

# 一种自确认软测量方法的研究与应用\*

肖红军<sup>1</sup>, 刘乙奇<sup>2</sup>, 伍俊<sup>1</sup>

(1. 佛山科学技术学院 机械与电气工程学院, 广东 佛山 528000;

2. 华南理工大学 自动化科学与工程学院, 广东 广州 510640)

**摘要:** 软测量技术在过程工业中得到了广泛应用于代替或部分代替仪表的功能进行监控。但是, 长期困扰软测量应用的是其鲁棒性问题。因此, 提出了新型的软测量模型即自确认软测量模型。利用主元分析对输入传感器数据进行了在线校验和故障重构, 不仅能确保了输入传感器 (辅助变量传感器) 数据的可靠性。同时, 利用一定的状态参数来指示当前输入传感器的状态。此外, 通过输出方差和归纳区间估计两种方法对软测量输出的不确定性进行了描述。模型的输出将不是单个预测值的输出, 而是同时输出五种信息: 带概率区间的输出、模型的状态 (故障状态, 重构状态和迷失状态等等)、不确定性、故障信息和校验输出值, 并对发生故障的输入传感器进行数据重构和修复以达到模型自校验和自诊断的目的。所提出方法在污水处理过程中得到了有效应用。

**关键词:** 自确认; 软测量; 主元分析; 递归偏最小二乘

**中图分类号:** TP302.1   **文献标志码:** A   **文章编号:** 0529-6579 (2014) 04-0045-07

## Research and Application on a Novel Self-validating Soft Sensor

XIAO Hongjun<sup>1</sup>, LIU Yiqi<sup>2</sup>, WU Jun<sup>1</sup>

(1. Mechanical and electric engineering college, Foshan University, Foshan 528000, China;

2. School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China)

**Abstract:** A soft sensor is widely used as a alternative for process monitoring, but robust problems prevailing in soft sensor application hinder its further usage. A novel soft sensor such as self-validating soft sensor is proposed, which not only deals with input data online calibration and fault reconstruction problems, but also further ensures the reliability of input data by performing Principal Component Analysis (PCA) analysis. In addition, uncertainties are characterized by output variance and inductive confidence predictor (ICP). The output with uncertainties, model statuses uncertainties, faulty messages and validated output are implemented for ensuring proper fault reconstruction and model calibration. The effectiveness was proved in a wastewater treatment process.

**Key words:** self-validating; soft sensor; principal component analysis (PCA); recursive partial least squares (RPLS)

在过程控制中, 常规仪表常常会遇到购买及维护费用过高, 延迟过大甚至工作条件过于恶劣等问题, 因此, 软测量技术广泛应用于代替或部分代替仪表的功能以达到系统进行监控的目的<sup>[1-3]</sup>。软测

量是依据可测或易测变量 (辅助变量) 建立模型对难以直接测量变量 (主导变量) 进行测量和估计的方法<sup>[4]</sup>。但是, 软测量是通过输入传感器的输入变量和模型进行预测的, 如果某一个传感器发

\* 收稿日期: 2014-01-20

基金项目: 广东省科技计划资助项目 (2011B010100015); 佛山市科技发展专项基金资助项目 (2012AA100651)

作者简介: 肖红军 (1979年生), 男; 研究方向: 自动检测装置、电力电子; 通讯作者: 刘乙奇; E-mail: aulyq@scut.edu.cn

生故障, 将直接导致软测量预测的偏离甚至失效。因此, 在软测量进行模型预测之前, 有必要对其输入传感器进行有效的校验。尽管大量学者对传感器的故障诊断和辨识进行了大量研究<sup>[5-6]</sup>, 但是, 鲜有学者研究专门针对软测量的输入传感器的故障诊断和辨识, 故障重构更是少之又少。Martin 和罗中良等提出了经网络对输入传感器进行校验<sup>[7-8]</sup>, 在他们的方法中, 如果输入测量数据超过某一限值, 操作员则利用经验知识手动的替代故障的数据。可是, 这在过程工业中显得不切合实际。通过最小化 SPE 故障检测的原理, Qin 等提出了基于 SPE 和  $T^2$  的故障辨识和重构方法<sup>[9]</sup>。但是, 单纯的 SPE 或  $T^2$  故障检测方法容易产生误报警。Yue 和 Qin 等<sup>[6]</sup>混合了 SPE 和  $T^2$  指标进行故障检测, 并获得了不错的效果。但由于重构算法比较复杂和耗时, 在实际中不易实现, 同时混合指标的限值不容易确定。

