

基于 EMD 距离的燃煤机组煤耗评估分析方法

孙虹, 孙栓柱, 周春蕾, 代家元, 孙彬, 王其祥

(江苏方天电力技术有限公司, 南京 中国 211102)

摘要:燃煤机组班组间运行人员操作习惯的差异常常导致煤耗水平存在优劣之分,因此需要对班组进行对标管理,指导班组优化运行方式.传统上一般通过煤耗的均值、极值等统计指标来评估班组能耗水平,但是这些指标不能完整准确地反映班组的运行水平.为此采用了一种基于 EMD 距离(earth mover's distance)的燃煤机组煤耗评估分析方法,采用直方图来估计煤耗及其影响参数的数据分布,并用 EMD 来定量计算数据分布之间的差异,然后以煤耗最低为目标搜索标杆机组,分析导致其他班组与标杆班组煤耗差异的关键参数,为班组运行优化提供建议.这种方法通过对煤耗关键性能指标数据分布的定量分析,实现了对燃煤机组煤耗水平的精准评估,为电力企业降低机组煤耗提供分析手段,也为政府监管部门开展燃煤机组节能降耗精细化对标管理和性能评估提供了技术支撑.

关键词:燃煤机组;煤耗评估;对标管理;数据分布距离度量;基于直方图的 EMD 算法

中图分类号:TP391 **文献标识码:**A doi:10.3969/j.issn.0253-2778.2015.10.004

引用格式: SUN Hong, SUN Shuanzhu, ZHOU Chunlei, et al. An EMD-based method for assessing and analyzing coal consumption of coal-fired units[J]. Journal of University of Science and Technology of China, 2015, 45(10):822-828.

孙虹, 孙栓柱, 周春蕾, 等. 基于 EMD 距离的燃煤机组煤耗评估分析方法[J]. 中国科学技术大学学报, 2015, 45(10):822-828.

An EMD-based method for assessing and analyzing coal consumption of coal-fired units

SUN Hong, SUN Shuanzhu, ZHOU Chunlei, DAI Jiayuan, SUN Bin, WANG Qixiang

(Jiangsu Frontier Electric Technology CO. LTD, Nanjing 211102, China)

Abstract: Operation habits of coal-fired unit workers often affect significantly the level of coal consumption. In practice, different operation teams usually consume different amounts of coal, thus it is necessary to carry out benchmarking management to guide the workers towards optimal operations. Conventionally, the level of coal consumption within a time period is evaluated by statistics such as the mean, the minima, and the maxima. However, these simple statistics are not informative enough to reveal the operation habits of workers. an EMD-based assessment method is proposed to analyze operation habits of teams. By taking the coal consumption distribution of the best operation team (i. e., the one with minimal amount of total coal consumption) as the baseline comparator, the method first estimates, for non-optimal teams, how far away their operations are from the optimum, and then finds out the primary parameters that cause the difference. Based on such an analysis, suggestions are provided to the operation

收稿日期:2014-09-12; **修回日期:**2014-12-29

基金项目:国家自然科学基金(61473001, 71071045, 71131002); 安徽大学青年科学研究基金(33050054).

作者简介:孙虹(通讯作者),男,1971年生,教授,研究方向:智能信息处理. E-mail:15905166162@163.com

teams, such that their behaviors can be adjusted. The proposed method achieves accurate assessment of coal consumption and may provide an analytical method for electric power enterprises. to save energy and reduce coal consumption. And it also may provide technical support for government supervision departments to carry out fine benchmarking management and performance evaluation of energy-saving for coal-fired units.

Key words: coal-fired unit; coal consumption assessment; benchmarking management; distance metric for data distributions; histogram-based EMD algorithm

0 引言

目前,我国能源结构仍然以煤炭为主,其中火力发电消耗约 60% 的煤炭总量. 根据国家能源局于今年 1 月 16 日发布的我国电力工业统计数据,2014 年全国 6 000 kW 及以上电厂平均供电煤耗为 318 g/kWh,远高于 2020 年 310 g/kWh 的目标值.

