

DOI: 10. 20040/j. cnki. 1000-7709. 2023. 20230827

基于移动最小二乘法的水轮机运行特性曲面构建

李 鹏¹, 刘艳娜^{2,3}

(1. 甘肃省水利水电勘测设计研究院有限责任公司, 甘肃 兰州 730000; 2. 国电南京自动化股份有限公司, 江苏 南京 210032; 3. 南京河海南自水电自动化有限公司, 江苏 南京 210032)

摘要: 针对水轮机实测运行数据分布非结构性、疏密不均等特点,采用移动最小二乘法拟合机组实测运行参数获得水轮机工作区空间特性曲面,可体现数据曲面局部特性、避免数据分布密度差异导致的失真。通过拉依达准则进行初步的数据滤波,对离散样本点进行不同范围的局部拟合自适应确定支撑半径,结合马氏距离除噪法建立稳健移动最小二乘曲面重构模型。实例计算结果表明,与最小二乘法相比,移动最小二乘法对水轮机运行特性数据进行了较高精度的有效拟合和局部数据补充,可真实反映水轮机运行特性,为修正水轮机出厂运行特性曲线提供参考。

关键词: 移动最小二乘法; 水轮机运行数据; 曲面拟合; 支撑半径; 特性曲面

中图分类号: TK733; [TV734.1] **文献标志码:** A **文章编号:** 1000-7709(2023)11-0183-04

1 引言

水轮机特性是研究水泵水轮机水力设计、机组启动、工况转换、过渡过程等关键技术的基础。当前水电站运行控制与经济调度过程多采用水轮机生产厂家提供的模型曲线。然而因比尺效应和机组安装、运行条件的影响,水轮机出厂特性数据与其真实运行特性之间存在差别^[1],导致水电机组运行控制策略在工程应用时失效。因此研究水轮机在电厂中的实测运行特性,既可反映水轮机真实运行状况,又可有效利用监控系统数据、提高相关研究的可靠性和真实性,为修正水轮机出厂运行特性数据提供依据。目前对水轮机运行特性的研究多基于模型特性数据通过数学几何^[2-4]、原理方程^[5-7]或智能算法^[8]分析各物理量的变化规律。这些方法以规律性较好的水轮机生产厂家提供的模型特性曲线试验数据为例,较少涉及水电站运行中的实测机组参数。水轮机实际运行数据高度集中于高效率运行区,在一般处理过程中极易倾向表达高效率区特性、忽略低效率区特性,降低实测数据使用效果。为此,本文利用移动最小二乘法(MLS)^[9]结合支撑半径寻优和马氏距离(MD)^[10]滤波法构建水电机组实测数据运行特性

空间曲面,得到精度较高的水轮机工作区运行特性数据,并比较了水轮机实测特性数据和出厂特性数据的区别,以期水轮机全工况真实运行特性的构建提供依据。

2 移动最小二乘的曲面重构模型

2.1 移动最小二乘的曲面重构原理

2.1.1 拟合函数的建立

移动最小二乘法为已知节点 $\mathbf{X} = [x_1, x_2, \dots, x_n]^T$ ($x_i \in \mathbf{R}^D$, n 为节点数目, D 为节点空间维数), 对应的节点值 $\mathbf{Y} = [y_1, y_2, \dots, y_n]^T$, 求解拟合函数 $f(x)$ 使:

$$\sum_{i=1}^n \omega(x - x_i) [f(x_i) - y_i]^2 \rightarrow \min \quad (1)$$

式中, $\omega(\cdot)$ 为 x_i 的权函数; $x - x_i$ 为节点间欧氏距离。

在支撑域内随机离散节点上, 拟合函数 $f_{MLS}(x)$ 为:

$$f_{MLS}(x) = \sum_{i=1}^m \alpha_i(x) p_i(x) = p^T(x) \alpha(x) \quad (2)$$

其中 $p(x) = [p_1(x), p_2(x), \dots, p_m(x)]^T$

$$\alpha(x) = [\alpha_1(x), \alpha_2(x), \dots, \alpha_m(x)]^T$$

式中, $p(x)$ 为基函数, $p_1(x) \in \Pi_r^D, \Pi_r^D$ 为 D 维

收稿日期: 2023-04-22, 修回日期: 2023-06-09

作者简介: 李鹏(1980-), 男, 高级工程师, 研究方向为水利水电工程电气、信息化设计, E-mail: 13919463527@139.com

通讯作者: 刘艳娜(1984-), 女, 高级工程师, 研究方向为水电站安全控制与运行, E-mail: yannaliu@foxmail.com

空间中最高次幂为 r 的多项式空间; m 为基函数的项数, $m = (r + D)! / (r! D!)$, 二维常用基函数中, 当 $m = 3$ 时线性基为 $p(x) = [1, x_1, x_2]^T$, 当 $m = 6$ 时二次基为 $p(x) = [1, x_1, x_2, x_1^2, x_1 x_2, x_2^2]^T$; $\alpha(x)$ 为拟合系数。

解析后的拟合系数 $\alpha(x)$ 为:

$$\alpha(x) = \mathbf{A}^{-1}(x)\mathbf{B}(x)\mathbf{Y} \quad (3)$$

拟合函数 $f(x)$ 为:

$$f(x) = p^T(x)\mathbf{A}^{-1}(x)\mathbf{B}(x)\mathbf{Y} = \Phi(x)\mathbf{Y} \quad (4)$$

其中 $\Phi(x) = p^T(x)\mathbf{A}^{-1}(x)\mathbf{B}(x)$ (5)

$$A(x) = w(x - x_i)p(x_i)p^T(x_i)$$

$$B(x) = [w(x - x_1)p(x_1), w(x - x_2)p(x_2), \dots, w(x - x_n)p(x_n)]$$

式中, $\Phi(x)$ 为形函数。

为确保 \mathbf{A}^{-1} 存在, 要求矩阵 \mathbf{A} 非奇异, 则支撑域内至少有 m 个采样点。

2.1.2 权函数

$w(x - x_i)$ 为权函数, 其值非负并随 $\|x - x_i\|_2$ 的增加单调递减, 具有紧支性, 即在支撑域内不等于零, 支撑域外等于零。本文以高斯型权函数 $w(r)$ 为例进行研究:

$$w(r) = \begin{cases} \frac{\exp(-r^2\beta^2) - \exp(-\beta^2)}{1 - \exp(-\beta^2)} & r \leq 1 \\ 0 & r > 1 \end{cases} \quad (6)$$

其中 $r = |x - x_i| / d_{mi}$

式中, r 为相对距离; d_{mi} 为支撑半径; β 为形状参数。

权函数的主要参数为支撑域 s_{\max} 和形状参数 β 。在均匀节点下, $s_{\max} = s_{\text{scale}} C_i$, s_{scale} 为大于 1 的乘子, C_i 为节点的间距, 应保证每个待拟合点的支撑域内包含的节点数不少于 m ; 形状参数 β 反映权函数的形态, 增加 β 会导致权函数 $w(x)$ 衰减更快, 支撑域的实际半径更小, 增强拟合点附近节点对拟合结果的影响并减小远距离节点的作用。

2.1.3 确定支撑半径

因经济运行和避开振动区的运行要求, 水电机组真实运行工况点在高效率运行区与低效率区分布密度差别显著。当采用经验公式时, 会因支撑域过小而使局部过拟合导致严重偏离整体趋势。因此, 本文采用以下方法避免局部过拟合: ①根据样本节点 \mathbf{X} 、 \mathbf{Y} 分布范围划分网格, 确定待拟合节点在 xy 平面上的坐标。②根据网格的排列依次采用 MLS 计算各拟合点坐标。③对每个网格点, 采用 MLS 计算出当前网格对应的数据后, 再次使用 MLS 计算 k 倍支撑半径内网格点对应的数据, 两次计算结果的误差应不大于某一阈值 e 。若差别过大, 则 $s_{\text{scale}} + 1$, 重新确定支撑半径重