SEVA (自确认) 仪表是自带自检测参数的一种新型仪表, 由牛津大学和 Foxboro 提出的标准, 并在热电偶和溶解氧传感器中得到了应用<sup>[10]</sup>。本文在此基础上研究预测模型的综合输出机制, 模型的输出将不是单个预测值的输出, 而是同时输出五种信息: 带概率区间的输出、模型的状态 (故障状态, 重构状态和迷失状态等等)、不确定性、故障信息和校验输出值, 并对发生故障的输入传感器进行数据重构和修复以达到模型自校验和自诊断的目的。特别需要指出的是本文创新性的运用鲁棒统计学中的 CP (Confidence Predictor) 技术对预测模型的不确定进行定性的描述。此外, 可以将输入传感器的状态也作为模型的输出, 从而达到对预测模型从输入、模型和输出的三方位检测, 使得模型的输出不再只是单纯的预测值, 还有不确定区间和输入传感器的输入状态。

## 1 统计学分析方法

### 1.1 主元分析 (PCA)

PCA 根据保留最大数据方差的条件下对输入变量进行降维约简<sup>[11]</sup>。假设  $X$  是  $n \times m$  的数据矩阵, PCA 将多变量的输入数据矩阵进行奇异值分解后得到主元和特征向量:

$$X = TP^T = t_1 p_1^T + t_2 p_2^T + \cdots + t_m p_m^T + \varepsilon \quad (1)$$

其中  $t_i \in R^n$  被称为得分向量, 即矩阵  $X$  的主元,  $p_i \in R^m$  被称为负荷向量, 是残差矩阵, 它包含了系统的数据的主要噪声。 $T = [t_1, t_2, \cdots, t_n]$  和  $P = [p_1, p_2, \cdots, p_m]$  分别是得分向量矩阵和负荷向量矩阵。给定一个新的采样向量, PCA 的得分向量,

预测和残差向量的关系如下:

$$t_i = x p_i \quad (2)$$

$$\hat{x} = x C \quad (3)$$

$$\hat{x} = x(I - C) \quad (4)$$

其中  $C = P^T P$ ,  $\hat{x}$  是预测值,  $\hat{x}$  是残差。采样向量被映射到了主元空间和残差空间, 即,

$$x = \hat{x} + \hat{x} \quad (5)$$

统计量 SPE 主要用于测量残差数据的变化

$$SPE(x) = \|\hat{x}\|^2 = x(I - C)x^T \quad (6)$$

正常情况下, SPE 是可以忽略不计的。但是, 如果某一个传感器发生故障 SPE 将急剧变大, 可用于传感器故障的诊断。但是 SPE 对噪声过度敏感从而极易导致误报警。基于上述分析, 本文将通过最小化 SPE 改进传感器故障诊断指标, 并提出 SVI 指标。在软测量预测之前, SVI 可以用来确认软测量的输入数据的正常与否。与此同时, 故障的传感器数据能被 PCA 模型的估计值进行重构。另外, PCA 还能用于软测量建模过程中 JIT 算法的数据选择。

### 1.2 递归偏最小二乘法 (RPLS)

PLS 方法将高维数据空间投影到低维特征空间, 得到相互正交的特征向量, 再建立特征向量间的一元线性回归关系<sup>[12]</sup>。假定一个输入输出组成的矩阵数组  $\{X Y\}$ , 输入输出矩阵可表达为

$$X = TP^T + E \quad (7)$$

$$Y = UQ^T + F \quad (8)$$

其中  $T \in R^{n \times d}$  和  $U \in R^{n \times d}$  分别是  $X$  和  $Y$  的得分向量矩阵,  $d$  是负载向量的个数。而  $P \in R^{m \times d}$  和  $Q \in R^{m \times d}$  分别是相关的负载向量矩阵,  $E$  和  $F$  分别是  $X$  和  $Y$  的残差。 $T$  和  $U$  的内部关系则由公式 (9) 表示

$$U = TB + R \quad (9)$$

其中  $B \in R^{d \times d}$  是通过最小化残差  $R$  其中包含回归权值的对角矩阵。 $\{T U P Q B\}$  是  $\{X Y\}$  经过偏最小二乘法处理后获得的参数, 即

$$\{X Y\} \xrightarrow{PLS} \{T U P Q B\} \quad (10)$$

而回归系数由式 (11) 表示:

$$C^{PLS} = (X^T X)^+ + X^T Y = W^* B Q^T \quad (11)$$

其中

$$W^* = [w_1^* w_2^* \cdots w_m^*] \quad (12)$$

$$w_i^* = \prod_{h=1}^{i-1} (I_m - w_h p_h^T) w_i \quad (13)$$

$0^+$  是右逆,  $I_m$  表示  $m$  行的单位阵,  $w_i$  是 PLS 的权向量,  $Q$  是 PLS 的负载向量,  $p$  是  $Q$  的列向量,  $h$  表示 PLS 提取出来的主元个数, 而  $C^{PLS}$  是 PLS 算法的回归系数。当新数据到来时, 偏最小二乘模型据

