DOI:10.11918/j.issn.0367-6234.201701033

# 污水处理出水 BOD 区间预测建模

柴 伟[1,2,3], 纪镐南[1,2,3]

(1.北京工业大学 信息学部自动化学院,北京 100124;2.计算智能与智能系统北京市重点实验室(北京工业大学),北京 100124;3.数字社区教育部工程研究中心(北京工业大学),北京 100124)

**摘 要**: 生化需氧量(BOD)是评价水质的重要指标,也是污水处理过程中直接控制的参数.为了提高污水处理质量,需要寻找BOD的有效测量方法.本文给出一种新的BOD软测量方法,可以实现其保证估计.采用主元分析方法选取BOD预测的主要辅助变量.利用径向基函数神经网络的逼近能力,将其用于污水处理出水BOD软测量建模.径向基函数神经网络的中心被确定之后,考虑到建模误差有界,使用参数线性集员辨识算法得到网络输出权值的集合描述.在污水处理系统运行过程中,所建立好的软测量模型可以预测出水BOD的上下界.此外,建立多个软测量模型,并将多模型测量结果进行融合以降低单一模型所给结果的保守性.实验结果表明本文方法的有效性.

**关键词**: 软测量;污水处理;径向基函数神经网络;集员辨识;区间预测

**中图分类号**: TP273 **文献标志码**: A **文章编号**: 0367-6234(2018)02-0071-06

# Interval predictor models for effluent BOD of wastewater treatment

CHAI Wei[1,2,3], JI Haonan[1,2,3]

(1.Faculty of Information Technology, School of Automation, Beijing University of Technology, Beijing 100124, China;
2.Beijing Key Laboratory of Computational Intelligence and Intelligent Systems (Beijing University of Technology), Beijing 100124, China;
3.Engineering Research Center of Digital Community (Beijing University of Technology), Ministry of Education, Beijing 100124, China)

**Abstract**: Biochemical oxygen demand (BOD) is an important index for evaluating water quality, and a variable directly controlled in the wastewater treatment process. To improve the performance of wastewater treatment, it is necessary to find out an effective method for measuring BOD. This paper presents a new soft measurement which can provide guaranteed estimation of the effluent BOD. The principal component analysis is utilized to select the secondary variables for the soft sensor. In virtue of its simple topological structure and universal approximation ability, the radial basic function neural network (RBFNN) is utilized in the soft sensor modeling. Considering the bounded modeling error, linear-in-parameters set membership identification algorithm is used to obtain a description of the uncertain set of the output weights after the determination of centers of the RBFNN. The RBFNN model with uncertain output weights can predict the upper and lower bounds of the effluent BOD during the wastewater treatment. Besides, a bundle of soft sensors is constructed and the intersection of the results given by the soft sensors is used to lower the conservatism by using a single sensor. Experiment results show the satisfying performance of the proposed method.

**Keywords**: soft measurement; wastewater treatment; radial basic function neural network; set membership identification; interval prediction

随着中国水环境问题的日益突出,污水处理作为保护环境的重要措施,受到全社会越来越多的关注.生化需氧量(BOD)反映污水中生物可降解有机物含量.是评价水质的重要指标,也是污水处理过程中直接控制的参数,能否对BOD进行实时监测已成为提高治污质量的关键.由于仪器水平或成本方面的原因,目前,污水处理厂多采用离线人工化验的方法:$BOD_5$离线分析测定周期为5 d,检测存在滞后,直接影响污水处理过程的实时控制和优化操作.实现BOD的高品质在线实时监测是个难题.软测量技术具有成本低、时效性好和准确度高等优点,被引入到污水处理过程中,为出水BOD高品质实时监测提供一个有效途径[1].

污水处理过程具有机理复杂、强非线性和高度不确定性等特点,因而基于生化反应机理建立软测量模型的传统方法效果欠佳.而神经网络凭借其优良的逼近能力被广泛应用于污水处理出水BOD软测量建模中.祁国强等[2]提出基于径向基函数(RBF)神经网络的软测量方法,运用大量实测数据对RBF神经网络进行训练和仿真,通过实验证明基