燃煤机组基本采用倒班制度运行,即将机组运行人员分成多个运行班组,按照顺序依次倒班操作发电机组. 对于一台机组,不同运行班组间操作人员行为习惯的差异常常导致煤耗水平存在优劣之分,因此需要对班组进行对标管理,指导班组运行方式向标杆机组靠拢以实现节能降耗,进而减少发电成本、降低污染物排放. 以江苏省 135 MW 以上燃煤机组为例,总装机容量为 62 809 MW,如果所有机组供电煤耗整体降低 1 g/kWh,则一年可节约标煤约 34.5 万吨(单台机组年运行时间以 5 500 h 计),节约成本约 1.9 亿元(标煤单价以 560 元/吨计),同时可减少二氧化硫排放约 118.3 吨、减少氮氧化物排放约 51.4 吨,减少烟尘排放约 74.1 吨.

传统上一般通过煤耗及其主要影响参数的统计数据(如平均值、最大值、最小值等)来评估班组间煤耗水平和运行方式的优劣. 然而,上述统计指标仅反映了各参数的平均水平或瞬时状态,不能反映班组的整体运行情况,因此需要对煤耗及其主要影响参数长期的运行情况进行分析,以了解各班组的平均操作水平.

参数的数据分布直方图可以直观地反映参数的长期运行情况,图形的位置(平均值)、形状(方差)可以说明班组对运行参数的控制力度,肉眼无法精确地比较、评估直方图之间的差异,因此需要一种量化手段计算这种差异,对直方图进行定量分析.

本文提出一种基于 EMD 的煤耗自动评估分析方法,根据两个班组间煤耗及其主要影响参数数据分布之间的距离,对燃煤机组不同运行班组煤耗特

性进行对标分析,挖掘导致煤耗偏差的关键因素,为开展燃煤机组调整优化,进而降低机组供电煤耗水平,减少污染物排放奠定基础.

1 问题定义

1.1 问题描述

(I) 机组运行数据

一台机组运行数据是机组运行参数(主汽温度、主汽压力等)和性能指标(煤耗、厂用电率等)在线测量数据或性能计算数据的集合,形如: $S = \{(t_i, \vec{v}_i)\}_{i=1}^N$, 其中 t_i 是时间戳, $\vec{v}_i = (v_{i1}, v_{i2}, \dots, v_{iM})$ 是该时间点上机组运行参数和性能指标的实时数据集.

(II) 运行班组划分

燃煤机组采用倒班工作制度,即将机组运行人员分成 K 个运行班组,记为 c_1, c_2, \dots, c_K . 运行班组按照排班表轮流上班,在规定的时段内操作机组. 据此,我们可以将由 t_1, t_2, \dots, t_N 确定的时间域划分成 K 个区间,每个区间内的所有运行数据属于同一个班组,用 S_{c_k} 表示.

(III) 班组煤耗行为模式

主汽温度、主汽压力、给水流量等主要运行参数会对煤耗产生影响,不同班组煤耗及其影响参数的不同组合方式(取值)反映了该班组的煤耗行为模式. 每个班组操作时段内的煤耗及其影响参数的实时数据,形成了该班组的煤耗行为模式矩阵 $S_{c_k}^m$, 矩阵的列是煤耗或一个影响参数,矩阵的行对应一个时刻. 每个班组的每个参数值或煤耗值(矩阵的一列)的变化反映了该班组的操作行为.

(IV) 班组煤耗分析任务

班组煤耗分析任务,就是:分析不同班组煤耗行为模式 $S_{c_1}^m, S_{c_2}^m, \dots, S_{c_K}^m$ 之间的差异,找出导致煤耗行为模式产生差异的关键参数.