复这一步骤, 直到满足要求。

本文 $k \geq 5$, 阈值 $e \leq 5\%$ 。试验表明该方法可有效避免局部过拟合问题。

2.2 水轮机运行特性数据去噪

水电站实测水头、流量、有功功率等运行参数分别来自水位计、流量计、电能量系统等, 测量误差包括系统误差、随机误差、粗大误差及空载、开机、停机过程等非正常运行工况。构建水电机组运行特性曲面过程中, 首先采用拉依达准则去掉粗大误差和特殊工况数据, 再根据拟合数据与采样数据之间的马氏距离 (M_D) 设置适当的阈值进一步滤除噪点。

2.2.1 拉依达准则滤波

根据拉依达准则, 数值分布在 $(\eta - 3\sigma, \eta + 3\sigma)$ 中的概率为 0.997 4, η 为各维数据的均值, σ 为标准差。在三维空间曲面中, 对每个维度的采样数据依据拉依达准则保留 $\eta \pm 3\sigma$ 范围内的数据、剔除粗大误差和非正常运行数据。

对于三维曲面数据集 $X = [x, y, z]$, 内含节点 n 个, 各维数据 x, y, z 的均值 μ 分别为:

$$\begin{cases} \mu_x = \frac{1}{n} \sum_{i=1}^n x_i \\ \mu_y = \frac{1}{n} \sum_{i=1}^n y_i \\ \mu_z = \frac{1}{n} \sum_{i=1}^n z_i \end{cases} \quad (7)$$

各维数据 x, y, z 的标准差 σ 分别为:

$$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)^2} \quad (8)$$

$$\sigma_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \mu_y)^2} \quad (9)$$

$$\sigma_z = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_i - \mu_z)^2} \quad (10)$$

式中, i 为数据集 X 内数据点编号; x_i, y_i, z_i 为三维数据集 X 内数据点对应维度坐标值; μ_x, μ_y, μ_z 分别为 x, y, z 对应维度数据均值; $\sigma_x, \sigma_y, \sigma_z$ 分别为 x, y, z 对应维度数据的标准差。

2.2.2 基于马氏距离的滤波

剔除较大噪声后, 在运用移动最小二乘拟合曲面过程中, 基于马氏距离对数据进一步滤波。马氏距离为数据的协方差距离。与欧氏距离不同, 它考虑各种特性之间的联系并有效度量两个未知样本集的相似度, 且马氏距离为尺度无关的, 即独立于测量尺度, 可排除变量之间的相关性的干扰, 广泛用于分类和聚类分析中。

水轮机运行特性曲面拟合时以每组坐标 $(x,$

y, z) 中 x, y 作为自变量、 z 作为因变量计算求解。设 n 个采样节点的误差数据集为 U :

$$U = [X, Y, Z^e] \quad (11)$$

其中 $Z^e = Z' - Z$

式中, Z', Z 分别为因变量拟合值、实测值。

各误差数据点到均值向量 μ 的马氏距离 (M_D) 为:

$$M_D(U, \mu) = \sqrt{(U - \mu)^T \Sigma^{-1} (U - \mu)} \quad (12)$$

式中, Σ 为协方差矩阵。

滤波算法为衡量拟合节点与采样数据之间的马氏距离, 保留 M_D 小于阈值 H_D 的数据作为有效测量数据来构建水轮机运行特性曲面。当测量数据的随机误差导致数据点整体偏离支撑域内平均特性时, 判定该点为误差点并进行剔除。