**收稿日期**: 2017-01-10
**基金项目**: 北京市自然科学基金(4144067)
**作者简介**: 柴 伟(1981—),男,讲师,硕士生导师
**通信作者**: 柴 伟,chaiwei@bjut.edu.cn

于 RBF 神经网络的软测量模型的有效性. Huang 等[3]建立基于粒子群的改进最小二乘支持向量机回归模型,实现出水 BOD、化学需氧量(COD)、总氮(TN)、氨氮($NH_4$-N)、总固体悬浮物(TSS)等浓度的预测,取得较精确的结果.王树东等[4]提出基于混合递阶遗传算法优化 RBF 神经网络的 BOD 软测量方法,并通过实验证明此方法具有较好的实时性、稳定性和较高的预测精度.许玉格等[5]提出一种基于快速相关向量机(Fast-RVM)的在线软测量回归模型来实时预测出水 BOD,并对比支持向量机和相关向量机在离线和在线情况下的预测效果,通过实验证明基于 Fast-RVM 的在线软测量回归模型能取得更好的预测效果. Qiao 等[6]将 COD、固体悬浮物(SS)、pH、溶解氧(DO)作为输入变量,利用改进的 T-S 模糊神经网络模型实现出水 BOD 的预测,取得较好的结果.当前工作多是获得一个单一值预测出水 BOD 浓度,而未曾给出评估真实值和预测值之间偏差(即预测误差)的有效方法.即使已采用均方根误差(RMSE)来评估神经网络建模精度,但通过此指标不易直接导出各时刻预测误差的计算公式.目前,围绕污水处理区间预测模型开展研究的工作尚不多见[7-8].

系统辨识理论中的区间预测方法有很多[9-16],基本上可以分为两大类.一类方法基于随机误差假设,如最小二乘法或极大似然法,要求误差的统计特性已知或部分已知,然后通过概率论给出被估计量的置信区间[9-10].另一类方法基于有界误差假设,如集员辨识,只要求误差有界而不需要知道其统计特性,然后通过集理论给出被估计量的置信区间[11-16].文献[7-8]提出的污水处理软测量方法即属于第一类.而在实际应用中误差的统计特性可能难以获得,此时第一类方法给出的置信区间不一定可靠.为此,给出一种新的 BOD 软测量方法,可以实现其保证估计.利用径向基函数神经网络的逼近能力,将其用于污水处理出水 BOD 软测量建模.径向基函数神经网络的中心被确定之后,考虑到建模误差有界,使用参数线性集员辨识算法得到网络输出权值的集合描述.在污水处理系统运行过程中,所建立好的软测量模型可以预测出水 BOD 的上下界.此外,本文建立多个软测量模型,并将多模型测量结果进行融合以降低单一模型所给结果的保守性.

## 1 BOD 软测量模型

### 1.1 软测量模型整体结构

神经网络软测量模型的一般设计步骤为:辅助变量的初选、现场(原始)数据采集、数据预处理、软测量模型搭建.辅助变量的初选是在可测变量集中初步选择与被估计变量有关的原始辅助变量,这些变量中部分可能是相关变量.数据预处理利用现场的历史数据作统计分析计算,将原始辅助变量化成少数几个互不相关的变量作为神经网络模型的输入.这里现场数据指某污水处理厂记录的历史数据及实际运行数据.本文设计污水处理出水 BOD 软测量模型整体结构如图 1 所示.

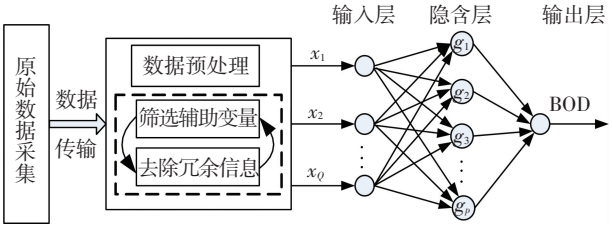



**图 1 BOD 软测量模型整体结构**

Fig.1 Structure of the soft sensor for BOD

### 1.2 数据预处理

数据预处理主要是去除冗余信息,选取 BOD 预测的主要辅助变量.由于取自污水处理现场的参数不是全部有助于出水水质参数 BOD 的预测,在预测前首先需要确定选择哪些变量作为软测量模型的输入.筛选辅助变量的主要目的是:简化神经网络模型的输入,提高网络的预测精度与计算效率.主元分析(PCA)是最常见的方法[17],该方法通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量,转换后的这组变量叫主成分.限于篇幅和突出重点,PCA 的具体计算公式参考文献[17].

### 1.3 RBF 神经网络模型

RBF 神经网络因为其简单的拓扑结构和强大的逼近能力,在软测量建模中得到广泛的应用.RBF 神经网络是典型的三层前馈网络,包括输入层、非线性隐含层和线性输出层,可以表示为

$$y = \sum_{i=1}^{p} w_i g_i(x, \boldsymbol{\sigma}, c_i). \tag{1}$$

式中:$y \in \mathbf{R}$ 和 $x \in \mathbf{R}^Q$ 分别为网络的输出和输入,在软测量模型中分别代表出水 BOD 和辅助变量,$Q$ 表示辅助变量的个数;$g_i(x, \boldsymbol{\sigma}, c_i)$ 为神经网络第 $i$ 个隐含层节点的输出,$c_i \in \mathbf{R}^Q$ 和 $\boldsymbol{\sigma} \in \mathbf{R}$ 分别为中心和宽度;$w_i \in \mathbf{R}$ 为线性输出权值;$p$ 为隐含层节点个数.RBF 神经网络的隐单元基函数

$$g_i(x, \boldsymbol{\sigma}, c_i) = \exp\left(-\frac{\|x - c_i\|^2}{\boldsymbol{\sigma}^2}\right), \tag{2}$$

式中$\|\cdot\|$表示欧式范数.从式(1)可以看出,RBF 神经网络从隐含层到输出层是线性传递.

