

# 渤海月平均水位对水文气象诸因子的 季节响应及其双筛回归模式

陈上及 马继瑞 于继业

(国家海洋局海洋科技情报研究所, 天津)

## 摘 要

本文根据渤海沿岸海洋站1965—1986年的月平均水位资料,进行了最大熵谱、交谱和频率响应函数分析。结果表明,渤海月平均水位的年变化是海洋水文气象诸因子季节变化耦合响应的结果。并用双重筛选逐步回归分析,求得了回归预报方程,经回报验证表明,所得结果是满意的。

渤海的海平面年变化在中国近海是最为显著的,其变幅达60cm,比南海的大一倍多。尤其当最高海平面与最高高潮潮位、最大风暴潮三者重合时,就会出现特大的危险水位。例如,1985年8月塘沽的一次特大风暴潮,就是由于台风风暴潮与大潮高潮、最高月平均海平面相叠加,塘沽的最高水位远远超过警戒水位,使新港的货场、仓库、码头和许多民房被淹,造成巨大损失。可见,海平面的变化与沿海工农业生产和人民生命财产的安危关系极大。我国海洋学者对此至为关切。李坤平<sup>[1]</sup>曾对1969年渤海的逐时水位作过谱分析,发现渤海水位存在着8d周期的低频振动,其峰值与垂直于海岸的风应力、大气压力有着密切的关系。梅丽明、陈宗镛<sup>[2]</sup>对黄、东海沿岸各站为时一年(1975年10月1日—1976年9月24日)的日平均水位进行了谱分析,发现在2—60d周期范围内,存在着半月周期的明显谱峰。表明1.4—1/2月周期的水位低频振动,均属水位对大气强迫力的响应。本文利用营口、秦皇岛、塘沽、龙口等海洋站1965—1986年的月平均水位和水文气象资料,进行最大熵谱和交谱分析,进一步阐明渤海月平均水位对水文气象诸因子的季节响应,并用双重筛选回归方法,求得其预报模式。

## 一、渤海月平均水位对各水文气象因子的季节响应

选取营口、秦皇岛、塘沽和龙口4个海洋站连续20年(1965—1986年)的逐月平均海平面资料,进行了最大熵谱分析,最大熵谱阶数取30,分析结果如图1所示。各站的谱形十分相近,最大峰值均出现在0.083cpm/cpm表示周、月处。通过红噪声假设检验<sup>[3]</sup>,

表明渤海月平均水位的主要振动周期均为年周期。次大峰值出现在0.17cpm处，但在显著水平 $\alpha$ 取0.15的红噪声假设检验中，它们均未达到显著性标准。

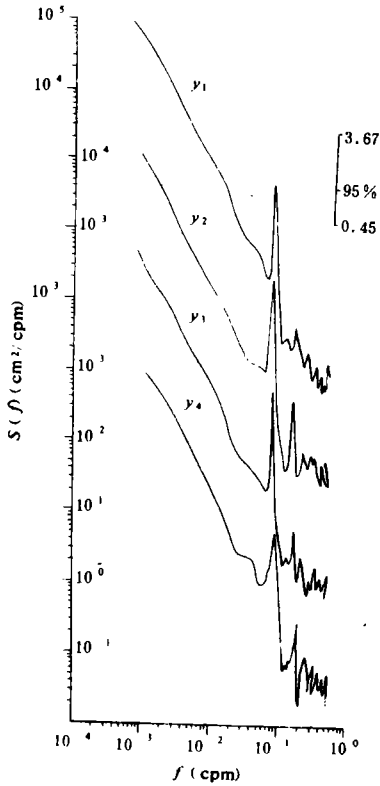


图1 营口( $y_1$ )、秦皇岛( $y_2$ )、塘沽( $y_3$ )和龙口( $y_4$ )月平均水位最大熵谱

为了分析月平均海平面季节振动的原因，对各站气压、气温、风速、降水、海面水温和盐度分别进行了最大熵谱估计。其中，风速分量，根据各站风力对水位显著作用方位的不同，营口取NE和W，秦皇岛取NE，塘沽取E，龙口取SW和W。因限于篇幅，图2只给出秦皇岛的气压、气温、水温，龙口的降水和W向风速及塘沽的E向风速等最大熵谱估计结果，其余谱形类同（从略）。红噪声假设检验结果（显著水平 $\alpha$ 取0.05）表明，渤海的气压、气温、水