式 (14) 进行更新。

$$X_{NEW} = [P^T; x_{new}^T], Y_{NEW} = [Q^T; y_{new}^T] \quad (14)$$

其中,  $x_{new}$  是新的数据输入点,  $y_{new}$  是新的数据输出点。 $X_{NEW}$  和  $Y_{NEW}$  则分别表示输入  $X$  的负载向量矩阵和输出  $Y$  的负载向量矩阵在新数据更新后的新矩阵。由上式可以看出, 随着新数据不断的到来, 矩阵的维数并没有发生变化, 也因此大大加快了数据的处理速度和加快模型的处理速度。

## 2 自确认软测量方法

### 2.1 基于 PCA 的辅助变量传感器故障诊断与数据重构

不同于传统的基于 PCA 模型的 SPE 或  $T^2$  的故障诊断方法, 本文结合了 SPE 因子和 SVI 因子, 并将它们用于故障的诊断。此外, 利用方差分析方法 URV (unreconstructed variance) 在 PCA 模型建立之初对数据进行了主元个数进行合理的分析。具体如图 1 所示。

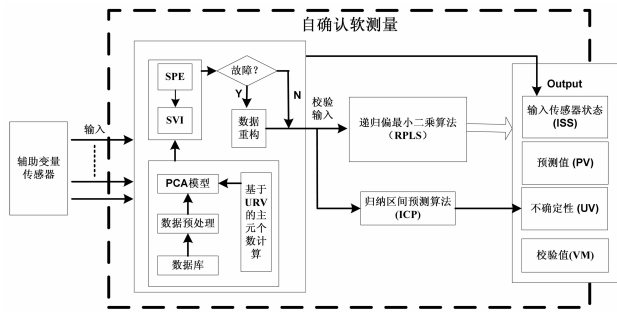


图 1 自确认软测量框架

Fig. 1 Schematic of self-validating a soft sensor

2.1.1 故障数据重构 假设软测量某一个传感器发生故障, 为了使得故障的软测量输入数据能恢复到正常状态, 本小节充分利用其它正常传感器的相关性对故障传感器进行数据重构。在每次重构之前或之后, 都需通过计算残差的 SVI 对重构的数据进行评估。如果故障传感器被确认, 残差会发生剧烈的衰减。因为, 传感器的故障诊断和数据重构是相互耦合且不可分的。根据式 (4), 能估算出第  $i$  个变量, 其中  $\hat{x}_i$  是  $x_i$  的重构数据。但是这种方法的缺陷在于故障传感器的数据也包含在了  $x$  中。为了消除故障传感器的影响, 第  $i$  个变量的预测值将被反馈重新输入到 PCA 的数据估计中, 直到收敛到一个稳定值  $z_i$ , 从而这个值能用于故障传感器的估计, 迭代方程的表达如下:

$$z_i^{new} = c_{ii} z_i^{old} + [x_{-i}^T \ 0 \ x_{+i}^T] = [x_{-i}^T z_i^{old} \ x_{+i}^T] \quad (15)$$

其中  $C = P^T P = [c_1 \cdots c_m], c_i^T = [c_{li} \cdots c_{mi}], x^T$  代表了  $X$  矩阵的行数, 下标  $-i, +i$  代表正交向量的前  $i-1$  个值和后  $m-i$  个值。Dunia 等<sup>[13]</sup> 提出迭代总是能收敛到一个稳定的值, 同时这个值能由式 (16) 计算而无需进行多次迭代。

$$z_i = \frac{[x_{-i}^T \ 0 \ x_{+i}^T] c_i}{1 - c_{ii}} \quad (16)$$

实际上, PCA 模型主元个数的确定是相对主观的。如果故障传感器输入数据通过大量主元重构所得, 则重构输入数据极大可能会包含大量主元所带来的噪声。相反, 如果所确定的主元过少, 所建立模型可能会由于故障数据得不到充分的重构而产生病态。此外, 在 PCA 用于数据重构的过程中, 主元的个数会进一步影响传感器的校验, 如检测小故障的能力, 故障辨识的自由度和重构的精确度。

Dunia 等<sup>[13]</sup> 发现当 PCA 模型用于重构故障传感器时, 方差分析方法 URV 是可以在分析方差的同时, 计算最优的主元个数, 并且对重构效果有一定的指示作用。例如, 变量  $i$  的重构方差  $u_i$  表示, 即  $x_i - z_i$  的方差计算如下:

$$u_i = \text{var}(x_i - z_i) = \text{var}(\varepsilon_i^T x - \frac{[x_{-i}^T \ 0 \ x_{+i}^T] c_i}{1 - c_{ii}}) \approx \frac{1}{(1 - c_{ii})^2} \varepsilon_i^T (I - C) R (I - C) \varepsilon_i = \frac{\varepsilon_i^T R \varepsilon_i}{(\varepsilon_i^T V_p V_p^T \varepsilon_i)^2} \quad (17)$$