将降维后的数据分成两部分.一份用于神经网

络模型建立,表示为 $D_1=\{x_k,y_k\}_{k=1}^{N}$;另一份用于模型预测,表示为 $D_2=\{x_k,y_k\}_{k=N+1}^{L}$.软测量模型建立好后,预测出水 BOD 的过程可以表示为

$$\hat{y}_k=\sum_{i=1}^{p}\hat{w}_i g_i(x_k,\sigma,c_i),N+1\leqslant k\leqslant L. \quad (3)$$

式中:$\hat{y}_k\in\mathbf{R}$ 为神经网络的预测输出,$\hat{w}_i\in\mathbf{R}$ 为神经网络输出权值的估计.

如果采用集员辨识算法估计 RBF 神经网络输出权值 $w_i$,则可以预测出水 BOD 的上下界.

## 2　BOD 置信区间计算

### 2.1　置信区间的基本计算

采用聚类算法[18]确定 RBF 神经网络的中心 $c_i$ 和宽度 $\sigma$.此算法的优点在于不需要知道数据集的任何先验知识,一旦神经网络隐含层中心和宽度被确定,就可以得到关于输出权值 $w_i$ 的线性方程

$$y_k=\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\theta}+e_k=\sum_{i=1}^{p}w_i g_i(x_k,\sigma,c_i)+e_k. \quad (4)$$

式中

$$\boldsymbol{\phi}_k=[g_1(x_k,\sigma,c_1),g_2(x_k,\sigma,c_2),\cdots,g_p(x_k,\sigma,c_p)]^{\mathrm{T}}, \quad (5)$$

$$\boldsymbol{\theta}=[w_1,w_2,\cdots,w_p]^{\mathrm{T}}. \quad (6)$$

$e_k$ 包含 RBF 神经网络中心、宽度和输出权值的选择带来的建模误差,而建模误差是与神经网络逼近能力有关的有界量,则可认为 $e_k$ 为有界误差,即 $|e_k|\leqslant\varepsilon$.

集员辨识是一种有效的可以替代随机参数估计(如最小二乘和极大似然估计)的方法,因为其在应用时只需要知道误差的上下界.这里采用集员辨识方法估计神经网络的输出权值.一方面是因为这里的估计问题符合集员辨识的应用条件,另一方面以集员辨识的估计结果为基础可以给出 BOD 的置信区间.

由样本数据 $D_1=\{x_k,y_k\}_{k=1}^{N}$、误差的有界假设 $|e_k|\leqslant\varepsilon$ 以及式(4)可以得到

$$S_k=\{\theta\in\mathbf{R}^p:|y_k-\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\theta}|\leqslant\varepsilon\},\ k=1,2,\cdots,N. \quad (7)$$

$N$ 个 $S_k$ 的交集可得到一个凸多面体 $\boldsymbol{\Theta}_N\subset\mathbf{R}^p$,即 $\boldsymbol{\Theta}_N=\cap_{k=1}^{N}S_k$.这里 $\boldsymbol{\Theta}_N$ 称为权值向量 $\boldsymbol{\theta}$ 的不确定集.权值不确定性 $\theta\in\boldsymbol{\Theta}_N$ 使模型输出为区间 $[\hat{y}_k^{\mathrm{m}},\hat{y}_k^{\mathrm{M}}]$,$k\geqslant N+1$.式中

$$\hat{y}_k^{\mathrm{m}}=\min_{\theta\in\boldsymbol{\Theta}_N}\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\theta},\hat{y}_k^{\mathrm{M}}=\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_N}\phi_k^{\mathrm{T}}\boldsymbol{\theta}. \quad (8)$$

这个区间表示出神经网络模型输出 $\hat{y}_k$ 的不确定性,即 $\hat{y}_k\in[\hat{y}_k^{\mathrm{m}},\hat{y}_k^{\mathrm{M}}]$,$k\geqslant N+1$.此区间称为神经网络输出的置信区间.根据 $\theta\in\boldsymbol{\Theta}_N$、式(4)和误差有界假设 $|e_k|\leqslant\varepsilon$,可得到出水 BOD 的置信区间

$$y_k\in[\hat{y}_k^{\mathrm{m}}-\varepsilon,\hat{y}_k^{\mathrm{M}}+\varepsilon],\ k\geqslant N+1. \quad (9)$$

出水 BOD 的置信区间式(8)和(9)可通过线性规划的方法得到,但是当隐含层节点数太多和/或样本数据量很大时,该方法计算成本过重.为了降低计算成本,采用集员辨识中的椭球外界算法给出一个包含权值不确定性集 $\boldsymbol{\Theta}_N$ 的椭球

$$E_N=\{\theta\in\mathbf{R}^p:(\theta-\hat{\theta}_N)^{\mathrm{T}}\boldsymbol{P}_N^{-1}(\theta-\hat{\theta}_N)\leqslant 1\}. \quad (10)$$

式中:$\hat{\theta}_N\in\mathbf{R}^p$ 为椭球 $E_N$ 的中心,$\boldsymbol{P}_N\in\mathbf{R}^{p\times p}$ 是一个表征椭球形状和大小的正定矩阵.椭球 $E_N$ 的计算过程如下[19-20].