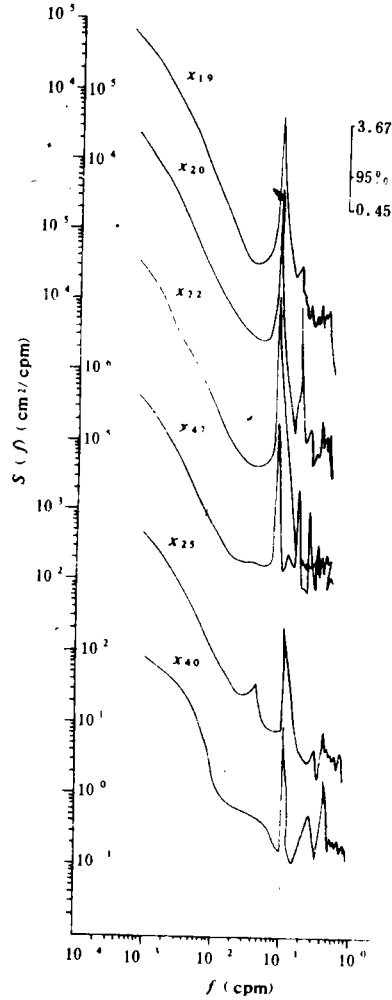


图2 诸水文气象因子的最大熵谱

$x_{19}$ : 秦皇岛气压谱( $\text{hPa}^2 \cdot \text{cpm}$ );  $x_{20}$ : 秦皇岛气温谱( $\text{C}^2/\text{cpm}$ );  $x_{22}$ : 秦皇岛水温谱( $\text{C}^2 \cdot \text{cpm}$ );  $x_{47}$ : 龙口降水量谱( $\text{mm}^2 \cdot \text{cpm}$ );  $x_{25}$ : 塘沽E向风频速因子 $f_b$ 谱( $\text{m}^2 \cdot \text{cpm}$ );  $x_{40}$ : 龙口W向风频速因子 $f_b$ 谱( $\text{m}^2 \cdot \text{cpm}$ )

温、降水和有效方位的风速，均以年周期振动最为显著。此外，气温、水温、龙口W方位的风速和降水在半年周期处的谱峰值，以及龙口降水在4个月周期处的峰值也均达到了显著性（显著水平 $\alpha$ 取0.05）要求。比较图1和图2可知，这些水文气象因子的谱图特征与

水位谱图非常相似, 相互对应。反映了月平均水位的变化受各水文气象因子的影响, 两者间有着密切关系。

水位与各水文气象因子的相关性, 可通过它们间的交谱分析进一步来揭示。分析结果表明, 各站水位与水文气象因子的年周期振动相关性极为显著, 年振动周期对应的频域相关系数(即凝聚谱函数)  $\bar{r}_{xy}(f)$  均大于 0.63, 超过了显著相关的临界值。如图 3 所示, 秦皇岛的气压、气温与秦皇岛、龙口的水位之间的凝聚函数达 0.9 以上; 龙口降水与塘沽、营口水位之间的凝聚函数也可达 0.85。可见渤海水位年周期变化主要是由水文气象因子的年周期振动所致。除了水位与气压呈反相位振动外, 与其他要素均为同相位变化。各站水位的变化, 从季节尺度来说几乎是同步的, 如表 1 所示, 两站间的相位差仅  $2^\circ$ — $18^\circ$ 。