2.3 水轮机运行特性曲面重构

根据移动最小二乘法的曲面拟合原理, 构建算法模型流程见图 1。其具体步骤如下。

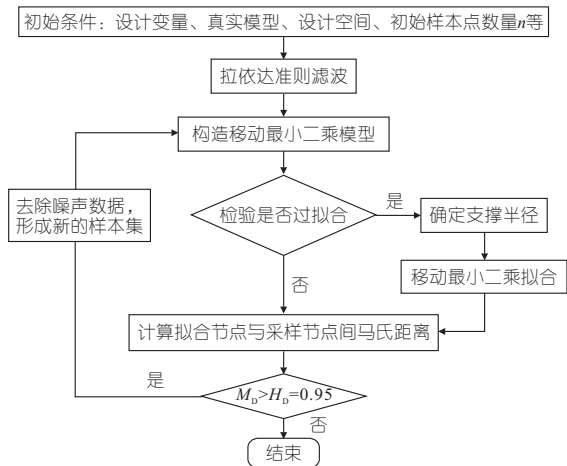


图 1 移动最小二乘拟合模型

Fig. 1 Fitting model of MLS

步骤 1 根据测量数据设置初始条件。设计变量、真实模型、设计空间、初始样本点数量 n 等。

步骤 2 根据拉依达准则剔除采样点中的离群孤立数据。

步骤 3 确定支撑半径。根据经验公式确定待拟合节点支持域半径初始值, 若 A 为奇异矩阵, 则支撑域乘子 $s_{scale} + 1$, 直到 A 非奇异, 计算

该节点拟合值。

步骤 4 对每个待拟合节点完成步骤 3 的计算, 得到拟合数据。

步骤 5 根据拟合结果, 采用样条插值法得到初始样本点自变量对应的因变量, 计算插值结果 Z' 与实测值 Z 的之间的马氏距离 M_D 。

步骤 6 设定滤波阈值 H_D , 剔除 $M_D \geq H_D$ 的样本点。

步骤 7 重复步骤 3~6, 直到所有样本点与拟合点间的 M_D 均小于 H_D 。

3 实例分析

3.1 实例计算

某不完全多年调节水电站引水系统采用两机一洞布置型式, 装机 4×127.6 MW。采集 2021 年 7 月 1 日~9 月 30 日的上游水位、尾水位、发电流量和发电出力, 采集频率 1 个数据/时, 共计 2 202 个数据, 其中共水力单元机组同时发电的工况有 1 472 个数据, 见图 2(a)。图 2(a) 中水头值为水轮机工作水头。因水库的不完全多年调节特性和机组避开振动区、尽量在高效率区运行的需求, 导致机组实测数据水头范围分布较窄 (116.00~125.00 m), 高度集中在 48~62 MW 功率区间内。为验证本文方法对全运行水头运行数据的处理效果, 引入水轮机出厂运行特性曲线 (图 2(b)) 进行同步计算验证本文方法对全运行水头运行数据的处理效果。根据水轮机基本工作方程 ($N = gQH\eta$, N 为功率, g 为重力加速度, Q 为发电流量, H 为水轮机工作水头, η 为水轮机效率), 将图 2(b) 转换为图 2(c) 所示数据。

建立移动最小二乘模型时, 设置 20×20 个拟合节点, 采用二次基, 以出力和水头为自变量、流量为因变量构建曲面拟合模型, 进行基于最小二乘法与移动最小二乘法的曲面拟合对比计算分析。

