

非平稳时间序列的组合模型及其 在中长期水文预报中的应用

韩光文 谢维敏
(华中工学院)

摘 要

基于水文序列是一类周期性非平稳时间序列的特点,本文提出了一种适合描述这一特点的组合模型.这个组合模型的优点在于它能较充分地利用历史数据中蕴含的信息,即它既能利用以年为采样周期的时间序列的相关信息,又能利用以月为采样周期的时间序列的相关信息;编制了 FORTRAN 语言程序,将它应用于河流的中长期水文预报.从文中所列出的汉口和龙滩水文站的向前 1 至 12 个月的水文预报结果可以看出,组合模型的预报效果是较好的.

一、非平稳时间序列的组合模型

$\{y_t\}_1^N$ 为某一非平稳时间序列, N 为样本长度. 如果在 $\{y_t\}_1^N$ 中含有明显的趋势分量和周期分量,那么,这一非平稳时间序列,可以一般地用下述组合模型予以描述,即:

$$\begin{cases}
 y_t = F_t + P_t + x_t, & (1.1) \\
 F_t = \sum_{\sigma=0}^s f_{\sigma} t^{\sigma}, & (1.1a) \\
 P_t = \sum_{i=1}^l P_t^{(i)}, & (1.1b) \\
 P_t^{(i)} = g_t^{(i)} + h_t^{(i)}, & (1.1c) \\
 g_t^{(i)} = g_{m_i}^{(i)} = \frac{l_i}{N} \sum_{j=1}^{N/l_i} z_{T_i(j-1)+m_i}, \quad m_i = 1, 2, \dots, T_i, & (1.1d) \\
 h_t^{(i)} = h^{(i)}(m_i, j) = (1 - \Phi_{m_i}(z^{-1}))h^{(i)'}(m_i, j) + (\Theta_{m_i}(z^{-1}) - 1)a_j, \\
 m_i = 1, 2, \dots, T_i, \quad j = 1, 2, \dots, \frac{N}{T_i}, & (1.1e) \\
 x_t = \frac{\Theta(z^{-1}, t)}{\Phi(z^{-1}, t)} a_t. & (1.1f)
 \end{cases}$$

方程 (1.1) 描述了组合模型的总框架及其信息结构,即 y_t 可分解成趋势分量 F_t , 若干不同频率的周期分量 $P_t^{(i)}$ 之和 P_t , 零均值的平稳时间序列分量 x_t ; 而 $g_t^{(i)}$ 和 $h_t^{(i)}$ 又分别是 $P_t^{(i)}$ 的确定性部分和随机性部分. 为了便于描述周期性分量,在组合模型中,采用了参

变量 j 和 m_i , 其定义为

$$t \triangleq T_i(j-1) + m_i. \quad (1.2)$$

式中, $j \triangleq \left[\frac{t-1}{T_i} \right] + 1$, $m_i \triangleq t - T_i \left[\frac{t-1}{T_i} \right]$, $[d]$ 表示取 d 的整数部分, T_i 是第 i 个周期分量的周期, $i = 1, 2, \dots, l$.

结合水文数据的特点, 且不失建模的一般性, 现在令 $l = 1$ (如果还有其它不明显的周期分量的话, 其影响可以让 $h_i^{(i)}$ 予以补偿), 这样 $P_i^{(i)}$ 可简记为 P_t , $g_i^{(i)}$ 、 $h_i^{(i)}$ 、 $h^{(i)}(m_i, j)$ 、 $h_i^{(i)'}(m_i, j)$ 、 m_i 和 T_i 分别简记为 g_t 、 h_t 、 $h(m, j)$ 、 $h'(m, j)$ 、 m 和 T . 以月水文数据为例, 它的周期是 12 个月, 即 $T = 12$, m 就是月份数, 随着时间的增加, 它总是重复着由 1 变到 12 的循环数, j 是第 j 年, 因此 (1.2) 式的含意是: 时点 t 就是第 j 年的第 m 月份.

(1) 趋势分量 F_t 的建模

趋势分量用多项式模型 (1.1a) 予以描述, 其系数 f_σ , $\sigma = 0, 1, \dots, s$ (s 通常取 2 或 3) 用逐步多项式回归方法予以确定^[1]. 并为下一步建模生成 $\{z_t = y_t - F_t\}_1^N$.