1.2 解决方案

上述问题的解决方案如下:

(I) 根据班组煤耗模式矩阵 $S_{c_k}^m$, 利用直方图法估计班组煤耗分布 $D_{c_k}^m$, 将班组煤耗模式数据转化成概率分布来处理;

(II) 比较各班组煤耗分布 $D_{c_k}^m$ 的煤耗平均值, 输出煤耗均值较低的班组及其煤耗均值;

(III) 用 EMD 算法计算任意两个班组煤耗分布 D_c^m 和 $D_{c'}^m$ 之间的距离;

(IV) 计算影响煤耗的各参数数据分布之间的 EMD 距离;

(V) 按从大到小的顺序对上述参数的 EMD 距离排序, 其中 EMD 距离最大的参数, 就被认为是导致这两个班组煤耗差异的关键因素。

2 数据分布的距离度量

统计学上, 为了比较两个数据分布之间的差异, 引入了很多距离度量指标, 常用的有: KL 距离、Pearson- χ^2 距离、全变差距离 (total variation metric) 和 EMD (earth mover's distance) 距离等。下面将简单介绍前三种距离度量, EMD 距离将在下文详细介绍。

2.1 KL 距离(相对熵)

KL 距离 (Kullback-Leibler divergence) 又叫 KL 散度, 它衡量了两个概率分布 P 和 Q 之间的差异性, 其定义如下^[1]:

$$D(P, Q) = \sum P(i) \log \frac{P(i)}{Q(i)}.$$

KL 距离越小, P 和 Q 之间就越接近。KL 距离更关注两个分布之间的元素比率, 而不是增幅。当两个元素之间的比率较小时, 这个比率的对数几乎为零^[2]。当 $Q(i) = 0$ 时, KL 距离就不再适用了。

2.2 Pearson- χ^2 距离

Pearson- χ^2 距离常被作为衡量标准来判断一个密度函数列是否收敛到某个确定的密度函数^[3]。如果 P 和 Q 是离散随机变量的概率密度分布, 则 Pearson- χ^2 距离定义如下:

$$d^2(P, Q) = \sum \frac{P^2(i)}{Q(i)} - 1.$$

与 KL 距离类似, Pearson- χ^2 距离不适用于 $Q(i) = 0$ 的情况。

2.3 全变差距离

全变差距离是将每个点上的概率分布之差进行累加, 其计算公式如下:

$$T(P, Q) = \frac{1}{2} \sum (P(i) - Q(i)) \quad (1)$$

从式(1)可以看出, 全变差距离和前两种距离一样, 都没有考虑到变量值的差异对分布距离度量的影响。例如, 有三个煤耗概率分布直方图 A 、 B 、 C , 假设 A 在 300 g/kWh 处的频率为 $A(300) = 0.5$; B 在 301 g/kWh 处的频率为 $B(301) = 0.5$; C 在 310 g/kWh 处的频率为 $C(310) = 0.5$, 则 $T(A, B) = T(A, C) = 0.5$ 。根据我们的专业知识, 这个结论是不正确的, A 和 B 之间的距离应该要小于 A 和 C 之间的距离, 因为对于煤耗这个指标, 1g 的差异就能使得机组经济效益发生明显差异, 因此 10g 的差距是巨大的, 这时频率的差异不再是主要的影响因素, 因此在计算煤耗数据分布的差异时, 必须同时考虑煤耗的数值和分布频率。基于这个原因, 我们选择 EMD 距离来衡量煤耗及其影响参数的数据分布距离。

3 基于 EMD 的班组煤耗评估分析方法

3.1 直方图密度估计

(I) 概率密度函数

在数学中, 连续型随机变量 X 的概率密度函数 (probability density function, PDF) 是一个描述变量 X 在某个确定的取值点 x 附近的可能性的函数。用 f 表示这个函数, 那么 $f(x)$ 具有以下性质:

$$f(x) \geq 0,$$

$$\int_{-\infty}^{+\infty} f(x) dx = 1,$$

$$P(a < x \leq b) = \int_a^b f(x) dx.$$

已知一组样本数据, 若要估计其概率密度函数 f , 主要有两类方法: 参数估计和非参数估计。

参数估计需要事先知道分布 (假定样本数据服从某种分布), 再用样本数据计算其中的参数。常用的参数估计方法有最小二乘法、极大似然法和贝叶斯估计法。参数估计要求密度函数的形式已知, 但这种假定有时并不成立, 这种情况下就很难拟合出实际的概率密度函数。因此我们主要采用非参数估计。