3.2 结果分析

表 1 及图 3、4 分别为最小二乘和移动做最小二乘拟合结果和误差。图 3 中移动最小二乘法的

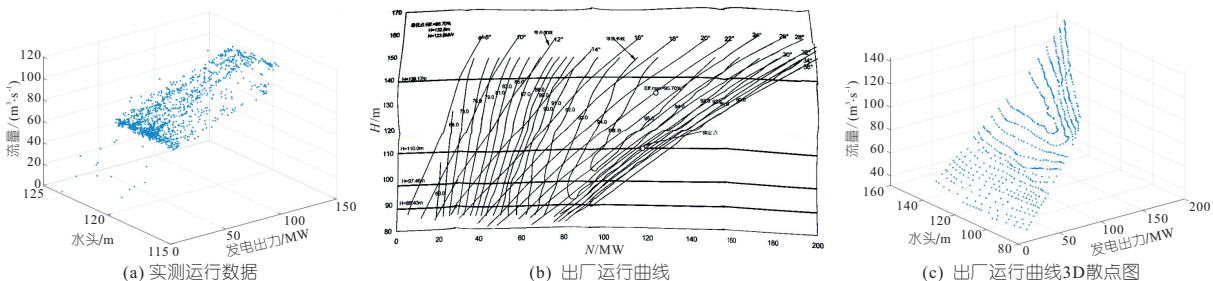


图 2 水电机组运行数据

Fig. 2 Operation data of hydropower unit

表 1 拟合结果和滤波方法对比

Tab. 1 Comparison of fitting results and filter method

数据类别	拟合方法	最大绝对误差	最大相对误差/%	相对误差均值/%	相对误差标准差
运行采样数据	最小二乘法	7.84	14.31	0.07	0.054
出厂运行曲线	移动最小二乘法	3.60	6.58	0.06	0.039
运行曲线	最小二乘法	5.57	0.05	0.06	0.043
运行曲线	移动最小二乘法	3.69	0.04	0.05	0.026

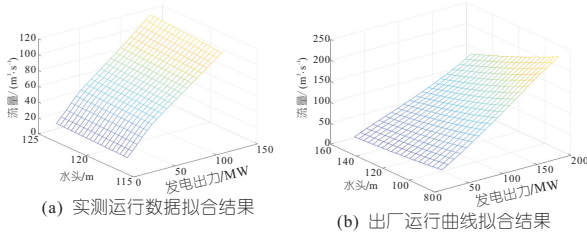


图 3 MLS 拟合结果

Fig. 3 Fitting results of MLS

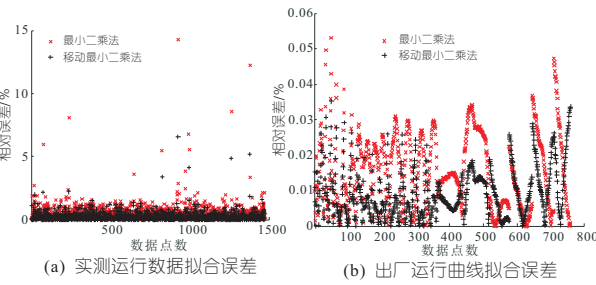


图 4 计算误差比较

Fig. 4 Computational error comparison

拟合曲面整体光滑。由表 1 及图 3、4 可知,移动最小二乘法在支撑域内更好地表达了数据点的局部特征,拟合结果与实测数据更为接近,拟合误差优于最小二乘法。图 3 中超出 5% 误差的数据点有 2/1 472 个,超出 2% 误差的数据点有 9/1 472 个,均可在马氏距离除噪过程中通过限定阈值得到处理。

图 5 为机组实测运行数据与其在出厂运行曲线拟合曲面中插值所得相同工况点的对比。由图 5 可知,在 40~47 MW 区间内,二者流量值相差最小且系相对大小关系发生变化。在功率大于 47 MW 时二者差别较小,前者实测流量值大于后者,差别不大于 3.61%; 小于 40 MW 时差别较大,前者实测流量值小于后者;在这两个区间内二