(2) 周期分量 P_t 中确定性部分 g_t 的建模

g_m 的周期 T 可以通过检验数据的相关性来确定, 也可以根据经验直接确定, 如月水文数据可取 $T = 12$.

g_m 用离散的周期函数 (1.1d) 予以拟合, 因为

$$g_t = g(T(j-1) + m) = g_m, \quad m = 1, 2, \dots, T, \quad (1.3)$$

所以 g_m 在 $m = 1, 2, \dots, T$ 各时点的值反映了 $\{y_t\}$ 的周期特征. 有了 g_m 的模型, 就能够生成 $\{h'_t = z_t - g_t\}_1^N$.

(3) 周期分量 P_t 中随机性部分 h_t 的建模

由 $h'_t = z_t - g_t$ 知:

$$h'(m, j) = h'_t = z_{T(j-1)+m} - g_m, \quad m = 1, 2, \dots, T.$$

注意到 (1.1d) 式, 即知: $E\{h'(m, j)\} = 0$, $m = 1, 2, \dots, T$, 所以 T 个以年为采样周期的序列 $\{h'(m, j)\}$, 至少可以认为是均值平稳的. 于是, 就能把 $h'(m, j)$ 表示成在零均值白噪声序列 $\{a_j\}$ (因为 $h'(m, j)$ 的均值为零) 作用下, 具有传递函数 $\Theta_m(z^{-1})/\Phi_m(z^{-1})$ 的线性时不变系统的输出, 即

$$h'(m, j) = \frac{\Theta_m(z^{-1})}{\Phi_m(z^{-1})} a_j, \quad m = 1, 2, \dots, T, \quad (1.4)$$

式中, $\Theta_m(z^{-1}) = 1 - \sum_{k=1}^{g_m} \theta_k^m z^{-k}$, $\Phi_m(z^{-1}) = 1 - \sum_{k=1}^{p_m} \varphi_k^m z^{-k}$, $z^{-k} x_t = x_{t-k}$. 现今

$$h(m, j) = h'(m, j) - a_j, \quad j = 1, 2, \dots, \quad (1.5)$$

并注意到 (1.4) 式, 立得组合模型中的 (1.1e) 式.

应该指出: 就水文序列而言, $h'(m, j)$ 是以年为采样周期的平稳随机过程, 它是以年为采样周期的众多独立随机变量的影响叠加的结果. 根据著名的中心极限定理: 在一定条件下, 独立随机变量之和的分布 (尽管每一个有任意的分布), 当变量无限增加时, 将趋于正态分布. 因此, $h'(m, j)$ 及与它同分布的 a_j 均可近似认为是正态的.

显然, 当 $h(m, j) = 0$ 时, 则 $h'(m, j) = a_j$ 为白噪声, 它说明周期分量的确定性部分

g_m 已最好地反映了 $\{y_t\}$ 的周期性特征。为下一步建模, 现生成 $\{x_t = h'_t - h_t\}_1^N$, 由于

$$\begin{aligned} E\{x_t\} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N (h'_t - h_t) = \frac{1}{T} \sum_{m=1}^T \left[\lim_{N \rightarrow \infty} \frac{T}{N} \sum_{j=1}^{N/T} (h'(m, j) - h(m, j)) \right] \\ &= \frac{1}{T} \sum_{m=1}^T (Ea_j) = 0, \end{aligned}$$

所以 $\{x_t\}$ 至少是均值平稳的。

(4) 平稳时间序列 $\{x_t\}$ 的建模

应指出的是, $a_j = h'(m, j) - h(m, j)$, 对于年份 j 而言是白噪声, 在相邻的 j 之间, a_j 是不相关的; 但 $x_t = h'_t - h_t$, 对于相邻的 t (月份) 而言, x_t 之间则可能是相关的, 这一相关成分往往为预报未来提供重要信息。由于 $\{x_t\}$ 至少是均值平稳的时间序, 基于 (3) 中所述的同样理由, 可以用 ARMA 模型 (1.1f) 描述 x_t , 并考虑了其参数的时变性质¹⁾。显然, a_j 也可认为是正态的。

组合模型的上述论述, 充分证明了其在结构上的合理性, 及其能充分利用 $\{y_t\}$ 中的信息的优点。

关于组合模型的辨识方法, 可参阅文献 [2—5]。

二、组合模型在河流水文预报中的应用

笔者建立了红水河上的天生桥、龙滩、岩滩、大腾峡和长江上的汉口、宜昌等水文站的组合模型。预报结果表明, 组合模型在中长期水文预报中, 是有实际应用价值的。这里仅以龙滩站和汉口站为例, 将组合模型的预报效果作一简单介绍和分析。