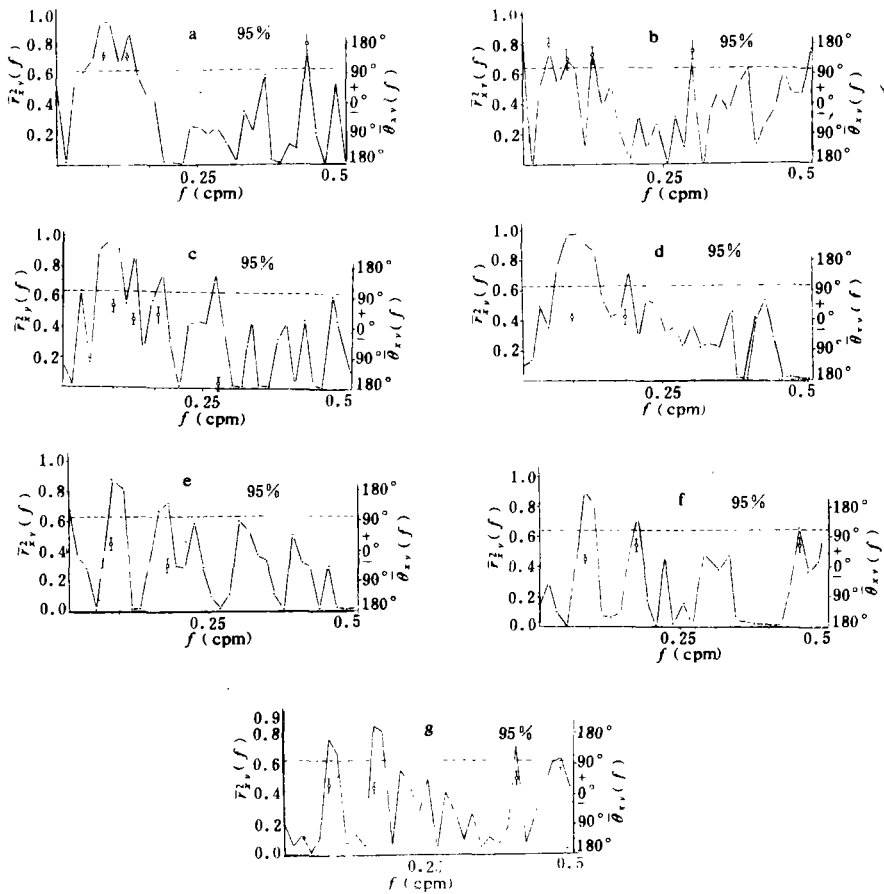


图 3 各站水文气象因子与水位间的凝聚谱  $\bar{r}_{xy}(f)$  和相位谱  $\bar{\theta}_{xy}(f)$

(a) 秦皇岛气压—秦皇岛水位; (b) 秦皇岛气压—龙口水位; (c) 秦皇岛气温—秦皇岛水位; (d) 秦皇岛气温—龙口水位; (e) 龙口降水—塘沽水位; (f) 龙口降水—营口水位; (g) 龙口西风风速—秦皇岛水位

月平均水位对各水文气象要素的响应, 还可从频率响应函数中看出。今以各站水位对秦皇岛各气象因子、塘沽东风和龙口降水的响应为例, 计算了各频率响应函数。结果表明,

表 1 各站水位的相关函数及相位差

相关测站	营口—龙口	营口—秦皇岛	营口—塘沽	龙口—塘沽	秦皇岛—塘沽	秦皇岛—龙口
凝聚函数	0.970	0.990	0.990	0.970	0.996	0.970
相位差(度)	342.0	358.0	1.7	18.3	3.4	345.0

最大增益因子几乎多出现在年振动周期的最大谱峰值频域处。图 4 仅列出部分计算结果，说明营口和塘沽水位的年周期振动，对龙口降水的相应振动的响应是显著的。综上分析，可见渤海月平均水位的年变化是对各水文气象因子年变化的耦合响应。

