x_{t-12} - 0,13x_{t-23}$$

и т.д.

Как только становится известно значение  $x_{t+1}$  прогноз подправляется:

$$\hat{x}_{t+1}(1) = x_{t+1} + 1,13x_{t-10} - 1,13x_{t-11} - 0,13x_{t-22} - 0,13x_{t-23} + \lambda_{t+1},$$

где  $\lambda_{t+1} = x_{t+1} - \hat{x}_t(1)$  и т.д.

На рис. 3 показаны истинные и прогнозируемые значения температуры.

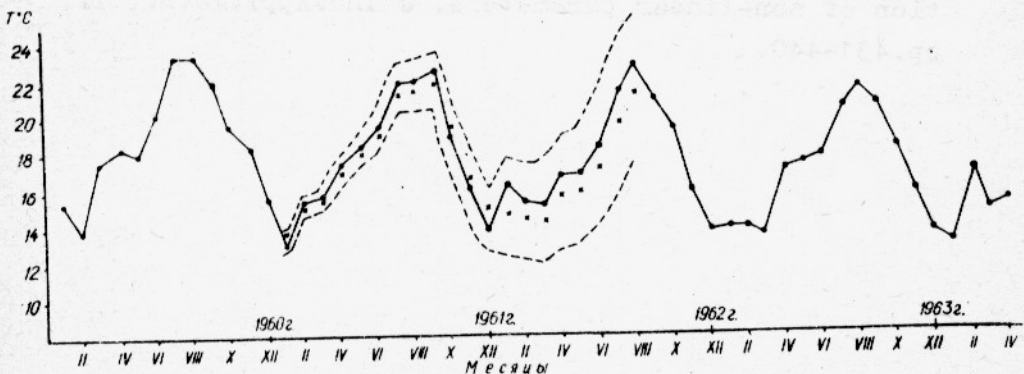


Рис. 3. Прогнозируемые (—•—) и фактические (××××) средне-месячные значения температуры. Пунктиром показаны 90%-ные вероятностные пределы

Модель временного ряда температуры оказалась простой и вместе с тем достаточно эффективной.

Программы для расчета порядка модели, значений коэффициентов и прогноза будущих значений ряда написаны нами на языке ФОРТРАН-IV для машины "Минск-32". Эти программы дают возможность анализировать периодичность и для других масштабов времени.

#### Л и т е р а т у р а

- Б о к с Дж., Д ж е н к и н с Г. Анализ временных рядов; прогноз и управление. М., "Мир", 1974, 405 с.
- Г р и г о р к и н а Р.Г., Г у б е р И.К., Ф у к с В.Р. Прикладные методы корреляционного и спектрального анализа крупномасштабных океанологических процессов. Л., изд-во ЛГУ, 1973, 170 с.
- Д ж е н к и н с Г., В а т т с Д. Спектральный анализ и его приложения. М., "Мир", 1971, 310 с.
- А н с о м б е, F.J. Examination of residuals, Proc. 4th Berkeley Symp., 1, 1961, p.1-36.
- А н с о м б е, F.J., Т у к е у, J.W. The examination and analysis of residuals, Technometrics, 5, 1963, p.141-160.
- В о т h, G.W., П е т е р с о н, T.J. Non-linear estimation, IBM Shore Program Pa. No. 618 WLNL1, 1968, p.1-12.
- Д а н и е л, С. Use of half normal plots in interpreting factorial experiments, Technometrics 1, 1959, p.311-322.
- М а р q u a r d t, D.W. On algorithm for least squares estimation of non-linear parameters. J. Ind. Appl. Math., II, 1963, pp.431-440.

Application of the parametric method for  
computation of future values of time series  
of hydrometeorologic characteristics

S.N.Rodionov

S u m m a r y

The application of the parametric method to predict time series of hydrometeorologic characteristics is shown with reference to modelling and prediction of future values of series of mean monthly temperatures of air in San-Diego, USA. Seasonal fluctuations in temperature and non-stationary series (non-constant means) are taken into consideration. Programmes written on the FORTRAN language make it possible to analyse also periodicity of other time periods.