非参数估计不需要预先假定密度函数的数学模型, 对密度函数模型的限定较少, 直接利用样本中的一般信息, 对总体的形状、特征进行估计^[4]。非参数估计具有以下优点: 假设条件少, 适用范围广泛; 运算简单, 可以快速得到结果; 方法直观, 容易理解; 能够适应名义尺度和顺序尺度等对象。常用的非参数

估计方法有:直方图密度估计、核密度估计、最近邻估计等。

(II)直方图密度估计

直方图密度估计是运用最早、最广泛的密度估计方法,它是用一组样本构造概率密度的经典方法^[5]。直方图密度估计通常的做法如下:

①将数据值覆盖的数据区间等分成几个子区间(bin);

②每当一个数据值落到某个子区间,这个子区间块的高度就相应地增加一个单位。

假设有 N 个样本数据,区间宽度为 V ,那么单位高度为 $1/(NV)$;若一个区间内有 K 个样本,那么它的高度为 $K/(NV)$ 。这样我们就利用样本数据构造出了概率密度函数。当数据维度较少时(如 1~2 维),我们一般使用直方图密度估计法。

直方图密度估计法概念简单易于使用,在样本容量较大、区间宽度较小的情况下,所得的密度函数能够较好地接近真实的函数。

由于在我们的应用中需要计算煤耗和相关运行参数分布的距离,而直方图密度估计法得到的分布能够很容易地转化为离散分布,便于分布间距离的计算,因此我们采用直方图密度估计方法。

3.2 EMD 距离计算

(I) EMD 距离

EMD 最初由 Rubner 等于 1997 年提出^[6-7],它是一种衡量在域 D 上的两个概率分布间距离的度量。首先,定义域 D 上有 N 个元素的离散概率分布为: $S = \{s_j = (\omega_j, m_j)\}_{j=1}^N$,其中 m_j 是第 j 个元素的位置, ω_j 是该元素的权重。

假设有两个离散分布 $P = \{(p_i, u_i)\}_{i=1}^n$ 、 $Q = \{(q_j, v_j)\}_{j=1}^n$,则这两个分布之间的距离被认为是一个运输问题(transportation problem),即将 P 中的元素看作 u_i 处的供应者,将 Q 中的元素看作 v_j 处的需求者,则 p_i 、 q_j 分别代表供应量和需求量,EMD 就被定义为解决供需运输问题的最小工作量,计算公式如下:

$$EMD(P, Q) = \min_{F \in \{f_{i,j}\}} \frac{\sum_{i,j} f_{i,j} d_{i,j}}{\sum_{i,j} f_{i,j}} \quad (2)$$

$$s. t. \quad \sum_j f_{i,j} = p_i \quad (3)$$

$$\sum_i f_{i,j} = q_j \quad (4)$$

$$\sum_{i,j} f_{i,j} = \min\left\{\sum_i p_i, \sum_j q_j\right\} \quad (5)$$

$$f_{i,j} \geq 0 \quad (6)$$

式中, $F = \{f_{i,j}\}$ 代表一个流的集合,其元素 $f_{i,j}$ 是从第 i 个供应者到第 j 个需求者之间的运输量, $d_{i,j}$ 是位置 u_i 和 v_j 间的地面距离。