**算法 1**

初始化.置 $\hat{\theta}_0=0$,$\boldsymbol{P}_0=\delta^{-1}\boldsymbol{I}$,$\boldsymbol{I}$ 为单位阵,$\delta$ 取小的正数,如 $10^{-5}$.

递推:　对 $k=1,2,\cdots,N$.

**步骤 1(a)**　计算

$$\beta_k=\boldsymbol{\phi}_k^{\mathrm{T}}P_{k-1}\boldsymbol{\phi}_k, \quad (11)$$

$$v_k=y_k-\boldsymbol{\phi}_k^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{k-1}. \quad (12)$$

**步骤 1(b)**　如果 $\varepsilon\geqslant\sqrt{\beta_k}+|v_k|$,置 $\hat{\theta}_k=\hat{\theta}_{k-1}$,$P_k=P_{k-1}$,转步骤 3.

**步骤 1(c)**　如果 $p(\varepsilon+v_k)<\sqrt{\beta_k}<\varepsilon-v_k$,则由式(13)和(14)计算 $\varepsilon'$ 和 $v_k'$ 的值,并由式(15)求得 $q_k$,将 $\varepsilon'$ 和 $v_k'$ 的值分别赋给 $\varepsilon$ 和 $v_k$ 之后,转步骤 2.

$$\varepsilon'=\frac{v_k+\varepsilon+\sqrt{\beta_k}}{2}, \quad (13)$$

$$v_k'=\frac{v_k+\varepsilon-\sqrt{\beta_k}}{2}, \quad (14)$$

$$q_k=\begin{cases}\dfrac{\varepsilon'(\sqrt{\beta_k}-p(\varepsilon+v_k))}{(p-1)\beta_k}, & \text{当}\sqrt{\beta_k}>p(\varepsilon+v_k);\\ 0, & \text{其他.}\end{cases} \quad (15)$$

如果 $\sqrt{\beta_k}<\varepsilon-v_k$ 且 $\sqrt{\beta_k}\leqslant p(\varepsilon+v_k)$,则置 $\hat{\theta}_k=\hat{\theta}_{k-1}$,$P_k=P_{k-1}$,转步骤 3.

**步骤 1(d)**　如果 $p(\varepsilon-v_k)<\sqrt{\beta_k}<\varepsilon+v_k$,则由式(16)和(17)计算 $\varepsilon'$ 和 $v_k'$ 的值,并由式(18)求得 $q_k$,将 $\varepsilon'$ 和 $v_k'$ 的值分别赋给 $\varepsilon$ 和 $v_k$ 之后,转步骤 2.

$$\varepsilon'=\frac{-v_k+\varepsilon+\sqrt{\beta_k}}{2}, \quad (16)$$

$$v_k'=\frac{v_k-\varepsilon+\sqrt{\beta_k}}{2}, \quad (17)$$

$$q_k=\begin{cases}\dfrac{\varepsilon'(\sqrt{\beta_k}-p(\varepsilon-v_k))}{(p-1)\beta_k}, & \text{当}\sqrt{\beta_k}>p(\varepsilon-v_k);\\ 0, & \text{其他.}\end{cases} \quad (18)$$

如果 $\sqrt{\beta_k}<\varepsilon+v_k$ 且 $\sqrt{\beta_k}\leqslant p(\varepsilon-v_k)$，则置 $\hat{\theta}_k=\hat{\theta}_{k-1}$，$P_k=P_{k-1}$，转步骤 3.

**步骤 1(e)** 如果 $\beta_k\leqslant p(\varepsilon^2-v_k^2)$，则置 $\hat{\theta}_k=\hat{\theta}_{k-1}$，$P_k=P_{k-1}$，转步骤 3. 否则，解方程(19)，求其正实根 $q_k$，转步骤 2.

$$(p-1)\beta_k^2q_k^2+((2p-1)\varepsilon^2-\beta_k+v_k^2)\beta_kq_k+\varepsilon^2(p(\varepsilon^2-v_k^2)-\beta_k)=0. \tag{19}$$

**步骤 2** 将 $q_k$ 代入式(20)~(22)中，计算 $\hat{\theta}_k$ 和 $P_k$.

$$P_k=(1+q_k-\frac{q_kv_k^2}{\varepsilon^2+q_k\beta_k})P'_{k-1}, \tag{20}$$

$$\hat{\theta}_k=\hat{\theta}_{k-1}+\frac{q_kv_k}{\varepsilon^2+q_k\beta_k}P_{k-1}\phi_k, \tag{21}$$