## 二、双重筛选回归模式

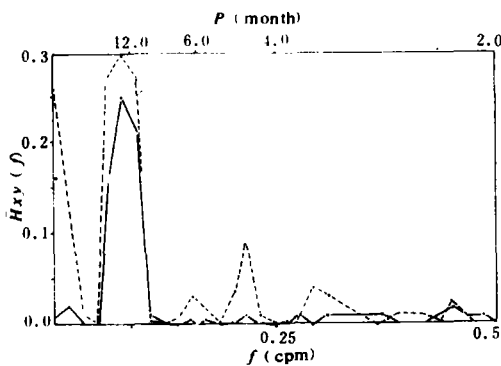


图 4 营口(实践)、塘沽(虚线)水位对龙口降水的频率响应函数

重筛选逐步回归分析方法<sup>[4]</sup>，把 $x_i$ 、 $y_j$ 都作为变量来筛选，进行多对多的逐步回归。其中， $i = 1, 2, \dots, m$ ， $j = 1, 2, \dots, p$ 均表示变量的序号。

### (一) 基本方法

用矩阵 $X$ 、 $Y$ 分别表示各影响因子 $x_1, x_2, \dots, x_m$ 及各站水位 $y_1, y_2, \dots, y_p$ 的 $n$ 次测值，即

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}, \quad Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix}.$$

若 $y_1, \dots, y_p$ 与 $x_1, \dots, x_m$ 之间有线性关系式

既然渤海月平均水位是受水文气象诸因子耦合响应的结果，建立回归方程时，若只考虑单个因子与水位的关系，由单因子求得 1 对 1 的回归方程，其效果必然很差。当然，用多元逐步回归，求得多对 1 的回归方程要好得多，但它只考虑 $\{x_i\}$ 与 $\{y_j\}$ 的关系，未考虑 $\{y_j\}$ 本身相互间的关系，其回归效果也是不够理想的。事实上，由各站水位的交谱分析表明，各站间水位的相关系数几乎均大于 0.97(表 1)，其相关性是非常好的。说明建立水位的回归模式时，考虑 $\{x_i\}$ 与 $\{y_j\}$ 相关的同时，必须考虑 $\{y_j\}$ 本身的相关。本文应用双

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} \beta_{01} & \beta_{11} & \cdots & \beta_{m1} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{0p} & \beta_{1p} & \cdots & \beta_{mp} \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix}, \quad (1)$$

则矩阵  $Y$  与  $X$  之间有关系式

$$E(Y) = (\mathbf{1}_n X) \begin{pmatrix} \beta_0 \\ \beta \end{pmatrix}, \quad rk(\mathbf{1}_n X) = m + 1, \quad (2)$$

I.

$\mathbf{1}_n$  是元素全为 1 的  $n$  维向量,  $Y$  中的各行  $y_{(1)}, y_{(2)}, \dots, y_{(n)}$  独立、正态、同协差阵  $\Sigma$ . 选取自变量  $x_i$  与选取因变量  $y_j$  时, 可归结为模型 I 与下列模型 II、III 之间的关系:

$$E(Y) = \begin{pmatrix} \mathbf{1}_n & X & u \\ n \times p & n \times m & n \times 1 \end{pmatrix} \begin{pmatrix} \beta_0^* \\ \beta^* \\ \beta_u \end{pmatrix}, \quad rk(\mathbf{1} \times u) = m + 2, \quad (3)$$

II.

$Y$  中的各行  $y_{(1)}, y_{(2)}, \dots, y_{(n)}$  独立、正态、同协差阵  $\Sigma$ .

$$E \begin{pmatrix} Y & z \\ n \times p & n \times 1 \end{pmatrix} = (\mathbf{1}_n \ X) \begin{pmatrix} \beta_0 & \beta_{0z} \\ \beta & \beta_z \end{pmatrix}, \quad rk(\mathbf{1} X) = m + 1. \quad (4)$$

III.

$$(Yz) \text{ 的各行独立、正态、同协差阵 } \begin{pmatrix} \Sigma & \sigma \\ \sigma' & \sigma_{zz} \end{pmatrix}.$$