分布比较问题可以转化成运输问题来求解。运输问题是一个特殊的线性规划问题,其约束矩阵具有非常稀疏的结构。基于标准单纯型法的 TS (transportation simplex)算法^[8]充分利用约束矩阵的这种特殊结构,极大地降低了运算量。对于有 N 个元素的分布,这种算法具有 $O(N^3 \log N)$ 的时间复杂度,计算成本比较昂贵。

(II)直方图间的 EMD 距离

由于我们选用直方图法来估计参数的概率密度函数,因此需要寻找新的 EMD 算法,以便计算两个直方图分布之间的距离^[9]。

在用直方图表示的数据分布问题中,每个柱子可以被看做是离散概率分布 S 中的一个元素 s_j ,柱子代表的值对应于权重 ω_j ,而柱子的位置(下标)则对应于位置 m_j 。

以两维直方图为例,假设:

①将直方图分成 m 行、 n 列,则直方图共有 $N = m \times n$ 个柱子;

②这些柱子的下标形成集合 $L = \{(i, j): 1 \leq i \leq m, 1 \leq j \leq n\}$ 。在下文中用 (i, j) 来代表一个柱子;

③流 $f_{i,j}$ 的下标集合定义如下: $J = \{(i, j, k, l): (i, j) \in L, (k, l) \in L\}$;

④两个待比较的直方图为 $P = \{p_{ij}: (i, j) \in L\}$ 和 $Q = \{q_{kl}: (k, l) \in L\}$ 。

则两个直方图 P 和 Q 之间的 EMD 公式如下:

$$EMD(P, Q) = \min_{F \in \{f_{i,j,k,l} \in J\}} \sum_j f_{i,j,k,l} d_{i,j,k,l} \quad (7)$$

$$s. t. \quad \sum_{\langle k,l \rangle \in L} f_{i,j,k,l} = p_{ij}, \forall (i, j) \in L \quad (8)$$

$$\sum_{\langle i,j \rangle \in L} f_{i,j,k,l} = q_{kl}, \forall (k, l) \in L \quad (9)$$

$$f_{i,j,k,l} \geq 0, \forall (i, j, k, l) \in J \quad (10)$$

式中, F 是从 P 到 Q 的流, $f_{i,j,k,l}$ 则代表柱子 (i, j) 到柱子 (k, l) 的流。满足上式的流 F 被称为可行解。

式(7)中的地面距离 $d_{i,j,k,l}$ 一般采用 L_p 距离,其计算公式为:

$$d_{i,j,k,l} = \| (i, j)^T - (k, l)^T \|_p = (|i-k|^p + |j-l|^p)^{1/p}$$

为了进一步降低计算量和时间复杂度,我们选

用 L_1 距离作为地面距离, 则 $d_{i,j;k,l}$ 计算公式简化为如下形式:

$$d_{i,j;k,l} = |i-k| + |j-l|.$$

以 L_1 距离作为地面距离的 EMD 算法, 仅有 $O(N)$ 个未知数, 远少于原始 EMD 距离中的 $O(N^2)$ 个未知数, 约束条件的数量也减半, 时间复杂度也降至 $O(N^2)$, 显著改善了 TS 算法的效率^[9].

4 方法应用效果

我们将基于 EMD 的班组煤耗评估分析方法应用到 6 台不同容量等级的典型机组(300 MW、600 MW、1000 MW 容量等级各 2 台机组)上, 以 2014 年机组煤耗数据为样本进行实验. 我们以其中一台

330 MW 机组为例, 详细说明方法的应用效果.

该机组执行五班三运倒班方式, 机组运行人员被分成 5 个班组($k=5$), 分别记为班组 1、班组 2、……、班组 5.

4.1 班组煤耗比较

以该机组 2014 年煤耗模式实时数据为样本进行班组煤耗比较分析, 各班组煤耗数据分布直方图如图 1 所示. 从图 1 可以看出, 班组 2 和班组 5 的煤耗分布差异最大, 这也与表 1 中各班组煤耗分布 EMD 距离计算结果相符. 同时, 班组 5 的煤耗至少有 50% 分布在小于 301 g/kWh 的区域内, 是所有班组中煤耗表现最好的, 因此选择班组 5 作为标杆班组.

