

文章编号: 1006-2467(2023)01-0036-09

DOI: 10.16183/j.cnki.jsjtu.2021.300

基于 MWSA 的热力系统单参数时序预测方法

肖鹏飞^a, 倪 何^a, 金家善^b

(海军工程大学 a. 动力工程学院; b. 船舶与海洋工程学院, 武汉 430033)

摘 要: 热力系统的状态参数变化可以实时反映系统的运行状态, 针对热力系统参数运行数据预测手段匮乏的现状, 基于 4 种算法提出一种单参数预测方法并简称 MWSA, 对当前设备状态参数进行分解降噪、趋势提取和时序预测, 并将预测结果作为下一步运行管理策略和装备维修的参考, 对系统的长期安全稳定运行具有重要意义. 首先, 利用中值回归经验模态分解(MREMD)方法将监测得到的运行状态参数分解为若干个本征模态函数(IMF)和残余分量. 然后, 对不符合筛选条件的分量进行小波阈值降噪(WTD), 并将去噪后的分量与原本符合筛选条件的分量重组成为新的 IMF 分量. 最后, 利用基于奇异值分解(SVD)和优化参数排列熵(PE)的 K-means 聚类算法, 对重组后的 IMF 分量进行分类, 取熵值较低的一类分量重构为趋势项并采用整合滑动平均自回归模型(ARIMA)进行预测. 经实际案例验证, 该方法能够有效克服原始参数时序中高频噪声的干扰, 与不采用降噪处理的同类方法相比, 该方法预测的准确度更高.

关键词: 时序预测; 模态分解; 阈值降噪; 聚类算法; 自回归模型

中图分类号: N 37; TK 221

文献标志码: A

Sequential Prediction Method of Single Parameter for Thermal System Based on MWSA

XIAO Pengfei^a, NI He^a, JIN Jiashan^b

(a. College of Power Engineering; b. College of Naval Architecture and Marine Engineering, Naval University of Engineering, Wuhan 430033, China)

Abstract: The changes in the status parameters of the thermal system reflect the operating status of the system in real time. The forecast results of the trend extraction and time series prediction of the current equipment status parameters can be used as a reference for the next operation management strategy and equipment maintenance, which can be used for the long-term system safe and stable operation. In this paper, a method which is described as MWSA, based on the midpoint and regression based empirical mode decomposition (MREMD), the wavelet threshold denoising (WTD) and midpoint and techniques, the singular value decomposition (SVD) and optimized parameter permutation entropy (PE), and an auto regressive integrated moving average model (ARIMA), is applied to the single-parameter time series prediction of thermal systems. First, the MREMD is used to decompose the monitored operating state parameters into a number of intrinsic mode functions (IMF) and residual components. Next, the components that do not meet the screening conditions are subjected to wavelet thresholding. The denoised components and the components that originally meet the filtering conditions are recomposed into new IMF

收稿日期: 2021-08-05 修回日期: 2021-10-06 录用日期: 2021-12-10

作者简介: 肖鹏飞(1996-), 硕士, 从事热力系统的状态监测与故障诊断研究.

通信作者: 倪 何, 副教授; E-mail: elegance2006@sina.com.

components. Finally, the K -means clustering algorithm based on SVD and PE is used to classify the recomposed IMF components, the component with a lower entropy value is selected and reconstructed into a trend item, and ARIMA is used to predict. An actual case verifies that this method can effectively overcome the interference of high-frequency noise in the original parameter timing, and the prediction accuracy is higher than that of similar methods without noise reduction treatment.

Key words: sequential prediction; mode decomposition; threshold denoising; clustering algorithm; autoregressive model

热力系统具有部件众多、结构复杂和工况多变等特点,设备故障的耦合性强且排故成本高^[1]. 为了提高系统的全寿命健康管理水平,如何通过系统的历史运行参数预测系统未来一段时间内的运行状态,并将得到的预测结果作为系统运行管理策略和装备维修的参考,已成为当前热力系统科学管理的重要研究方向之一^[2-3].

热力系统的运行参数大多为非平稳时间序列. 在非平稳时间序列趋势项的提取方面,文献[4]对风速时序数据进行预测并基于排列熵(PE)的经验模态分解(EMD)趋势项提取算法,结合人工神经网络(ANN)的预测模型对提取的趋势进行预测,通过计算预测数据与实际数据的相关系数来验证预测精度,相比未引入排列熵算法的 EMD 结合 ANN 的 EMD-ANN 算法具有更好的预测效果;陈亮等^[5]提出基于集合经验模态分解(EEMD)和奇异值分解(SVD)的排列熵非平稳时序趋势提取方法,并通过与文献[6]和文献[7]中的方法进行对比,验证了该算法的优越性,但由于 EEMD 分解后的各个本征模态函数(IMF)含噪信息不一,且在分解后的噪声分量和信号分量的分界问题方面选择直接去掉最高频分量的方法而未做量化分析,导致最高频 IMF 分量中的信号信息丢失,其余分量未进行降噪处理,使后续趋势提取误差增加;刘海江等^[8]提出利用 EMD 结合小波阈值去噪(WTD)的降噪方法,在保留原信号信息的基础上对噪声进行了更有效地消除,但因 EMD 算法中的“端点效应”问题导致重构信号误差较大;黄礼敏^[9]在 EMD 算法的基础上提出了中值回归经验模态分解(MREMD)方法,有效解决了 EMD 分解过程中的“端点效应”问题,并且优化了包络线生成方式,但同样未对分解后的 IMF 分量进行定量分界,导致后续计算结果误差较大. Behera 等^[10]利用整合滑动平均自回归(ARIMA)模型,通过对公共部门每天接到的电话次数进行预测,并与

实际数据对比分析,验证了模型的有效性,其预测结果可作为部门管理层规划员工日常工作量的参考. 但由于热力系统中的参数历史数据中包含较多干扰,直接采用 ARIMA 模型预测误差较大;仇琦等^[11]在 ARIMA 预测模型的基础上提出了基于 EMD 结合 ARIMA 的 EMD-ARIMA 算法预测模型,通过对光电功率信号经 EMD 分解后的预测验证了模型的有效性,但这两种方法都未对原始非平稳信号的趋势进行提取,导致预测干扰较大.