选入自变量时, 以式 (2) 作为原始模型, 式 (3) 为增加一个自变量  $u$  后的模型. 选入因变量时, 以式 (2) 作为原始模型, 式 (4) 为增加一个因变量  $z$  后的模型. 反之, 剔除自变量或因变量时, 只需将式 (3) 或式 (4) 看作原始模型, 把式 (2) 作为剔除  $u$  或  $z$  后的模型, 它们间的关系式不变. 剔除变量时, 均用扫除变换和矩阵求逆四块公式, 求得

模型 I 中参数  $\beta_0$  与  $\beta$  的最小二乘估计

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \bar{x}' \hat{\beta}, \text{ 其中 } \bar{y} = \frac{1}{n} Y' \mathbf{1}, \\ \hat{\beta} = L_{xx}^{-1} L_{xy}, \text{ 其中 } L_{xy} = X' (I - \frac{1}{n} \mathbf{1} \mathbf{1}') Y, \end{cases} \quad (5)$$

$L_{xx} = X' (I - \frac{1}{n} \mathbf{1} \mathbf{1}') X$ ,  $L_{xx}^{-1}$  是  $L_{xx}$  的逆矩阵,  $I$  是单位矩阵.

模型 II 中的参数为:

$$\begin{cases} \hat{\beta}_0^* = \bar{y}' - \bar{x}' \hat{\beta}^* - \bar{u} \hat{\beta}_u, \text{ 其中 } \bar{u} = \frac{1}{n} u' \mathbf{1}, \\ \hat{\beta}^* = \hat{\beta} - L_{xx}^{-1} L_{xu} \hat{\beta}_u, \text{ 其中 } L_{xu} = X' \left( I - \frac{1}{n} \mathbf{1} \mathbf{1}' \right) u, \\ \hat{\beta}_u = L_{uu}^{-1} (X) L_{uy} (X), \text{ 其中 } L_{uu} (X) = L_{uu} - L_{ux} L_{xx}^{-1} L_{xu} \end{cases} \quad (6)$$

$$L_{uy} (X) = L_{uy} - L_{ux} L_{xx}^{-1} L_{xy}, \quad L_{uu} = u' \left( I - \frac{1}{n} \mathbf{1} \mathbf{1}' \right) u,$$

$$L_{uy} = u' \left( I - \frac{1}{n} \mathbf{1} \mathbf{1}' \right) y.$$

模型Ⅲ中的参数为:

$$\begin{cases} \hat{\beta}_{0z} = \bar{z} - \bar{x}' \hat{\beta}_z, \text{ 其中 } \bar{z} = \frac{1}{n} z' \mathbf{1}, \\ \hat{\beta}_z = L_{xx}^{-1} L_{xz}, \text{ 其中 } L_{xz} = X' \left( I - \frac{1}{n} \mathbf{1} \mathbf{1}' \right) z. \end{cases} \quad (7)$$

## (二) 所用资料

所用资料, 取自营口 ( $y_1$ )、秦皇岛 ( $y_2$ )、塘沽 ( $y_3$ ) 和龙口 ( $y_4$ ) 1970—1975 年的月平均水位、气压、风、气温、降水、水温、盐度。其中, 风向对上述港口水位影响较显著的 NE、E、SW 和 W 4 个方位, 并以各个方位的频率  $f$  与其平均风速  $v$  之积  $fv$  为风频速因子, 以各方位频率与其平均风速的平方  $v^2$  之积  $fv^2$  为风频能因子。这样, 共取 53 个水文气象因子参加筛选 (详见表 2)。