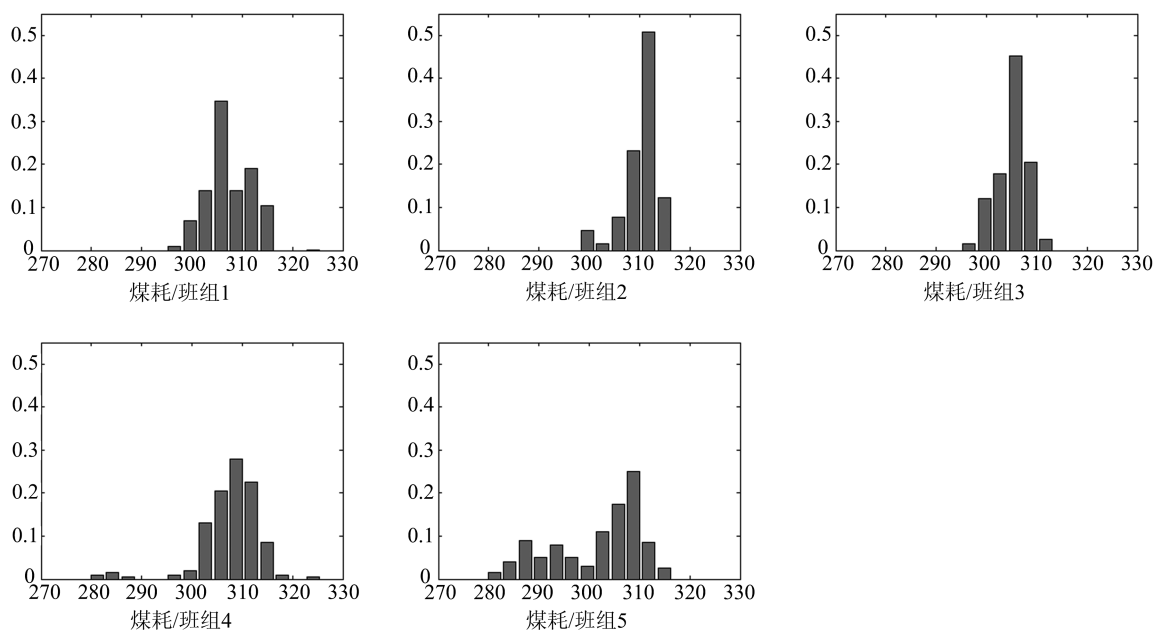


图 1 各班组煤耗数据分布直方图

Fig. 1 Coal consumption histograms of 5 teams

表 1 各班组间煤耗直方图 EMD 距离

Tab. 1 EMD of coal consumption histograms between teams

	班组 1	班组 2	班组 3	班组 4	班组 5
班组 1	—	0.069 8	0.052 9	0.030 8	0.133 2
班组 2	—	—	0.122 7	0.065 5	0.203 0
班组 3	—	—	—	0.082 1	0.114 2
班组 4	—	—	—	—	0.141 1
班组 5	—	—	—	—	—

4.2 班组煤耗分析

在确定班组 5 作为标杆班组后, 需要进一步分

析各班组影响煤耗的参数分布情况与标杆班组参数分布情况的差距, 定位导致煤耗发生差异的关键因素.

为了进行说明, 在本次应用中仅选取主蒸汽温度、再热蒸汽温度、给水温度、主蒸汽压力、再热蒸汽压力、排烟温度、过热减温水流量、再热减温水流量共 8 个影响煤耗的主要参数, 以班组 5 的参数分布为比较基准, 计算煤耗表现最差的班组 2 与标杆班组之间上述 8 个指标的 EMD 距离, 根据计算结果确定影响煤耗的关键因素. 这两个班组的 EMD 距离按从大到小的顺序排列, 结果如表 2 所示.