式中

$$P'_{k-1}=P_{k-1}-\frac{q_k}{\varepsilon^2+q_k\beta_k}P_{k-1}\boldsymbol{\phi}_k\boldsymbol{\phi}_k^{\mathrm{T}}P_{k-1}. \tag{22}$$

**步骤 3** 如果 $k<N$，则 $k$ 增 1 并返回步骤 1.

由于有 $\boldsymbol{\Theta}_N\subseteq E_N$，所以得 $[\hat{y}_k^{\mathrm{m}},\hat{y}_k^{\mathrm{M}}]\subseteq[\bar{\hat{y}}_k^{\mathrm{m}},\bar{\hat{y}}_k^{\mathrm{M}}]$.式中

$$\bar{\hat{y}}_k^{\mathrm{m}}=\min_{\theta\in E_N}\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\theta},\bar{\hat{y}}_k^{\mathrm{M}}=\max_{\theta\in E_N}\boldsymbol{\phi}_k^{\mathrm{T}}\boldsymbol{\theta}. \tag{23}$$

可以推导出

$$\bar{\hat{y}}_k^{\mathrm{m}}=\boldsymbol{\phi}_k^{\mathrm{T}}\hat{\boldsymbol{\theta}}_N-\sqrt{\boldsymbol{\phi}_k^{\mathrm{T}}P_N\boldsymbol{\phi}_k}, \tag{24}$$

$$\bar{\hat{y}}_k^{\mathrm{M}}=\boldsymbol{\phi}_k^{\mathrm{T}}\hat{\boldsymbol{\theta}}_N+\sqrt{\boldsymbol{\phi}_k^{\mathrm{T}}P_N\boldsymbol{\phi}_k}. \tag{25}$$

替换式(9)，出水 BOD 的置信区间可以表示为

$$y_k\in[\bar{\hat{y}}_k^{\mathrm{m}}-\varepsilon,\bar{\hat{y}}_k^{\mathrm{M}}+\varepsilon],\ k\geqslant N+1. \tag{26}$$

### 2.2 置信区间的保守性降低

借鉴集员估计文献[21]的思想，设计多个 BOD 神经网络软测量模型(对于这些神经网络模型，隐含层节点个数、中心、宽度以及建模误差界设不同值)，并将多模型测量结果进行融合以降低单一模型所给结果的保守性.设软测量模型个数为 $M$，第 $i$ 个软测量模型的输出为置信区间 $I^{(i)}$.可以得到关于实际出水 BOD $y_k$ 的如下关系式

$$y_k\in I_k^{(i)},i=1,\cdots,M. \tag{27}$$

因而有 $y_k\in\cap_{i=1}^{M}I_k^{(i)}$，并且有

$$\forall i,\ \cap_{i=1}^{M}I_k^{(i)}\subseteq I_k^{(i)}. \tag{28}$$

即将各模型输出的交集作为最终结果，并且最终结果比任意单一模型的输出具有更低的保守性.该方法如图 2 所示.另外，有

$$\forall\Sigma\subseteq\{1,2,\cdots,M\},\cap_{i=1}^{M}I_k^{(i)}\subseteq\cap_{i\in\Sigma}I_k^{(i)}. \tag{29}$$

由上式可以看出：子模型个数越多，所得结果的保守性越低.因此，在时间允许的前提下，可建立更多的模型以进一步降低保守性.

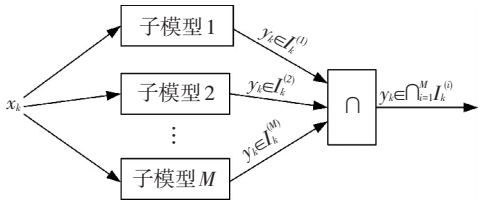



**图 2 置信区间的保守性降低方案**

Fig.2 Scheme for reducing the conservatism of the confidence interval

## 3 BOD 置信区间的使用

所计算出的置信区间描述了出水 BOD 真实值 $y_k$ 的存在范围，即 $y_k\in[y_k^{\mathrm{m}},y_k^{\mathrm{M}}]$.式中：$y_k^{\mathrm{M}}$ 和 $y_k^{\mathrm{m}}$ 分别表示置信区间的上下界.通过此置信区间可给出一个点估计[22]

$$y_k^{\mathrm{e}}=\arg\min_{y\in\mathbf{R}}\max_{y_k\in[y_k^{\mathrm{m}},y_k^{\mathrm{M}}]}|y_k-y|=(y_k^{\mathrm{M}}+y_k^{\mathrm{m}})/2. \tag{30}$$