基于上述研究提出一种热力系统单参数时序预测模型,由 MREMD、WTD、基于奇异值分解和排列熵的 K -means 聚类算法(SVD-PE-KCA)和 ARIMA 共 4 部分组成,简称 MWSA. 利用对原始信号的分解、降噪和趋势提取,使预测结果更加精确,可为热力系统的运行管理、控制优化和维修保障研究提供底层输入,具有一定的理论创新和工程应用价值.

1 预测步骤和算法

基于 MWSA 的单参数时序预测流程如图 1 所示.

首先,利用 MREMD 算法将历史运行参数的时序数据信号分解为若干个 IMF 分量和 1 个残余分量(R_n);其次,计算各个 IMF 分量与原信号的互相关系数和均方根误差,筛选出不符合阈值条件的分量进行 WTD,并将去噪后的 IMF 分量与原本符合筛选条件的分量以及残余分量重组新的 IMF 分量;再次,利用 SVD 将重组后的新 IMF 分量分解,重构成按照奇异值升序排列的奇异值分量,并采用基于优化参数的 PE 算法计算各个 IMF 分量的熵值,利用 K -means 聚类算法针对各 IMF 分量熵值进行分类;最后,选取熵值较低的一类 IMF 分量重构为当前运行参数的变化趋势项,并采用 ARIMA 模型对其进行预测.

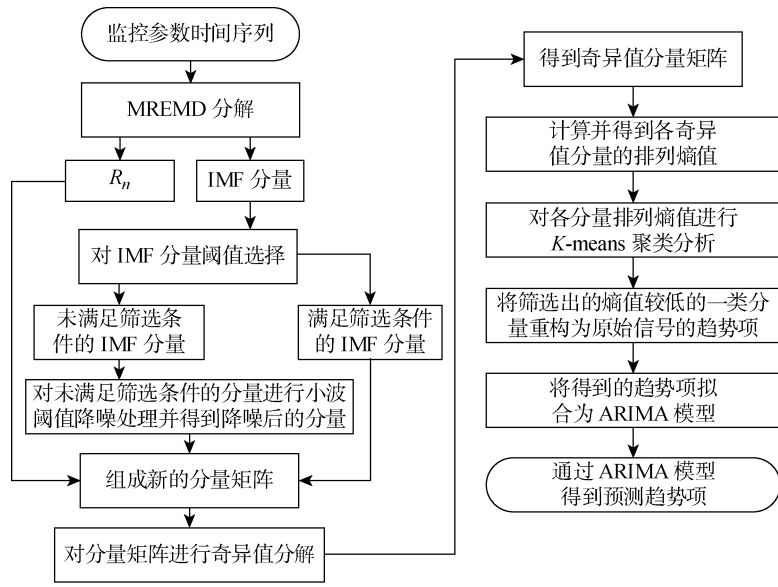


图 1 基于 MWSA 的单参数时序预测流程图

Fig. 1 Flow chart of single parameter sequential prediction method based on MWSA

1.1 MREMD 分解

MREMD 采用自回归 (AR) 模型对信号端点延拓,并用优化包络线拟合的方法改善了 EMD 算法中的“端点效应”问题,具体步骤和算法如下。

步骤一 对于长度为 N 的初始时间序列 $x_0(t)$,采用如下式所示的 p 阶 AR 模型对 $x(t)$ 的左右端点进行预测延拓,即 $x(t)$ 为延拓后的新序列,使左右端点处于延拓后时间序列的相邻两个极值点之间。

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \mu_t \quad (1)$$

式中: $x_t(t = p + 1, p + 2, \dots, N)$ 为均值点; $\phi_0, \phi_1, \dots, \phi_p$ 为 $p + 1$ 个实数; μ_t 为零均值的白噪声序列。

步骤二 计算延拓后新时序中相邻极值点间的均值,得到初始信号 $x_0(t)$ 的均值点序列 $\{x_{1,0}^m, x_{2,0}^m, \dots, x_{k,0}^m\}$,其中, k 为均值点个数,共 $k + 1$ 个;通过 3 次样条插值得到 $x_0(t)$ 的信号均值序列 $m_{1,0}(t)$,由 $x_0(t) - m_{1,0}(t)$ 得到 $x_0(t)$ 的 1 阶信号分量为

$$h_{1,0}(t) = x_0(t) - m_{1,0}(t) \quad (2)$$

步骤三 将计算得到的信号分量 $h_{1,0}(t)$ 作为原始信号,重复步骤一和步骤二进行迭代计算,直到经过 l 次迭代后的信号分量 $h_{1,l}(t)$ 满足如下式所示的终止条件。

$$\left. \begin{aligned} \sigma^* \left(= \frac{|\sigma_{l-1} - \sigma_l|}{X_\sigma^M} \right) &\leq 0.2 \\ P\{\vartheta_l \mid \vartheta_l \leq \vartheta_0\} &\geq 95\% \end{aligned} \right\} \quad (3)$$

式中: σ^* 为归一化标准差; σ_{l-1} 和 σ_l 分别为第 $l - 1$

次迭代后信号分量 $h_{1,l-1}(t)$ 和第 l 次迭代后信号分量 $h_{1,l}(t)$ 的均值点序列标准差; P 为条件概率;根据 Sigma 原则^[12] 定义 X_σ^M 为原始信号 $x_0(t)$ 中极值点绝对值的有义值; ϑ_0 和 ϑ_l 为初始信号和第 l 次迭代后的信号均值点与 X_σ^M 之比。具体有

$$P\{|x_z| \mid (|x_z| \leq X_\sigma^M), z = 1, 2, \dots, k + 1\} \geq 68.5\% \quad (4)$$

$$\left. \begin{aligned} \vartheta_0 &= \left\{ \frac{x_{1,0}^m}{X_\sigma^M}, \frac{x_{2,0}^m}{X_\sigma^M}, \dots, \frac{x_{k,0}^m}{X_\sigma^M} \right\} \\ \vartheta_l &= \left\{ \frac{x_{1,l}^m}{X_\sigma^M}, \frac{x_{2,l}^m}{X_\sigma^M}, \dots, \frac{x_{k,l}^m}{X_\sigma^M} \right\} \end{aligned} \right\} \quad (5)$$