表 2 班组 2 与标杆班组间各参数分布 EMD 距离

Tab. 2 EMD of parameters between the team 2 and the benchmarking team

参数	单位	EMD
过热减温水流量	t/h	0.209 3
主蒸汽压力	MP	0.165 8
给水温度	℃	0.159 5
再热蒸汽温度	℃	0.156 7
再热蒸汽压力	MP	0.139 8
排烟温度	℃	0.138 0
再热减温水流量	t/h	0.110 0
主蒸汽温度	℃	0.045 5

从表 2 可以看出,两个班组间过热减温水流量的 EMD 距离最大,主蒸汽温度的 EMD 距离最小,说明前者是导致这两个班组煤耗差异的首要因素,后者对这两个班组煤耗的煤耗影响最小.图 2 和图 3 分别是这两个班组过热减温水流量、主蒸汽温度的直方图,其数据分布情况与 EMD 距离的计算结果一致.

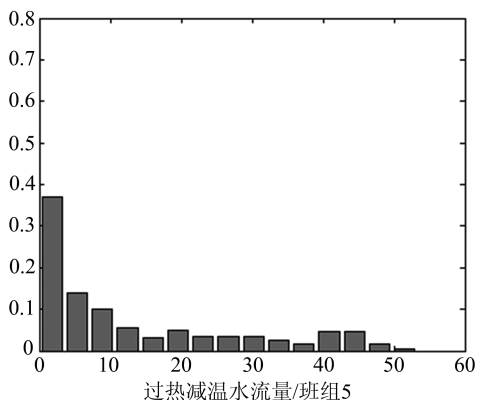
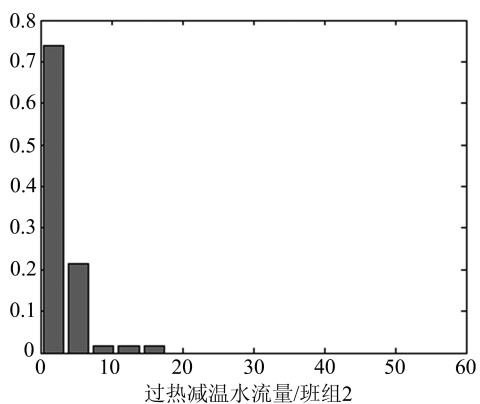


图 2 班组 2、班组 5 的过热减温水流量直方图

Fig. 2 Superheated water flow histograms of team 2 and team 5

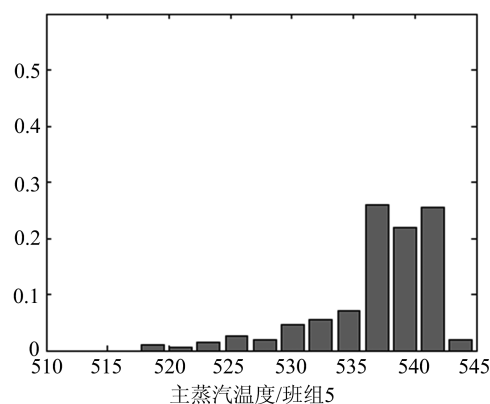
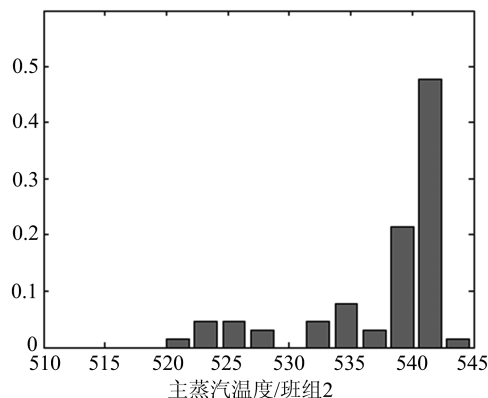


图 3 班组 2、班组 5 的主蒸汽温度直方图

Fig. 3 Main steam temperature histograms of team 2 and team 5

4.3 经济效益分析

该 330 MW 机组年运行时间以 5 500 h 计,则平均每个班组运行 1 100 h.如果其他班组能按照标杆班组(班组 5)的方式运行机组,则 2014 年该机组能节省燃煤成本 503.44 万元.各班组与标杆班组的煤耗及燃煤经济效益指标如表 3 所示.