可以作为 BOD 真实值 $y_k$ 的一个预测.而且，不难推导出 $|y_k-y_k^{\mathrm{e}}|\leqslant(y_k^{\mathrm{M}}-y_k^{\mathrm{m}})/2$.即 BOD 真实值 $y_k$ 与预测值 $y_k^{\mathrm{e}}$ 之间的偏差小于置信区间的半径.因此，置信区间的保守性越低，预测误差的最大值就越小.可以看出，通过出水 BOD 置信区间不仅可给出 BOD 真实值的一个预测，还可以评估预测误差.现有工作多是获得一个 BOD 预测值，而未曾给出评估各时刻预测误差的有效方法.关于水质参数 BOD 是否达标或满足运行调控要求，直接使用预测值而不参考预测误差(预测精度没有保证)可能会导致错误的分析和判断.

## 4 实 验

实验数据共 200 组，即 $L=200$，来源于北京市某污水处理厂.使用 PCA 方法选取辅助变量.在累计方差贡献率大于 90%的前提下，出水 BOD 软测量模型的辅助变量选取为 COD、SS、pH 和 DO.将降维后的数据分成两部分，一份用于神经网络模型建立，共 150 组，即 $N=150$；另一份用于模型预测，共 50 组.

实验建立 3 个软测量模型，然后将各模型输出的交集作为最终结果.在神经网络建模之前，先对数据进行归一化处理，即式(4)中的 $x_k$ 和 $y_k$ 都是归一化后的数据.试凑隐节点数使建模误差界尽可能小.对于模型 1，神经网络隐节点数 $p$、宽度 $\sigma$ 以及建模误差界 $\varepsilon$ 分别为 50、0.9 和 0.477 5；对于模型 2，3 个参数分别为 55、1 和 0.442 5；对于模型 3，3 个参数分别为 60、0.7 和 0.395 3.在权值估计过程中，集员辨识重复处理所有数据 200 次以进一步降低椭球描述的保守性.

图 3~5 分别为已建立好的模型 1~3 给出的出

水 BOD 置信区间.图 6 是模型 1、2 以及 3 所给结果的交集.图中实线为测量值 $y_k$,虚线和点线分别为反归一化后的区间上下界 $y_k^{\mathrm{M}}$ 和 $y_k^{\mathrm{m}}$.可以看出,出水 BOD 实测值完全落入 3 个模型所给出的置信区间中,以及其进一步完全落入 3 个模型所给结果的交集中,并且融合后的结果比单一结果保守性更低.不难看出,在时间允许的前提下,可建立更多的模型以进一步降低保守性.第 3 节已指出,置信区间的保守性越低,预测误差的最大值就越小,从而结果就越能反映真实值.