式中: x_z 为延拓后新序列的第 z 个极值点; $x_{1,0}^m, x_{2,0}^m, \dots, x_{k,0}^m$ 为未进行迭代时初始信号 $x_0(t)$ 的均值点; $x_{1,l}^m, x_{2,l}^m, \dots, x_{k,l}^m$ 为第 $l - 1$ 次迭代后信号分量 $h_{1,l-1}(t)$ 的均值点。

步骤四 将 $h_{1,l}(t)$ 作为 1 阶本征模态函数 (即 I_1 分量) 输出, $x_0(t) - I_1$ 得到分量 R_1 ,并将 R_1 作为原始信号重复步骤一至步骤四,直到残余信号成为单调函数或分离不出新的 IMF 分量为止,残余信号提取过程表示为

$$\left. \begin{aligned} R_1 &= x_0(t) - I_1 \\ R_2 &= R_1 - I_2 \\ &\vdots \\ R_n &= R_{n-1} - I_n \end{aligned} \right\} \quad (6)$$

式中: 下标 n 为能够分解出的 IMF 分量个数。

经过上述分解后,原始信号 $x_0(t)$ 可表示为所有 IMF 分量和最终残余分量之和,表达式为

$$x_0(t) = \sum_{i=1}^n I_i + R_n \quad (7)$$

1.2 IMF 分量选取

对于 MREMD 分解后的 IMF 分量,噪声主要集中在高频分量中,在工程实际应用时一般通过舍弃高频分量的方法来达到降噪的目的,但这同样也会丢失部分的信号信息,使得分析结果产生误差。此外,虽然低频分量通常包含了较多的信息成分,但同样也存在一些噪声,仅靠舍弃高频分量的方法并不能够实现完全降噪的目的。为此,本文采用王彬蓉等^[13]提出的 IMF 分量筛选方法对所有分量进行筛选,在降噪的同时尽可能多地保留原始信号的信息,具体步骤和算法如下。

步骤一 计算各阶 IMF 分量与原始信号的相关系数 c_i 和均方根误差 RMSE_i ,表达式为

$$\left. \begin{aligned} c_i &= \frac{\sum_{j=1}^N (x_0(t) - \bar{x}_0(t))(I_i - \bar{I}_i)}{\sqrt{\sum_{j=1}^N (x_0(t) - \bar{x}_0(t))^2 \sum_{j=1}^N (I_i - \bar{I}_i)^2}} \\ \text{RMSE}_i &= \sqrt{\frac{1}{N} \sum_{j=1}^N (x_0(t) - I_i)^2} \end{aligned} \right\} \quad (8)$$

($i = 1, 2, \dots, n$)

式中: $\bar{x}_0(t)$ 和 \bar{I}_i 分别为原始信号 $x_0(t)$ 和各阶 IMF 分量的均值。其中,相关系数值越大,各分量与原始信号的相关度越高;均方根误差值越小,各分量与原始信号的密切程度越高。

步骤二 计算 IMF 分量相关系数筛选阈值和均方差筛选阈值,表达式为

$$\left. \begin{aligned} \lambda_{\text{IMF}} &= \frac{1}{n-1} \sum_{j=1}^{n-1} c_i \\ \delta_{\text{IMF}} &= \frac{1}{n-1} \sum_{j=1}^{n-1} \text{RMSE}_i \end{aligned} \right\} \quad (9)$$

步骤三 选取满足如下条件的 IMF 分量,进行小波阈值降噪为

$$c_i \leq \lambda_{\text{IMF}}, \quad \text{RMSE}_i \geq \delta_{\text{IMF}} \quad (10)$$

1.3 WTD

WTD 是利用小波变换将原始信号分解成多层近似系数和细节系数,由于经过小波变换后的噪声信息主要集中在绝对值较小的细节系数中,所以可通过将绝对值小于规定阈值的细节系数设置为 0,并将剩余小波系数(即分解得到的近似系数和保留的细节系数)通过小波逆变换重构回原始信号的方法,以达到去除噪声的目的,具体步骤和算法如下。

步骤一 阈值选取。传统 WTD 的阈值选取方法有固定阈值法、自适应阈值法、启发式阈值法和极大极小阈值法等^[14]。本文采用改进的固定阈值法,通过细节系数对每一层的噪声进行估计,同时利用

系数 $\ln(j+1)$ 逐层降低细节系数的阈值^[15],从而尽可能多地保留蕴含在高频分量中的真实信号,其阈值选取准则为

$$\lambda_j = \frac{2.1 \text{median}(|d_{(j)}|) \sqrt{\ln N}}{\ln(j+1)} \quad (11)$$

式中: λ_j 为第 j 层小波细节系数的噪声阈值; $d_{(j)}$ 为第 j 层小波细节系数; median 为中间值函数,即将每层系数按照降序排列后,取中间数的值(当系数个数为奇数时)或中间两个数的均值(当系数个数为偶数时)。

步骤二 阈值处理。常用阈值处理函数有硬阈值函数、软阈值函数和复合阈值函数^[16]等,依次表示为

$$\begin{aligned} \hat{d}_{(j)} &= \begin{cases} d_{(j)}, & |d_{(j)}| \geq \lambda_j \\ 0, & |d_{(j)}| < \lambda_j \end{cases} \\ \hat{d}_{(j)} &= \begin{cases} (|d_{(j)}| - \lambda_j) \text{sign } d_{(j)}, & |d_{(j)}| \geq \lambda_j \\ 0, & |d_{(j)}| < \lambda_j \end{cases} \\ \hat{d}_{(j)} &= \begin{cases} (|d_{(j)}| - a\lambda_j) \text{sign } d_{(j)}, & |d_{(j)}| \geq \lambda_j \\ 0, & |d_{(j)}| < \lambda_j \end{cases} \end{aligned}$$

式中: $\hat{d}_{(j)}$ 为降噪处理后的第 j 层小波细节系数; sign 为符号函数; $a \in [0, 1]$ 为调节系数,当 $a = 0$ 时 $\hat{d}_{(j)}$ 为硬阈值函数, $a = 1$ 时 $\hat{d}_{(j)}$ 为软阈值函数。