1. 衡量月水文预报的标准

根据长江流域规划办公室水文局的建议: 用 RVE 衡量中长期水文预报的精度, 如果有 75% 以上的预报值的 $RVE < 30\%$, 则其预报精度是可以接受的。RVE 定义为

$$RVE = \frac{|e_t|}{\Delta y}$$

式中, $e_t = y_t - \hat{y}_t$, $\Delta y = y_m^{\max} - y_m^{\min}$, $m = 1, 2, \dots, 12$, y_t 为实际值, \hat{y}_t 为预报值, y_m^{\max} 和 y_m^{\min} 分别是历年第 m 月份中水文数据中的最大值和最小值。

程序中采用下述四个指标来检验预报精度, 它们是: (1) $RVE < 30\%$ 的百分比;

(2) $RVE < 20\%$ 的百分比; (3) 平均 RVE $\left(= \frac{1}{N} \sum_{t=1}^N \frac{|e_t|}{\Delta y} \right)$; (4) 最大

$$RVE \left(= \max_i \left(\frac{|e_t|}{\Delta y} \right) \right).$$

2. 预报结果及其分析

这里 x_t 是用定常 ARMA 模型进行拟合的。

1) 在水文序列中, F_t 和 P_t 的时变性不明显, 所以, 可以把它们均看成是时不变的, 而将其可能的时变因素均归结到 x_t 的模型中。

表 1 汉口站向前 1 至 12 步的预报精度

向前预报步数	RVE < 20%	RVE < 30%	平均 RVE	最大 RVE
1	93.750%	97.917%	10.234%	31.994%
2	91.667%	97.917%	10.400%	32.192%
3	87.500%	95.833%	10.866%	32.060%
4	85.416%	95.833%	11.039%	32.536%
5	85.417%	95.833%	11.044%	32.854%
6	85.417%	95.833%	11.047%	32.779%
7	85.417%	95.833%	11.049%	32.828%
8	85.417%	95.833%	11.074%	32.804%
9	85.417%	95.833%	11.074%	32.801%
10	85.417%	95.833%	11.074%	32.804%
11	85.417%	95.833%	11.074%	32.805%
12	85.417%	95.833%	11.074%	32.804%

表 2 龙滩站向前 1 至 12 步的预报精度

向前预报步数	RVE < 20%	RVE < 30%	平均 RVE	最大 RVE
1	91.666%	93.750%	11.937%	35.210%
2	81.250%	91.667%	13.810%	36.963%
3	81.250%	89.583%	14.131%	38.674%
4	79.167%	89.583%	14.500%	37.706%
5	79.167%	89.583%	14.641%	38.400%
6	79.167%	89.583%	14.734%	39.259%
7	79.167%	87.500%	14.739%	39.169%
8	77.083%	87.500%	14.768%	39.028%
9	77.083%	87.500%	14.770%	39.087%
10	77.083%	87.500%	14.776%	39.179%
11	77.083%	87.500%	14.764%	39.191%
12	77.083%	87.500%	14.757%	39.137%