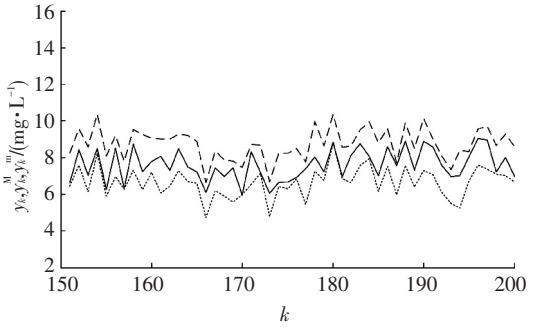



**图 3　模型 1 给出的出水 BOD 置信区间**

Fig.3　Confidence interval of the effluent BOD given by Model 1

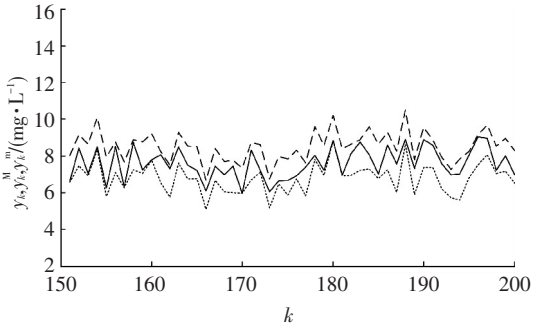



**图 4　模型 2 给出的出水 BOD 置信区间**

Fig.4　Confidence interval of the effluent BOD given by Model 2

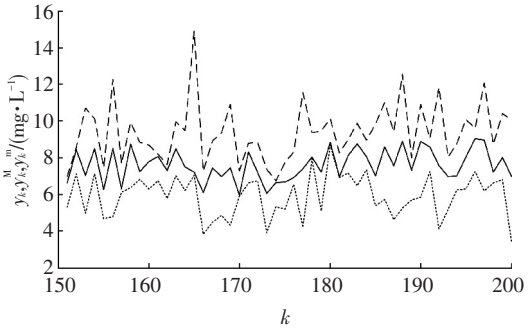



**图 5　模型 3 给出的出水 BOD 置信区间**

Fig.5　Confidence interval of the effluent BOD given by Model 3

此外,本实验还与传统的神经网络软测量建模方法进行对比.该方法采用最小二乘法确定 RBF 神经网络输出权值,并且试凑中心和宽度使建模误差尽可能小.神经网络隐节点数 $p$ 和宽度 $\sigma$ 分别为 60 和 0.7.针对归一化后的数据,最终 RSME 为 0.227 5.置信区间计算公式见文献[16].图 7 给出该方法的结果.图中实线为测量值 $y_k$,点划线为单点估计 $y_k^{\mathrm{e}}$,虚线和点线分别为反归一化后的区间上下界 $y_k^{\mathrm{M}}$ 和 $y_k^{\mathrm{m}}$.可以看出,点估计不能保证预测效果始终可靠,尽管建模误差已被降低,比如在时刻 165 出现较大预测误差.并且,该方法基于随机误差假设计算置信区间,而这里误差的统计特性难以获得,因而所计算出的置信区间并不理想,其保守性大于本文方法所得结果.

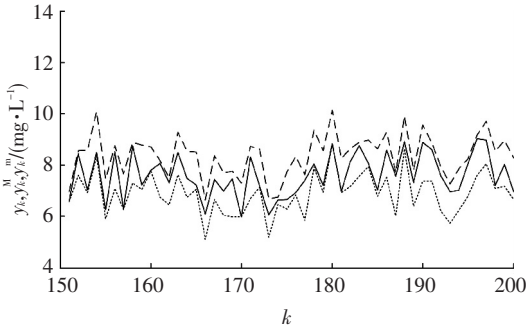



**图 6　3 个模型给出结果的交集**

Fig.6　Intersection of the results given by the three models

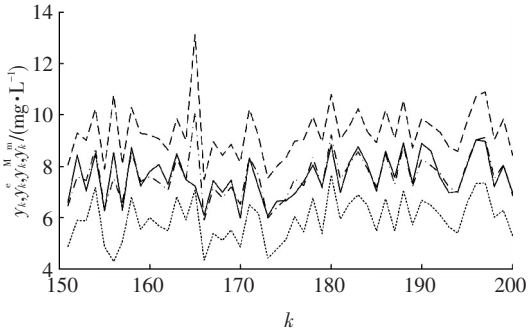



**图 7　采用最小二乘法所建模型给出的结果**

Fig.7　Results given by the model built by the least squares

## 5　结　语

污水处理过程要求实时监测出水水质指标,以防止超标造成环境污染.本文给出一种新的软测量方法,可以计算出水 BOD 的置信区间,从而实现水质指标的可靠监测.该方法通过 PCA 选取辅助变量;采用 RBF 神经网络建立软测量模型;使用参数线性集员辨识算法给出网络输出权值的不确定集合描述.此外,本文借鉴集员估计文献[21]的思想,设计多个 BOD 神经网络软测量模型,并将多模型测量结果进行融合以降低单一模型所给结果的保守性.

## 参考文献

[1] HAIMI H, MULAS M, CORONA F, et al. Data-derived soft-sensors for biological wastewater treatment plants: An overview [J]. Environmental Modelling & Software, 2013, 47: 88-107. DOI:10.1016/j.envsoft.2013.05.009.

[2] 祁国强, 刘载文, 崔莉凤. 基于 RBF 人工神经网络的生活污水处理软测量方法[J]. 北京工商大学学报(自然科学版), 2004,

22(3): 36-38. DOI: 10.3969/j.issn. 1671-1513. 2004. 03. 010.
QI Guoqiang, LIU Zaiwen, CUI Lifeng. The soft measuring method based on RBF neural network for municipal wastewater treatment [J]. Journal of Beijing Technology and Business University (Natural Science Edition), 2004, 22(3): 36-38. DOI: 10.3969/j.issn. 1671-1513. 2004. 03. 010.

[3] HUANG Zhiqing, LUO Jianxu, LI Xiongjun, et al. Prediction of effluent parameters of wastewater treatment plant based on improved least square support vector machine with PSO[C]//Proceedings of the 1st International Conference on Information Science and Engineering. Piscataway, NJ: IEEE Press, 2009: 4058 - 4061. DOI: 10.1109/ICISE.2009.846.

[4] 王树东, 葛珉昊, 陈明明. 基于混合递阶遗传算法优化 RBF 神经网络的 BOD5 软测量方法[J]. 给水排水, 2014, 40(3): 149-153. DOI: 10.3969/j.issn. 1002-8471. 2014. 03. 036.
WANG Shudong, GE Minhao, CHEN Mingming. Soft measuring method for BOD5 based on RBF neural network optimized by hybrid hierarchy genetic algorithm[J]. Water Supply and Drainage, 2014, 40(3): 149-153. DOI: 10.3969/j.issn. 1002-8471. 2014. 03. 036.