1.4 趋势项提取

采用 SVD、PE 和 K-means 聚类算法的组合以实现信号的进一步降噪和趋势项的提取,具体步骤和算法如下。

步骤一 SVD 分解。将经过小波阈值降噪处理后的 n 个 IMF 分量 $I_1 \sim I_n$ 和残余分量 R_n 去均值,然后以列向量的形式组成矩阵 $\mathbf{A}_{N \times (n+1)}$,并计算其协方差矩阵 $\mathbf{B}_{(n+1) \times (n+1)}$:

$$\mathbf{A}_{N \times (n+1)} = \begin{pmatrix} I_1 - \bar{I}_1 & I_2 - \bar{I}_2 & \cdots & I_n - \bar{I}_n & R_n - \bar{R}_n \end{pmatrix} \quad (12)$$

$$\mathbf{B}_{(n+1) \times (n+1)} = \mathbf{U}_{(n+1) \times (n+1)} \mathbf{A}_{(n+1) \times (n+1)} \mathbf{U}_{(n+1) \times (n+1)}^T \quad (13)$$

式中: $\mathbf{A} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_{n+1}})$ 为 \mathbf{B} 的奇异值组成的对角阵,其中, $\lambda_1, \lambda_2, \dots, \lambda_{n+1}$ 为按照降序排列的矩阵 \mathbf{B} 对应的特征值; \mathbf{U} 为 \mathbf{B} 的特征值对应的特征向量组成的矩阵。

将矩阵 \mathbf{U} 和 \mathbf{A} 中 0 特征值所对应的向量去掉,重构后的 K 个奇异值分量表示为

$$\mathbf{I}_{N \times K} = \mathbf{A}_{N \times K} \mathbf{U}_{K \times K} = (P_1 \ P_2 \ \cdots \ P_K) \quad (14)$$

式中: P_K 为第 K 个奇异值分量。

步骤二 PE 计算。利用延迟相空间重构法^[17]对 $\mathbf{I}_{N \times K}$ 中各个奇异值分量进行重构,如对分量 P_1 重构表示为

$$\mathbf{I}_{N \times K} = \begin{pmatrix} p_{1,1} & p_{1,1+\tau} & \cdots & p_{1,1+(m-1)\tau} \\ p_{1,2} & p_{1,2+\tau} & \cdots & p_{1,2+(m-1)\tau} \\ \cdots & \cdots & \ddots & \cdots \\ p_{1,l} & p_{1,l+\tau} & \cdots & p_{1,l+(m-1)\tau} \end{pmatrix} \quad (15)$$

式中: $l = N - (m-1)\tau$; $p_{1,i}$ ($i = 1, 2, \dots, N$) 为第 1 个奇异值分量的第 i 个元素; m 为嵌入维数; τ 为延迟时间。

将重构矩阵的每行元素按升序排列, 得到每行元素的索引值, 然后将每行元素的索引值组成索引向量并构成索引矩阵。索引矩阵中每行元素的排列方式共有 $m!$ 种可能, 假设第 b 种排列方式在索引矩阵中的出现次数为 u_b , 统计其出现的频率为

$$f_b = u_b / m!, \quad b \in (1, K) \quad (16)$$

由式(16)可计算得到各奇异值分量的排列熵为

$$\text{PE}_i = - \sum_{b=1}^{m!} f_b \lg f_b \quad (17)$$

步骤三 K -means 聚类。根据计算得到的排列熵值对 IMF 分量进行聚类, 选取熵值较低的分量重构原始信号的趋势项。首先, 任意选取其中 1 个点 ρ_1 作为 1 个类别中心, 并选择与 ρ_1 距离最远的 ρ_2 点为另 1 个类别中心, 计算熵值序列中其余对象点 x 与 ρ_1 和 ρ_2 的距离为

$$d(\rho, x) = |x - \rho| \quad (18)$$

然后, 把其余各点划分到与两个类别中心距离近的一类, 将两类点的均值作为新的类别中心进行分类, 重复上述步骤, 直至两个类别中心不再发生变化为止。最后, 将两类熵值点中平均熵值较低一类所对应的奇异值分量筛选出来^[18], 重构为原始信号的趋势项为

$$\mathbf{T}_{N \times 1} = \sum (\mathbf{E}_{N \times K} + \mathbf{S}_{N \times K} \mathbf{Q}_{K \times K}^T) \quad (19)$$

式中: $\mathbf{E}_{N \times K}$ 为通过聚类分析选取的熵值点对应的 IMF 分量取均值扩展后的矩阵, 未选取分量对应的均值用 $N \times 1$ 替换; $\mathbf{S}_{N \times K}$ 为筛选出的奇异值分量矩阵, 未选取的分量用 $N \times 1$ 替换; $\mathbf{Q}_{K \times K}^T$ 为将 $\mathbf{U}_{K \times K}^T$ 中未选取分量所对应的特征向量用 $K \times 1$ 替换后的矩阵。

1.5 ARIMA 预测模型

ARIMA 常用于差分平稳时间序列的拟合和预测, 具体步骤和算法如下。

步骤一 对式(19)计算得到的趋势项 $\mathbf{T}_{N \times 1}$ 进行平稳性 (Phillips-Perron) 检验^[19], 然后对检验结果为非平稳的时间序列作差分平稳化处理, 表示为

$$\hat{\mathbf{T}}_{N \times 1} = L^D(\text{diff}^f(\mathbf{T}_{N \times 1})) \quad (20)$$

式中: $\hat{\mathbf{T}}_{N \times 1}$ 为差分平稳化处理后的趋势项; L 为延迟

算子; D 为延迟阶数; diff^f 为 f 次差分。

步骤二 对经过平稳化处理后的趋势项的 $\hat{\mathbf{T}}_{N \times 1}$ 进行赤池信息准则 (AIC) 定阶^[20], 其定阶准则表示为

$$\text{AIC}(s, q) = \ln \hat{\sigma}_a^2 + \frac{2r}{N} \quad (21)$$

式中: s 和 q 分别为模型的自回归系数多项式阶数和移动平均系数多项式阶数; $\hat{\sigma}_a^2$ 为模型残差标准差的极大似然估计; r 为模型的独立参数个数。