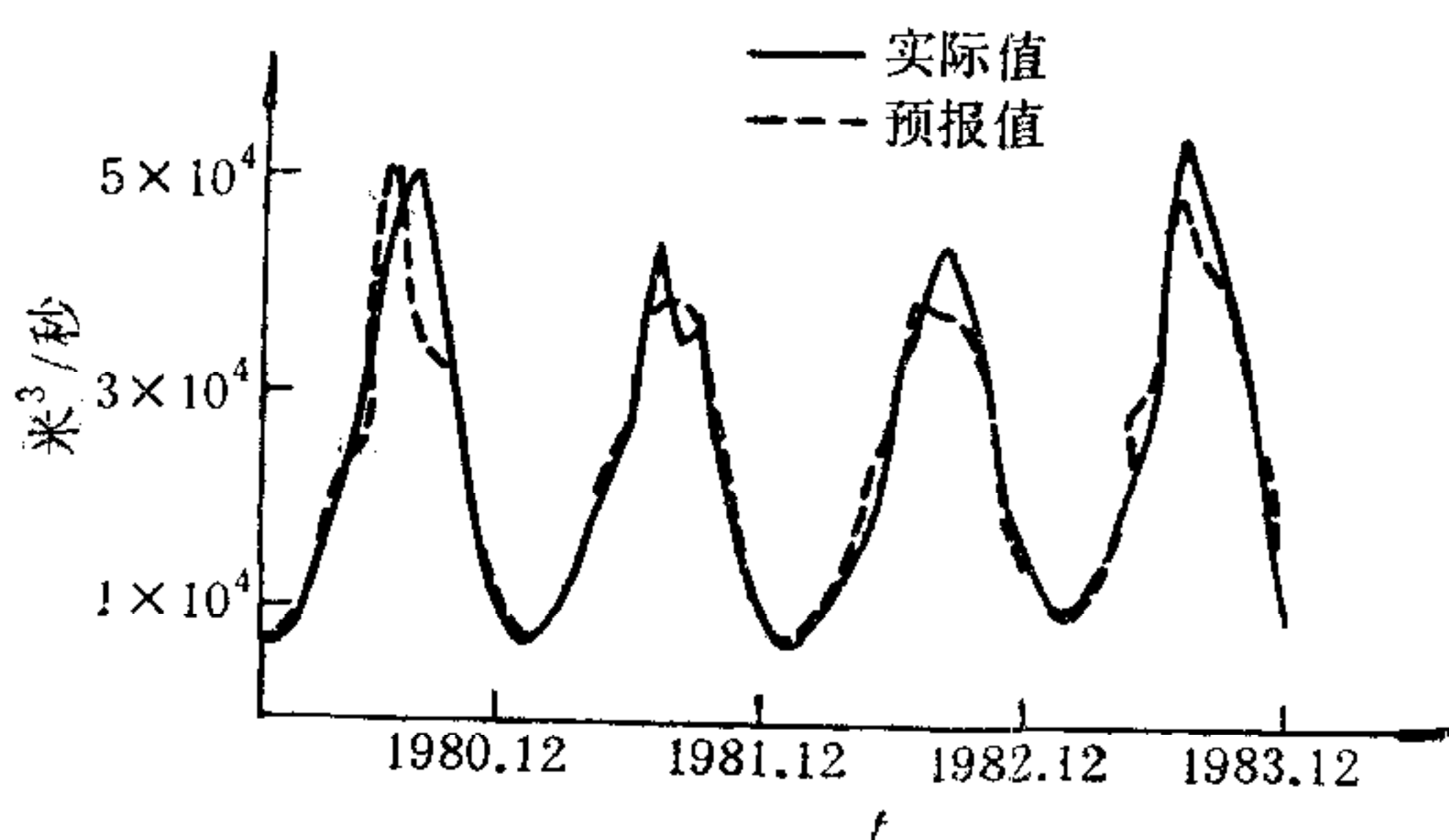


图 1

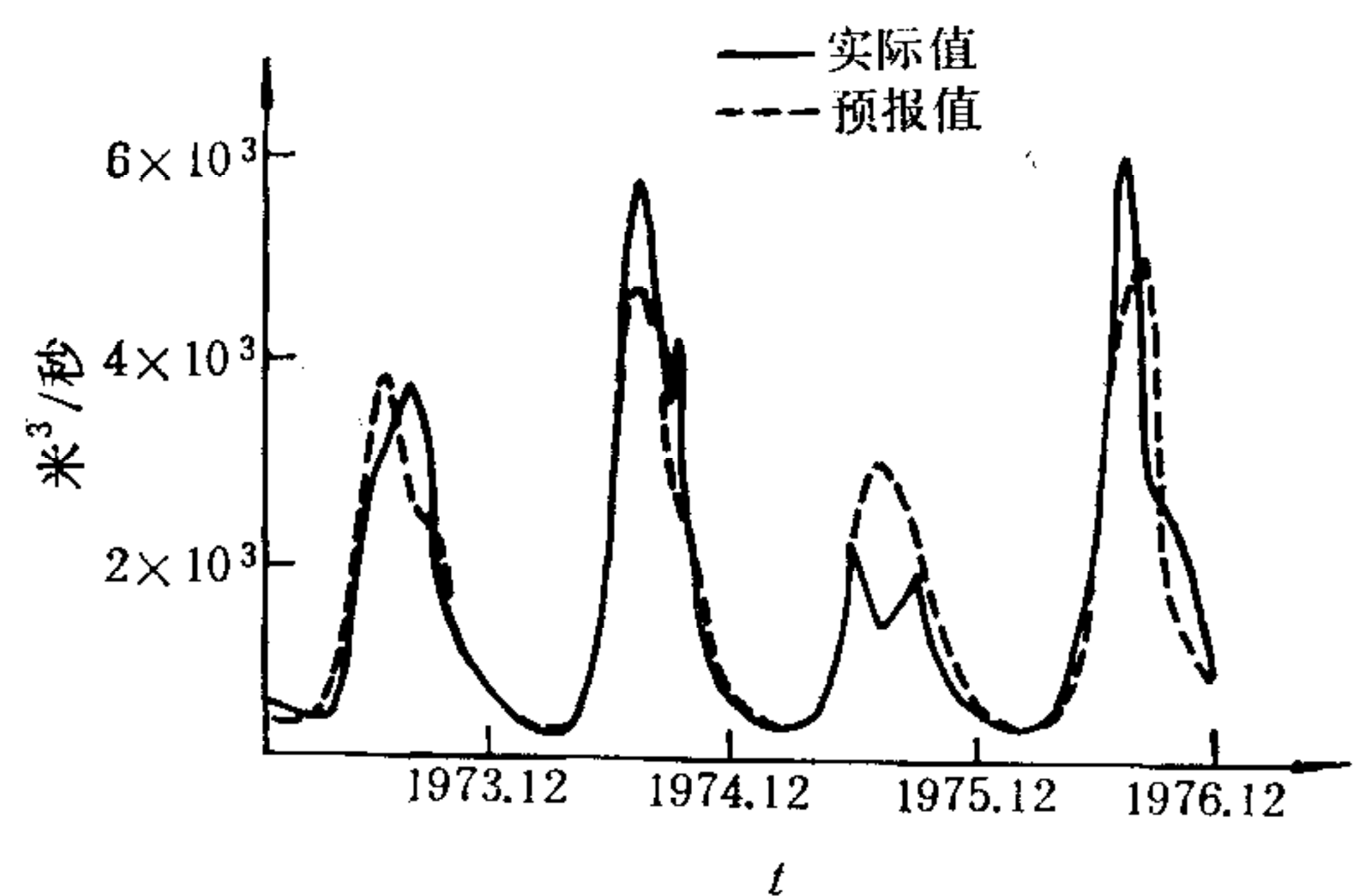


图 2

从表 1 和表 2 所示的结果可以看出,组合模型的预报精度是较高的。值得指出的是:预报步数小于 $T(=12)$ 时,组合模型的结构特点保证了预报精度不随预报步数的增加而明显降低,因为预报步数增加对 F_t 、 g_m 和 $h(m, j)$ 都没有影响。多步预报精度的降低,是 x_t 模型的预报精度随着预报步数增加而减少所引起的。

组合模型的预报效果, 还可以从图 1 和图 2 所示的汉口站和龙滩站月流量的实际值与预报值的对比中看出.

参 考 文 献

- [1] 茆诗松等, 回归分析及其试验设计, 华东师范大学出版社, 1981.
- [2] 安鸿志等, 时间序列的分析与应用, 科学出版社, 1983.
- [3] 韩光文, 辨识与参数估计, 国防工业出版社, 1980.
- [4] 韩志刚, 动态系统预报的一种新方法, 自动化学报, 9(1983).
- [5] Box, G. E. P., Jen Kins, G. M., Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, 1970.

THE COMBINATORY MODEL OF NON-STATIONARY TIME SERIES AND ITS APPLICATION TO MEDIUM-LONG TERMS RIVER FLOW FORECASTING

HAN GUANGWEN XIE WEIMIN

(*Huazhong Institute of Technology*)

ABSTRACT

Based on the nature that river flow series is a kind of seasonal non-stationary time series, a proper combinatory model describing this nature is developed in this paper. The advantage of the combinatory model is that information in historical data can be fully utilized, i.e., the correlated information of time series sampled both yearly or monthly can all be used. FORTRAN programs have been written to forecast medium-long terms flow of river by means of this combinatory model. The monthly mean flow forecasted one to twelve months ahead of time at Hankou Station on the Long River and Longtan Station on the Hongshuhe River show that the forecasting effect is satisfactory.

The work is part of a water resource research work mandated by the Ministry of Hydro-power.