表 2 各影响因子及其符号

影响因子	营 口 $y_1$	秦 皇 岛 $y_2$	塘 沽 $y_3$	龙 口 $y_4$
NE 向的 $fv$	$x_1$	$x_{11}$	$x_{24}$	$x_{37}$
E 向的 $fv$	$x_2$	$x_{12}$	$x_{25}$	$x_{38}$
SW 向的 $fv$	$x_3$	$x_{13}$	$x_{26}$	$x_{39}$
W 向的 $fv$	$x_4$	$x_{14}$	$x_{27}$	$x_{40}$
NE 向的 $fv^2$	$x_5$	$x_{15}$	$x_{28}$	$x_{41}$
E 向的 $fv^2$	$x_6$	$x_{16}$	$x_{29}$	$x_{42}$
SW 向的 $fv^2$	$x_7$	$x_{17}$	$x_{30}$	$x_{43}$
W 向的 $fv^2$	$x_8$	$x_{18}$	$x_{31}$	$x_{44}$
月平均气压		$x_{19}$	$x_{32}$	$x_{45}$
月平均气温	$x_9$	$x_{20}$	$x_{33}$	$x_{46}$
月降水量	$x_{10}$	$x_{21}$	$x_{34}$	$x_{47}$
月平均水温		$x_{22}$	$x_{35}$	$x_{48}$
月平均盐度		$x_{23}$	$x_{36}$	$x_{49}$
月平均水位	$x_{50} = y_1$	$x_{51} = y_2$	$x_{52} = y_3$	$x_{53} = y_4$

### (三) 筛选过程和最后结果

先选入因变量 $y_4$ , 再剔选自变量; 自变量剔选完了再剔选因变量; 因变量剔选好了, 再剔选自变量, 这样反复交替剔选, 直至再无因变量和自变量可选入时为止, 此时, 第一组因变量和自变量的筛选已告完成。再进行第二组因变量和自变量的筛选。筛选时, 显著性检验的临界值, 取 $F_{进} = F_{出} = 3.5$ 。

筛选结果, 将渤海 4 个水位站分成两组:

第一组 秦皇岛 ( $y_2$ ) 与龙口 ( $y_4$ ) 划为一组, 对其影响的主要因子共有 7 个, 即 $x_5$ 、 $x_{10}$ 、 $x_{19}$ 、 $x_{20}$ 、 $x_{22}$ 、 $x_{25}$  和 $x_{39}$ , 不相关系数 $u = 0.036 29$ 。求得 $y_2$ 和 $y_4$ 的回归方程分别为:

$$\begin{cases} y_2 = 37.412 17 + 0.001 19x_5 + 0.014 39x_{10} - 0.035 85x_{19} \\ \quad - 0.036 53x_{20} + 0.030 1 x_{22} - 0.045 79x_{25} + 0.093 34x_{39} \\ y_4 = 36.515 64 + 0.000 81x_5 + 0.012 12x_{10} - 0.034 25x_{19} \\ \quad - 0.030 62x_{20} + 0.026 37x_{22} - 0.047 8 x_{25} + 0.072 1 x_{39} . \end{cases} \quad (8)$$

第二组 营口 ( $y_1$ ) 与塘沽 ( $y_3$ ) 同归这一组, 它们的共同影响因子有 8 个:  $x_1$ 、 $x_4$ 、 $x_6$ 、 $x_{11}$ 、 $x_{22}$ 、 $x_{40}$ 、 $x_{46}$  和 $x_{47}$ , 不相关系数 $u = 0.093 59$ 。 $y_1$ 与 $y_3$ 的回归方程分别为:

$$\begin{cases} y_1 = 1.599 79 - 0.048 72x_1 + 0.065 29x_4 + 0.059 63x_6 - 0.068 59x_{11} \\ \quad - 0.032 02x_{22} - 0.002 18x_{40} + 0.009 2 x_{46} + 0.048 47x_{47}, \\ y_3 = 1.322 46 + 0.019 79x_1 - 0.114 88x_4 - 0.007 8 x_6 + 0.160 07x_{11} \\ \quad + 0.008 3 x_{22} + 0.139 73x_{40} + 0.004 36x_{46} + 0.009 35x_{47} . \end{cases} \quad (9)$$