步骤三 经过 AIC 检验得到模型的自回归系数多项式阶数 s 和移动平均系数多项式阶数 q , 并拟合 ARIMA(s, d, q) 模型表达为

$$y_t = c + a_1 y_{t-1} + \cdots + a_s y_{t-s} + \epsilon_t - b_1 \epsilon_{t-1} - \cdots - b_q \epsilon_{t-q} \quad (22)$$

式中: d 为非平稳序列经差分变为平稳序列的差分次数; $a_1 \sim a_s$ 为 AR 模型系数; $b_1 \sim b_q$ 为 MA 模型系数; $\epsilon_{t-q} \sim \epsilon_t$ 为白噪声序列; c 为常数; $y_{t-s} \sim y_t$ 为 $\hat{\mathbf{T}}_{N \times 1}$ 中的元素。

步骤四 对 ARIMA 模型的残差序列进行 LB (Ljung-Box) 统计量检验^[21], 表示为

$$R_{\text{LB}} = N(N+2) \sum_{i=1}^{k'} \frac{\hat{\epsilon}_i}{N-i} \sim \chi^2(k') \quad (23)$$

$$\forall k' > 0$$

式中: $N \in [1, k']$ 且为整数; R_{LB} 为残差序列的 Ljung-Box 统计量值; $\hat{\epsilon}_i$ 为模型残差的估计值; k' 为残差个数; χ^2 为卡方分布。

2 预测案例

以某型船舶蒸汽动力系统的除氧器水位为案例, 对上述基于 MWSA 的热力系统单参数时序预测方法进行验证。按 1.1 节的步骤和算法, 对一段时间内监控系统采集的实际水位数据进行 MREMD 分解, 得到 7 个 IMF 分量和 1 个残余分量, 如图 2 所示。

对于分解得到的分量, 根据式(8)计算其与原始信号的相关系数和均方根误差, 如表 1 所示。

根据式(9)计算得到 IMF 分量相关系数的筛选阈值 $\lambda_{\text{IMF}} = 0.3017$, 均方差的筛选阈值 $\delta_{\text{IMF}} = 7.3141$, 符合条件的分量有 $I_4 \sim I_7$ 。按 1.3 节的具体步骤和算法对 $I_1 \sim I_3$ 进行小波阈值降噪, 经反复测试, 选取 Sym4 小波基, 分解层数为 3 层。采用固定阈值选取方法, 根据式(11)计算各层的噪声阈值。采用不同阈值函数处理后的各 IMF 分量的信噪比及均方根误差, 如表 2 所示。

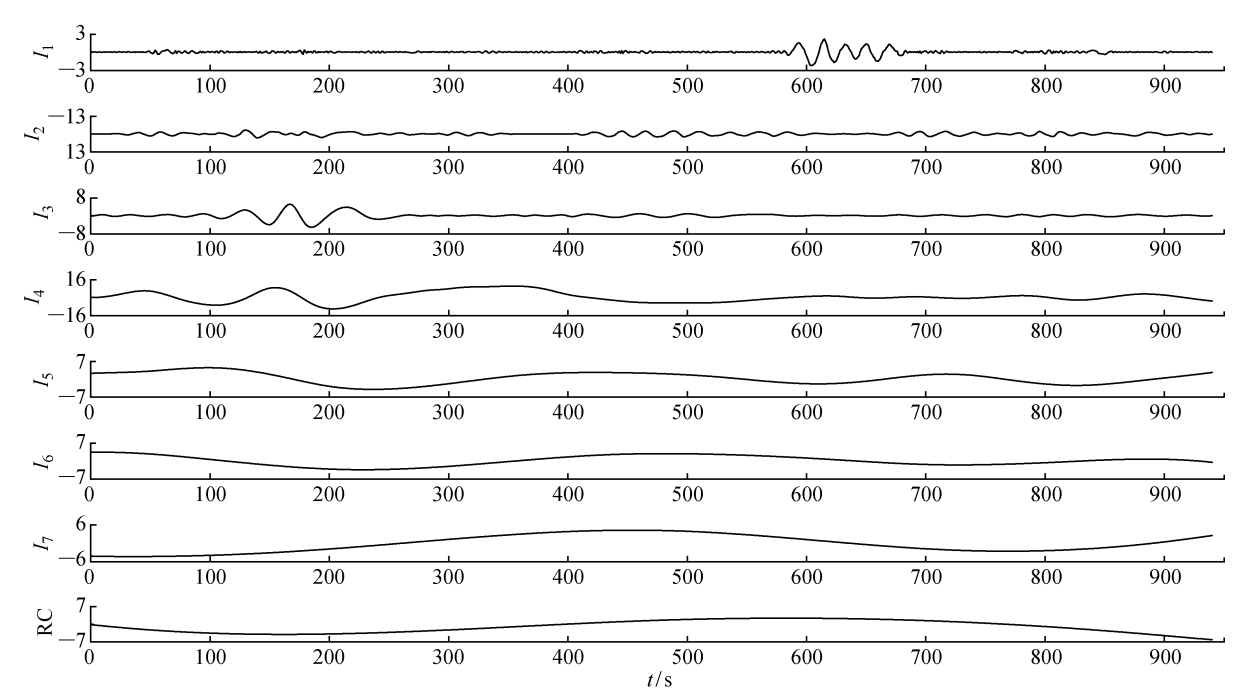


图 2 分解得到的 IMF 分量和 RC
Fig. 2 Decomposed IMF component and RC

表 1 各个 IMF 分量与原信号的相关系数和均方根误差
Tab. 1 Correlation coefficient and root mean square error of each IMF component relative to the original signal

IMF 分量	相关系数	均方根误差
I_1	0. 046 9	7. 772 7
I_2	0. 105 5	7. 739 0
I_3	0. 106 1	7. 742 4
I_4	0. 526 2	6. 679 1
I_5	0. 426 7	7. 153 2
I_6	0. 599 1	6. 797 9
I_7	0. 593 4	6. 494 4

表 2 去噪后各 IMF 分量的信噪比、均方根误差
Tab. 2 Signal-to-noise ratio and root mean square error of each IMF component after denoising

方法	IMF 分量	信噪比	均方根误差
复合阈值函数	I_1	39. 185 4	0. 004 1
	I_2	36. 503 3	0. 013 8
	I_3	56. 600 6	0. 001 6
硬阈值函数	I_1	14. 015 4	0. 073 5
	I_2	38. 981 8	0. 010 3
	I_3	83. 197 9	$7. 604 1 \times 10^{-5}$
软阈值函数	I_1	12. 118 4	0. 090 7
	I_2	32. 954 7	0. 020 7
	I_3	74. 748 9	0. 000 2

由表 2 可知,对于高频的 I_1 分量,采用复合阈值降噪的信噪比更高,均方差更小. 但对于较低频的 I_2 和 I_3 分量,采用硬阈值降噪的效果更佳. 按 1. 4 节的步骤和算法,将降噪后的 $I_1 \sim I_3$ 与 $I_4 \sim I_7$ 及残余分量 R_n 一起进行 SVD 分解,得到 8 个奇异值分量即 P1~P8,如图 3 所示.