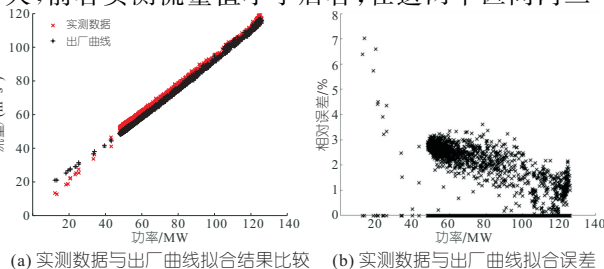


图 5 实测数据与出厂曲线拟合结果比较

Fig. 5 Comparison of measured data and factory curve fitting results

者的差别与功率大小反相关且具有总体一致性特点。由此可见,实测数据与出厂曲线的差别由机组特性和运行条件的差异造成,实测数据可为机组出厂数据的修正提供依据。

综上分析可知,最小二乘曲面拟合趋向于总体近似,当数据分布极不均匀时,会更加依赖数据量较多的区域、忽略数据量较少区域的局部特征。而移动最小二乘近似采用的是局部近似,着重于支撑域范围内对拟合式的影响,能更好地拟合局部区域,反映局部区域的变化趋势,减小不同数据密度对拟合结果的影响。

4 结论

- a. 运用马氏距离有效度量两个未知样本集的相似度和尺度无关的特点,对实测数据进行多次滤波,获得了精度较高的水电机组运行数据特性曲面。
- b. 针对实测水电机组运行数据分布不均匀的特点,移动最小二乘曲面重构法有限考虑局部和整体数据变化趋势的一致性,提出自适应确定支撑半径的方法,可有效避免局部过拟合,提高拟合精度。

参考文献:

- [1] 刘艳娜,郑源,王荣兴.水电机组工作参数测量失误时水轮机效率计算[J].水电能源科学,2021,39(7): 159-162.
- [2] 杨桀彬,杨建东,王超.基于空间曲面的水泵水轮机机组数学模型及仿真[J].水力发电学报,2013, 32(5): 244-245.
- [3] 邵卫云,张雄.水泵水轮机全特性曲线的拟合—移动最小二乘近似[J].水力发电学报,2004(5): 102-106.
- [4] 陈玉,王煜,戴凌全.水轮机模型综合特性曲线数值处理方法研究[J].水资源与水工程学报,2020, 31(3): 155-161.
- [5] 常近时.水力机械装置过渡过程[M].北京:高等教育出版社,2005: 102-104.
- [6] 马伟超,杨桀彬,赵志高,等.混流式水轮机特性曲线在多重边界条件下的分区方法[J].农业工程学报, 2021, 37(11): 31-39.
- [7] 孙佳祥,张成立,徐连奎,等.低比速混流式水轮机特性曲线的近似模拟[J].水电能源科学,2018, 36(7):127-129, 123.
- [8] 许力,田佳乐,齐鹏云,等.基于 MEA-BP 的水轮机运转特性曲线拟合及应用[J].人民长江,2019, 50(9): 141-145.
- [9] 程效军.海量点云数据处理理论与技术[M].上海:同济大学出版社,2014: 58-60.
- [10] 林毅,吉鸿江,韩佳佳,等.一种基于马氏距离的系统故障诊断方法[J].计算机科学,2020, 47(增刊 2): 57-63.

位变幅。枯季提前有利于增加水生植物的生物量,但枯季水位变幅减小可能降低枯季生物多样性。

参考文献:

- [1] 杨桂山,马荣华,张路,等. 中国湖泊现状及面临的重大问题与保护策略[J]. 湖泊科学, 2010,22(6): 799-810.
- [2] 张云昌,张业刚,宋秋龄,等. 江湖关系的历史和未来自[J]. 水利学报, 2021,52(10):1183-1192.
- [3] 周建军,张曼. 近年长江中下游径流节律变化、效应与修复对策[J]. 湖泊科学, 2018,30(6): 1471-1488.
- [4] HU Q, FENG S, GUO H, et al. Interactions of the Yangtze river flow and hydrologic processes of the Poyang Lake, China[J]. Journal of hydrology, 2007,347(1-2): 90-100.
- [5] ZHANG Z T, JIN G Q, TANG H W, et al. How does the three gorges dam affect the spatial and temporal variation of water levels in the Poyang Lake? [J]. Journal of hydrology, 2022, 605;

127356.