根据上述方程所得的月平均水位估计值与实测值之差, 算得的剩余标准差 $\sigma$  列于表 3。

表 3 双筛回归效果

	秦 皇 岛 $\Delta y_2$	龙 口 $\Delta y_4$	营 口 $\Delta y_1$	塘 沽 $\Delta y_3$
$u$	0.036 29		0.093 59	
$\sigma$	0.068 9	0.067 3	0.058 6	0.055 1

由该表可知, 各站的最大剩余标准差均小于 0.07, 双筛回归效果要比多对 1 的逐步回归好得多。以营口为例, 逐步回归的 $\sigma = 0.087 6$  m, 双筛回归的 $\sigma = 0.058 6$  m, 预报精度有了明显提高。尤其是塘沽的月平均水位, 估计值的剩余标准差 ( $\sigma$ ) 仅为 0.055 m。由此可见, 基于月平均水位是水文气象因子的耦合响应的出发点, 将这些因子和各站水位进行双重筛选, 所得的回归模式, 其预报精度要比其他方法好得多<sup>[5]</sup>。

### 三、讨论和结语

1. 月平均水位对海洋水文气象诸因子的耦合响应是导致渤海海平面季节变化的主要原因。各影响因子与水位的相关系数(凝聚函数)多在0.9以上。水位和诸影响因子功率谱的最大峰值,以及水位对诸因子响应的最大增益因子均出现在年周期的峰值频率处。水位除与气压呈反相位振动外,与其他影响因子均呈同相位变化。

2. 双筛回归分析结果表明,水位的年变化,在秦皇岛与龙口之间,营口与塘沽之间,有着密切的联系,各有其共同的影响因子,可分别建立两组回归模式。这是由于渤海特有的自然地理条件所决定的。秦皇岛和龙口几乎同位于渤海的横轴腹线上,两站的连线直指渤海海峡,港外海区开阔,所受的水文气象影响条件相似。因而,在式(8)的 $y_2$ 与 $y_4$ 两式中,各项回归系数的符号均相同。而营口和塘沽分别位于辽东湾和渤海湾的湾顶或端点,此两海湾分别属于西南向和东北向半封闭海湾,受东北风、东风和偏西风作用,易产生风增水和风减水效应。这一特征也可从回归模式(9)中看出, $y_1$ 和 $y_3$ 两式的前6项回归系数是反映风和水温的效应的,其符号在两式中均为异号,说明该回归模式充分反映了辽东湾、渤海湾的风增(减)水效应有反相的趋势。

3. 双筛逐步回归充分考虑了各站水位与各站水文气象因子间以及各站水位之间的关系,所建立的回归模式,机制关系清楚,预报精度较高,具有实用意义。

此外,必须指出,所用双筛逐步回归是把各变量作为离散随机变量来考虑的。若应用混合自回归模型进行多维时间序列分析,可明确各影响因子的主次关系,回归效果也可能会有所提高。这有待于进一步探索。

### 参 考 文 献

- [1] 李坤平,渤海低频水位振动谱,海洋学报, 6 (1984), 4: 433—440.
- [2] 梅丽明、陈宗镛,黄、东海沿岸海面低频振动对大气强迫力的响应,海洋学报, 10 (1988), 4: 389—398.
- [3] Ma Jirui and FuShijie, An estimation of the confidence interval and methods of significance period tests in the maximum entropy spectral analysis, *Collected Oceanic Work*, 11 (1988), 42—50.
- [4] 张尧庭、汪剑平,双重筛选的算法和依据,数学的实践与认识, 4 (1986), 60—70.
- [5] 常征、巴兰春,消除海平面变化中“干扰”因子影响的“绝对方法”,海洋通报, 2 (1983), 3: 22—27.\*
- [6] Ulrych, T. J. and T. N. Bishop, Maximum entropy spectral analysis and autoregression decomposition, *Reviews Geophysics and Space Physics*, 113 (1975), 1: 183—200.
- [7] Jenkins, G. M. and D. G. Watts, *Spectral Analysis and Its Application*, Holden-Day, 1968.