根据式(15)~(17)计算各奇异值分量的排列熵值. 在相空间重构时,运用互信息法^[22]选取最佳延迟时间,在 1~100 s 的延迟时间范围内,各奇异值分量的互信息值 Mi1~Mi8 变化如图 4 所示. 采用

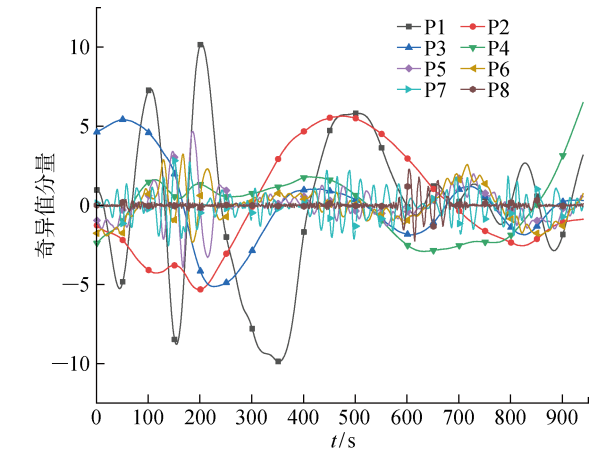


图 3 SVD 分解后的奇异值分量
Fig. 3 Singular value components after SVD decomposition

伪近邻法^[23]计算最佳嵌入维数,在 1~10 的嵌入维数变化范围内,各奇异值分量的伪近邻率 FNNP1~FNNP8 随嵌入维数 m 的变化如图 5 所示.

取图 4 中互信息值随时间变化曲线的第一个极值点为各奇异值分量的最佳延迟时间 τ_{opt} ;取图 5 中伪近邻率小于 5% 时的嵌入维数为各奇异值分量的最佳嵌入维数 m_{opt} ;然后将其代入式(15),并通过式(16)和式(17)计算得到各分量的排列熵值,如表 3 所示.

对各分量的排列熵值进行 K -means 聚类,选取 P2、P3 和 P4 共 3 个分量进行趋势项重构,结果如图 6 所示.

根据 1.5 节的步骤和算法,对除氧器水位进行拟合和预测.首先,对如图 6 所示的趋势项进行 Phillips-Perron 检验,结果为非平稳序列,再经过 4

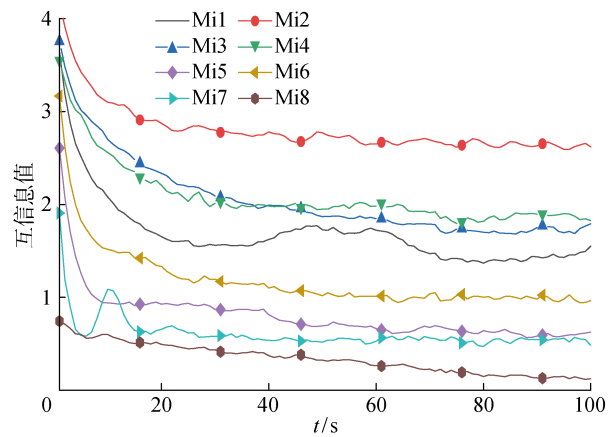


图 4 各分量互信息值随延迟时间的变化曲线
Fig. 4 Variation curve of mutual information value with delay time

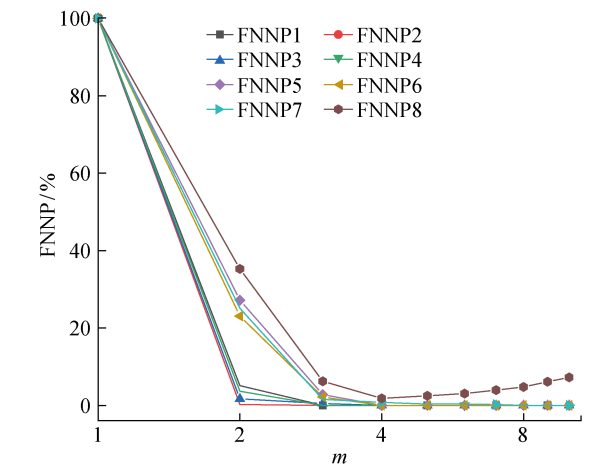


图 5 各分量伪近邻率随嵌入维数的变化曲线
Fig. 5 Variation of pseudo nearest neighbor rate with embedding dimension

次差分计算后转换为平稳序列;然后,对差分平稳化后的序列进行 AIC,得到自回归系数多项式阶数 s 以及移动平均系数多项式阶数 q 分别为 1 和 3;在经过 4 次反差分计算和 LB 统计量检验后,拟合得到如式(22)所示的 ARIMA(1, 4, 3)模型.

取实际运行数据的前 90% 为训练数据,取后 10% 为检验数据,预测除氧器水位在未来 94 s 的变化趋势,结果如图 7 所示.而在相同数据输入下,采用文献[23]中提出的无降噪处理的算法(MSOP)得到的预测结果如图 8 所示.

由图 7 和图 8 可知,随着预测时长的增加,两种方法的预测误差都会增大.经计算,在采用 MWSA 方法时,预测结果与实际数据(检验数据)的均方差为 1.388 7;而文献[23]中提出的方法其预测结果与实际数据(检验数据)的均方差为 2.133 4.可知,本文预测方法的准确度更高.