- [6] SU L, MIAO C Y, DUAN Q Y, et al. Multiple-wavelet coherence of world's large rivers with meteorological factors and ocean signals[J]. Journal of geophysical research: atmospheres, 2019,124(9): 4932-4954.
- [7] 吴帆. 石臼湖—固城湖圩区景观格局与聚落形态研究[D]. 南京:南京大学,2019.
- [8] 于忠华, 刘海滨,张涨. 石臼湖流域江苏段生态环境驱动因素分析[J]. 水资源保护, 2010,26(6): 70-74.
- [9] 董增川,梁忠民,李大勇,等. 三峡工程对鄱阳湖水资源生态效应的影响[J]. 河海大学学报(自然科学版), 2012,40(1): 13-18.
- [10] COOPS H, HOSPER S H. Water-level management as a tool for the restoration of shallow lakes in the Netherlands [J]. Lake and reservoir management, 2002,18(4):293-298.
- [11] 刘成林,谭胤静,林联盛,等. 鄱阳湖水位变化对候鸟栖息地的影响[J]. 湖泊科学, 2011,23(1):129-135.

Hydrological Rhythm Change of Shijiu Lake and Its Ecological Effect Under New River Lake Relationship

LUO Li-ya¹, TONG Jian¹, ZHOU Jie², PANG Mai-tian³, WANG Yong²

(1. Jiangsu Hydrological and Water Resources Survey Bureau, Nanjing 210029, China;

2. Hydraulic Engineering Department, Nanjing Hydraulic Research Institute, Nanjing 210029, China;

3. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China)

Abstract: In order to quantitatively analyze the impact of upstream reservoirs and the operation of the Three Gorges Project on the hydrological rhythms of the lower reaches of the Yangtze River and lakes, M-K test, cumulative anomaly method, wavelet coherence spectrum method and other hydrological time series trend analysis methods with STL time series decomposition method were used to analyze the 49-year historical water level data of Sheshan Station at Lake Shijiu. The long-term change trend and law of hydrological rhythm of Lake Shijiu under the new relationship between Yangtze River and Lake Shuijiu were analyzed. The coherent relationship between hydrological rhythm changes and local meteorological factors were analyzed along with the ecological effects under the new river-lake relationship. Under the new river-lake relationship, the dry season is earlier and longer and the averaged water level is higher, the extreme low water level is more likely to occur, the variation is gentler, and it is difficult for high water level to happen. The extreme high water level in the flood season is lower and happens later, the averaged water level is lower, ends earlier, and the amplitude is gentler. The early dry season is beneficial to increase aquatic plant biomass, but the decrease of water level variation in dry season may cause biodiversity to decrease.

Key words: hydrological rhythms; ecological effects; Lake Shijiu; time series decomposition

(上接第 186 页)

Construction of Hydraulic Turbine Operating Characteristic Surface Based on Moving Least Square Method

LI Peng¹, LIU Yan-na^{2,3}

(1. Gansu Water Resources and Hydroelectric Investigation and Design & Research Institute Co., Ltd., Lanzhou 730000, China; 2. Guodian Nanjing Automation Co., Ltd., Nanjing 210032, China; 3. Nanjing Hohai Nansi Hydropower Automation Co., Ltd., Nanjing 210032, China)

Abstract: The distribution of measured operating data of hydraulic turbines is non-structural and uneven density. The moving least squares method was used to fit the measured operating parameters to obtain the spatial characteristic surface of the water turbine, which can reflect the local characteristics of data surface and avoid the distortion caused by the differences of data distribution density. The Riyadh criteria was used to perform preliminary data filtering. The support radius was adaptively determined by local fitting across different ranges of discrete sample points. The robust moving least square surface reconstruction model was established combined with the Mahalanobis distance denoising method. Compared with the least squares method, the example calculation results show that the moving least squares method fits the operating characteristics of hydraulic turbines with higher accuracy and supplements the local data effectively. The result can reflect the operating characteristics of hydraulic turbines authenticity, which provides a reference for correcting the operating characteristic curve given by the turbine manufacturer.

Key words: moving least square method; turbine operation data; surface fitting; support radius; characteristic surface