表 3 各班组与标杆班组间节能经济效益比较

Tab. 3 Energy-saving economic benefits comparison between benchmarking team and other teams

	煤耗/ (克/千瓦时)	用煤量/ (万吨)	燃煤成本/ (万元)	超标杆 燃煤成本/ (万元)
班组 5	301.5	10.944	6 128.64	—
班组 1	307.5	11.162	6 250.72	122.08
班组 2	310.3	11.264	6 307.84	179.20
班组 3	305.2	11.079	6 204.24	75.60
班组 4	307.7	11.170	6 255.20	126.56
总计	—	—	—	503.44

5 结论

基于 EMD 的燃煤机组煤耗评估分析方法,利用海量历史运行数据,采用直方图方法估计煤耗及其影响参数的数据分布概率密度;然后,以此为基础,运用基于直方图的 EMD 算法,对煤耗及其影响参数之间的整体差异进行量化评估,以煤耗最低为目标搜索标杆机组,根据 EMD 距离快速查找影响煤耗的关键因素,为开展燃煤机组对标管理、运行优化,进而降低机组煤耗水平提供了分析手段。

基于 EMD 的燃煤机组煤耗评估分析方法,计算过程无需人工干预,可以在后台进行,适用于在线分析应用场景;同时,建模过程仅需提供煤耗及其影响参数的样本数据,无需其他专业知识,因此模型能提供客观的分析结果,不受个人经验影响,确保了算法分析结果的客观公正性。

在后续工作中,将进一步深化应用,在业务层面,横向上将应用范围扩展至环保指标,实现对全省燃煤机组节能减排 KPI 指标的全方位综合评估;纵向上,将所有影响煤耗的运行参数都纳入评估范围,进一步提高分析结果的可靠性。在技术层面,将充分利用大数据平台的并行处理能力,实现省级燃煤机组节能减排 KPI 指标的全方位在线评估,为政府监管部门实现对燃煤机组节能减排水平的动态实时对标管理提供强有力的技术支撑。

参考文献(References)

- [1] 许震. 基于 KL 距离的半监督分类算法[D]. 上海: 复旦大学, 2010.
- [2] 李中魁. 基于动态阈值的网络流量异常检测方法研究与实现[D]. 成都: 电子科技大学, 2010.
- [3] 李开灿. Pearson- χ^2 距离的若干性质[J]. 数学的实践与认识, 2003, 33(1): 49-53.
- LiK C. On the properties of the distance of Pearson- χ^2 [J]. Mathematics in Practice and theory, 2003, 33 (1): 49-53.
- [4] 于春蕾. 基于非参数统计的判别分析[D]. 济南: 山东大学, 2013.
- [5] 牛君. 基于非参数密度估计点样本分析建模的应用研究[D]. 济南: 山东大学, 2007.
- [6] Rubner Y, Tomasi C, Guibas L J. The Earth Mover's distance as a metric for image retrieval [J]. International Journal of Computer Vision, 2000, 40 (2): 99-121.
- [7] Rubner Y, Tomasi C, Guibas L J. A metric for distributions with applications to image databases[C] // Proceedings of the International Conference on Computer Vision. Bombay, India: Narosa Publishing House, 1998: 59-66.
- [8] Hillier F S, Lieberman G J. Introduction to Mathematical Programming[M]. New York: McGraw-Hill, 1990.
- [9] Ling H B, Okada K. An efficient earth mover's distance algorithm for robust histogram comparison [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(5): 840-853.
- [1] 许震. 基于 KL 距离的半监督分类算法[D]. 上海: 复