表 3 各奇异值分量的最佳延迟时间、最佳嵌入维数和排列熵值
Tab. 3 Optimal delay time, embedding dimension, and permutation entropy of each singular value component

奇异值分量	τ_{opt}	m_{opt}	PE
P1	23	3	2.467 2
P2	12	2	0.975 2
P3	25	2	1.263 9
P4	18	2	1.212 7
P5	12	3	3.123 8
P6	9	3	2.972 9
P7	6	3	2.955 8
P8	6	4	4.709 8

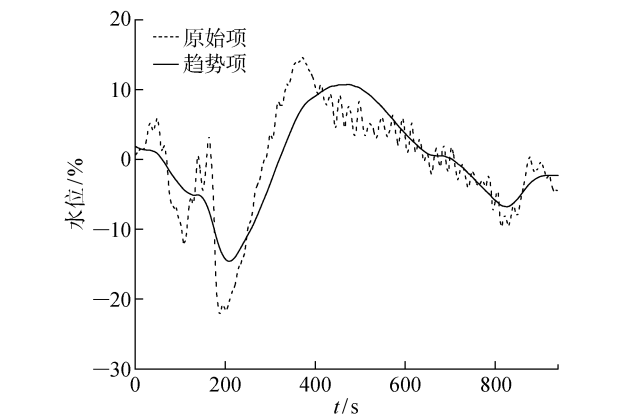


图 6 除氧器水位的原始项与趋势项
Fig. 6 Original item and trend item of deaerator water level

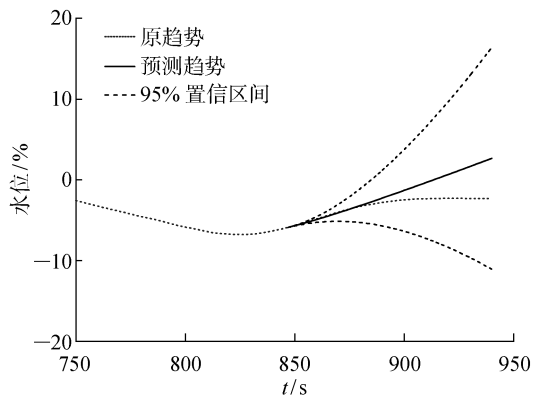


图 7 基于 MSOP 的除氧器水位趋势项与预测项

Fig. 7 Trend and prediction of deaerator water level based on MSOP

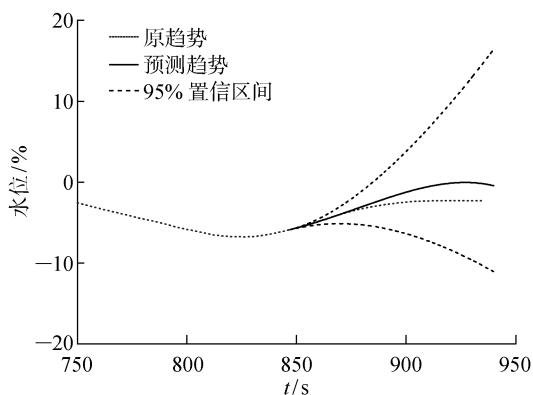


图 8 基于 MWSA 的除氧器水位的趋势项与预测项

Fig. 8 Trend and prediction of deaerator water level based on MWSA

3 结语

提出一种基于 MWSA 的热力系统单参数时序预测方法,并以某型号船舶蒸汽动力系统的除氧器水位为例,根据实际运行数据预测其在未来一段时间内的变化趋势。通过与基于 MSOP 的时序预测方法对比,证明所提方法的预测误差较小,预测结果较为准确。该算法完全基于基本物理规律,利用纯数学推导得出,在推导过程中未加入任何约束条件或有关具体设备的辅助方程,在数学上具有本质的普适性。利用滑动数据采集窗口可以对其他时段的数据进行分析验证,亦可通过添加其他设备的数据接口,对其他热力参数进行分析和预测。研究成果可用于热力系统的运行管理、控制优化和维修保障研究,达到优化运行管理和控制策略,合理安排维修保障资源、减少设备故障和延长系统使用寿命的目的。

参考文献:

[1] 倪何,覃海波,郑奕杨. 考虑给水泄漏的锅炉升负荷

仿真及其可靠性[J]. 上海交通大学学报, 2021, 55(4): 444-454.

NI He, QIN Haibo, ZHENG Yiyang. Simulation and performance reliability of boiler load raising process considering leakage of feed water[J]. **Journal of Shanghai Jiao Tong University**, 2021, 55(4): 444-454.

[2] 秦文学,王嘉兴,王继强,等. 基于 ARMA 模型的燃煤机组主蒸汽压力控制策略[J]. 热力发电, 2020, 49(9): 127-132.

QIN Wenxue, WANG Jiaying, WANG Jiqiang, *et al.* Control strategy of main steam pressure of coal-fired unit based on ARMA model[J]. **Thermal Power Generation**, 2020, 49(9): 127-132.

[3] 朱少民,夏虹,吕新知,等. 基于 ARIMA 和 LSTM 组合模型的核电厂主泵状态预测[J]. 核动力工程, 2022, 43(2): 246-253.

ZHU Shaomin, XIA Hong, LYU Xinzh, *et al.* Condition prediction of reactor coolant pump in nuclear power plants based on the combination of ARIMA and LSTM[J]. **Nuclear Power Engineering**, 2022, 43(2): 246-253.

[4] RUIZ-AGUILAR J J, TURIAS I, GONZÁLEZ-ENRIQUE J, *et al.* A permutation entropy-based EMD-ANN forecasting ensemble approach for wind speed prediction[J]. **Neural Computing and Applications**, 2021, 33(7): 2369-2391.

[5] 陈亮,刘宏立,郑倩,等. 基于 EEMD-SVD-PE 的轨道波磨趋势项提取[J]. 哈尔滨工业大学学报, 2019, 51(5): 171-177.

CHEN Liang, LIU Hongli, ZHENG Qian, *et al.* An EEMD-SVD-PE approach to extract the trend of track irregularity[J]. **Journal of Harbin Institute of Technology**, 2019, 51(5): 171-177.

[6] YANG Z J, LING B W K, BINGHAM C. Joint empirical mode decomposition and sparse binary programming for underlying trend extraction[J]. **IEEE Transactions on Instrumentation and Measurement**, 2013, 62(10): 2673-2682.