[5] 许玉格, 刘莉, 曹涛. 基于 Fast-RVM 的在线软测量预测模型[J]. 化工学报, 2015, 66(11): 4540-4545. DOI: 10.11949/j.issn.0438-1157.20150566.
XU Yuge, LIU Li, CAO Tao. On-line soft measuring model based on Fast-RVM[J]. CIESC Jorunal, 2015, 66(11): 4540-4545. DOI: 10.11949/j.issn.0438-1157. 20150566.

[6] QIAO Junfei, LI Wei, HAN Honggui. Soft computing of biochemical oxygen demand using an improved T-S fuzzy neural network[J]. Chinese Journal of Chemical Engineering, 2014, 22(11/12): 1254-1259. DOI: 10.1016/j. cjche. 2014.09.023.

[7] LIU Yiqi, HUANG Daoping, LI Yan. Development of interval soft sensors using enhanced just-in-time learning and inductive confidence predictor[J]. Industrial and Engineering Chemistry, 2012, 51(8): 3356-3367. DOI: 10.1021/ie201053j.

[8] ŠKRJANC I. Confidence interval of fuzzy models: An example using a waste-water treatment plant[J]. Chemometrics and Intelligent Laboratory Systems, 2009, 96(2): 182-187. DOI: 10.1016/j. chemolab.2009.01.009.

[9] LJUNG L. System identification-theory for the user[M]. 2nd ed. Englewood-Cliffs, NJ: Prentice-Hall, 1999.

[10] NINNESS B, GOODWIN G. Estimation of model quality[J]. Automatica, 1995, 31(12): 1771-1797. DOI: 10.1016/0005-1098(95)00108-7.

[11] MRUGALSKI M, WITCZAK M, KORBICZ J. Confidence estimation of the multi-layer perceptron and its application in fault detection systems[J]. Engineering Applications of Artificial Intelligence, 2008, 21(6): 895 - 906. DOI: 10.1016/j.engappai. 2007.09.008.

[12] FÁBREGA A J S, CARO J M B, HERRERA P J A, et al. Data-driven bounded-error fault detection[J]. International Journal of Adaptive Control and Signal Processing, 2014, 28(12): 1299-1324. DOI: 10.1002/acs.2443.

[13] FÁBREGA A J S, CARO J M B, HERRERA P J A, et al. Fault detection method based on bounded error and dynamic threshold techniques[J]. International Journal of Adaptive Control and Signal Processing, 2016, 30(2): 256-270. DOI: 10.1002/acs.2558.

[14] 柴伟, 孙先仿. 一种非线性系统集员辨识算法[J]. 北京航空航天大学学报, 2005, 31(11): 1237-1241. DOI: 10.3969/j.issn. 1001-5965.2005.11.018.
CHAI Wei, SUN Xianfang. Set membership identification algorithm of nonlinear systems[J]. Journal of Beijing University of Aeronautics and Astronautics, 2005, 31(11): 1237-1241. DOI: 10.3969/j.issn.1001-5965.2005.11.018.

[15] CHAI Wei, QIAO Junfei. Non-linear system identification and fault detection method using RBF neural networks with set membership estimation[J]. International Journal of Modelling, Identification and Control, 2013, 20(2): 114-120. DOI: 10.1504/IJMIC. 2013. 056183.

[16] CHAI Wei, QIAO Junfei. Passive robust fault detection using RBF neural modeling based on set membership identification[J]. Engineering Applications of Artificial Intelligence, 2014, 28: 1-12. DOI: 10.1016/j. engappai. 2013.10.005.

[17] 杨斌, 田永青, 朱仲英. 智能建模方法中的数据预处理[J]. 信息与控制, 2002, 31(4): 380-384. DOI: 10.3969/j.issn.1002-0411.2002.04.019.
YANG Bin, TIAN Yongqing, ZHU Zhongying. Data pretreatment in intelligent modeling methods[J]. Information and Control, 2002, 31(4): 380-384. DOI: 10.3969/j.issn.1002-0411.2002.04.019.

[18] BEZDEK J C. Pattern recognition with fuzzy objective function algorithms[M]. New York: Plenum Press, 1981.

[19] FOGEL E, HUANG Y F. On the value of information in system identification-bounded noise case[J]. Automatica, 1982, 18(2): 229-238. DOI: 10.1016/0005-1098(82)90110-8.

[20] BELFORTE G, BONA B. An improved parameter identification algorithm for signals with unknown but bounded errors[C]// Proceedings of the 7th IFAC/IFORS Symposium on Identification and System Parameter Estimation. Oxford: Pergamon Press, 1985: 1507-1511.

[21] MOISAN M, BERNARD O, GOUZE J L. Near optimal interval observers bundle for uncertain bioreactors[J]. Automatica, 2009, 45(1): 291-295. DOI: 10.1016/j. automatica.2008.07.006.

[22] MILANESE M, VICINO A. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview[J]. Automatica, 1991, 27(6): 997-1009. DOI: 10.1016/0005-1098(91)90134-N.

(编辑 刘 彤)