[7] 梁兵,汪同庆. 基于 HHT 的振动信号趋势项提取方法[J]. 电子测量技术, 2013, 36(2): 119-122.

LIANG Bing, WANG Tongqing. Method of vibration signal trend extraction based on HHT[J]. **Electronic Measurement Technology**, 2013, 36(2): 119-122.

[8] 刘海江,刘世高,李敏. 换挡加速度信号的 EMD 和小波阈值降噪方法[J]. 噪声与振动控制, 2018, 38(2): 198-203.

LIU Haijiang, LIU Shigao, LI Min. EMD and wavelet threshold denoising method of gear-shift acceleration signals[J]. **Noise and Vibration Control**, 2018,

- 38(2): 198-203.
- [9] 黄礼敏. 海浪中非平稳非线性舰船运动在线预报研究[D]. 哈尔滨: 哈尔滨工程大学, 2016.
HUANG Limin. On-line prediction of non-stationary and nonlinear ship motions at sea[D]. Harbin: Harbin Engineering University, 2016.
- [10] BEHERA A P, GAURISARIA M K, RAUTARAY S S, *et al.* Predicting future call volume using ARIMA models[C]// **2021 5th International Conference on Intelligent Computing and Control Systems**. Madurai, India: IEEE, 2021: 1351-1354.
- [11] 仇琦, 杨兰, 丁旭, 等. 基于改进 EMD-ARIMA 的光伏发电系统短期功率预测[J]. **电力科学与工程**, 2020, 36(8): 42-50.
QIU Qi, YANG Lan, DING Xu, *et al.* An improved short-term power prediction method of PV power generation system based on EMD-ARIMA model[J]. **Electric Power Science and Engineering**, 2020, 36(8): 42-50.
- [12] 尤保健. 基于六西格玛理论的轴流泵叶轮水力效率影响因子分析[J]. **水电能源科学**, 2020, 38(2): 168-171.
YOU Baojian. Influencing factors analysis of hydraulic efficiency of axial flow pump impeller based on six sigma theory[J]. **Water Resources and Power**, 2020, 38(2): 168-171.
- [13] 王彬蓉, 王维博, 周超, 等. 基于 EMD 自适应重构的心音信号特征筛选及分类[J]. **航天医学与医学工程**, 2020, 33(6): 533-541.
WANG Binrong, WANG Weibo, ZHOU Chao, *et al.* Feature selection and classification of heart sound based on EMD adaptive reconstruction[J]. **Space Medicine & Medical Engineering**, 2020, 33(6): 533-541.
- [14] 崔公哲, 张朝霞, 杨玲珍, 等. 一种改进的小波阈值去噪算法[J]. **现代电子技术**, 2019, 42(19): 50-53.
CUI Gongzhe, ZHANG Zhaoxia, YANG Lingzhen, *et al.* An improved wavelet threshold denoising algorithm[J]. **Modern Electronics Technique**, 2019, 42(19): 50-53.
- [15] CHANG F X, HONG W X, ZHANG T, *et al.* Research on wavelet denoising for pulse signal based on improved wavelet thresholding[C]// **2010 First International Conference on Pervasive Computing, Signal Processing and Applications**. New York, USA: IEEE, 2010: 564-567.
- [16] 徐晨, 赵瑞珍, 甘小冰. 小波分析应用算法[M]. 北京: 科学出版社, 2016.
XU Chen, ZHAO Ruizhen, GAN Xiaobing. Application algorithm of wavelet analysis[M]. Beijing: Science Press, 2016.
- [17] 谢忠玉, 张立. 相空间重构参数选择方法的研究[J]. **中国科技信息**, 2009(16): 42-43.
XIE Zhongyu, ZHANG Li. Selection of embedding parameters in phase space reconstruction[J]. **China Science and Technology Information**, 2009(16): 42-43.
- [18] 杨恒岳, 刘青荣, 阮应君. 基于 k-means 聚类算法的分布式能源系统典型日冷热负荷选取[J]. **热力发电**, 2021, 50(3): 84-90.
YANG Hengyue, LIU Qingrong, RUAN Yingjun. Selection of typical daily cooling and heating load of CCHP system based on k-means clustering algorithm[J]. **Thermal Power Generation**, 2021, 50(3): 84-90.
- [19] 田行宇, 李传金. PP 检验对异方差时间序列的伪检验[J]. **统计与决策**, 2018, 34(17): 74-76.
TIAN Xingyu, LI Chuanjin. Spurious tests to heteroscedastic time series with PP test[J]. **Statistics & Decision**, 2018, 34(17): 74-76.
- [20] 李国春, 王恩龙, 王丽梅, 等. 基于 AIC 准则判断锂电池最优模型[J]. **汽车工程学报**, 2019, 9(5): 352-358.
LI Guochun, WANG Enlong, WANG Limei, *et al.* Evaluating optimal models of lithium battery based on AIC[J]. **Chinese Journal of Automotive Engineering**, 2019, 9(5): 352-358.
- [21] 李颖若, 韩婷婷, 汪君霞, 等. ARIMA 时间序列分析模型在臭氧浓度中长期预报中的应用[J]. **环境科学**, 2021, 42(7): 3118-3126.
LI Yingruo, HAN Tingting, WANG Junxia, *et al.* Application of ARIMA model for mid-and long-term forecasting of ozone concentration[J]. **Environmental Science**, 2021, 42(7): 3118-3126.
- [22] 饶国强, 冯辅周, 司爱威, 等. 排列熵算法参数的优化确定方法研究[J]. **振动与冲击**, 2014, 33(1): 188-193.
RAO Guoqiang, FENG Fuzhou, SI Aiwei, *et al.* Method for optimal determination of parameters in permutation entropy algorithm[J]. **Journal of Vibration and Shock**, 2014, 33(1): 188-193.
- [23] 郑奕扬, 倪何, 金家善. 基于 MSOP 的蒸汽动力系统单参数运行稳定性评估方法[J]. **上海交通大学学报**, 2021, 55(11): 1438-1444.
ZHENG Yiyang, NI He, JIN Jiashan. An operation stability assessment method of a single-parameter in steam power system based on MSOP[J]. **Journal of Shanghai Jiao Tong University**, 2021, 55(11): 1438-1444.

(本文编辑: 孙伟)