其中  $R = E(xx^T)$  正常数据的协方差矩阵,  $\varepsilon_i$  是故障传感器的单位向量。因为每个变量  $x_i$  被标准化为方差为 1 的变量, 即  $u_i = 1$ 。如果令  $z_i = \bar{x}_i$ , 则  $u_i > 1$  说明重构效果比较差。

2.1.2 故障检测和辨识 因为多传感器同时发生故障的可能性比较小, 因此, 此方法很容易能推广到多个不同时发生故障传感器的故障诊断。当某个传感器发生故障的时候, 采样向量  $x$  可以表示为

$$x = x^* + f \varepsilon_i \quad (18)$$

其中  $x^*$  代表正常数据的部分,  $\varepsilon_i$  是故障传感器的单位向量, 而  $f$  是故障的大小, 可以为正值也可以为负值。为了说明方法的效果, 重构数据表示为:

$$x_j^T = [x_{-j}^T \ z_j \ x_{+j}^T] \quad (19)$$

被重构的向量能被映射到模型和残差空间, 从而计算出 SPE 如下:

$$\text{SPE}(x_j) = \|\hat{x}\|^2 = x_j^T (I - C) x_j \quad (20)$$

如果第  $i$  个传感器发生故障同时  $j \neq i$ ,  $\text{SPE}(x_j)$  不会发生明显变化。然而, 如果选到真正的故障传感器, 即  $i = j$ ,  $\text{SPE}(x_j)$  则会明显减小。同时,

故障 SPE ( $x$ ) 会急剧的变大, 因此, 本文设计的 SVI 因子如下:

$$SVI = \eta_j = \frac{SPE(x_j)}{SPE(x)} \quad (21)$$

其中  $1 \leq \eta \leq 0$ 。从以上式子中可以看出, 当发生故障时, SPE ( $x_j$ ) 急剧减小, 而 SPE ( $x$ ) 会急剧的变大, 因此 SVI 会近乎接近于 0, 即  $\eta_j$  近乎于 0, 同时比单纯的 SPE ( $x$ ) 更加敏感。Dunia 等<sup>[13]</sup>在他的文中给出了具体的讨论。此外, 基于输入传感器彼此间的相关性, 其它正常的相关传感器将被用来重构故障变量, 从而使得软测量的预测更加稳定。

## 2.2 自确认软测量的多维度输出

从传统角度来说, 软测量一般仅仅是作为未知信号的发生器。区别于传统的软测量, 本文的自确认软测量不仅如上述描述的能输入传感器数据进行自校验, 同时能产生多种信号: PV, ISS, OSS, VM 和 VU。

**2.2.1 预测值 (PV)** 预测值与传统传感器的输出并无差异, 即给定易测量传感器的输入在模型的基础上预测输出值。

**2.2.2 输入传感器的状态 (ISS)** 输入传感器的状态表示了输入传感器在系统的运行当中所呈现的故障或正常运行状态。状态类别如表 1 所示。ISS 能为现场操作人员提供现场仪表运行的直接决策依据。

表 1 输入传感器的状态值

Table 1 Input sensor status values

状态	说明
CLEAR	输入传感器正常
BLURRED	故障过渡状态
DAZZLED	输入传感器被历史数据修复
BLIND	没有可替代的数据
OFFLINE	仪表处于离线状态

**2.2.3 输出传感器状态 (OSS)** 由于污水厂现场条件的恶劣, 在线仪表很容易偏离正常状态。另外, 由于软测量模型独立于在线仪表, 因此, 本文提出了 PV 值, 其可对在线分析仪表进行在线校验。和表 1 类似, 共有 5 种状态对在线分析仪表进行故障显示。

**2.2.4 不确定性校验** 不确定, 即 UV 值, 能为真实的预测值提供可靠的置信空间, 可以用于衡量模型和过程的噪声以及故障的影响。为了便于置信区间的描述, 同时也是基于 ICP (Inductive Confi-

dence Predictor) 算法的高效性, 本文中利用 ICP 算法对软测量模型的不确定性进行了阐述。ICP 算法, 是主动的 (Inductive) 不确定性描述算法, 较之于被动算法 (Transductive) 有更高的计算效率, 具体可参见文献 [14]。对于 ICP 算法, 给定  $l$  个采样点的集合  $\{(x_1, y_1), \dots, (x_l, y_l), (x_l + g, y_l + g)\}$ , 其中  $i = 1, \dots, l$ , 和一个新的输入点  $x_{l+g} \in \mathcal{R}^m$ 。ICP 假定输入点  $x_{l+g}$  的可能输出值是  $y_{l+g}$ , 检验其拓展样本集合是独立同分布的, 集合如下:

$$\{(x_1, y_1), \dots, (x_l, y_l), (x_{l+g}, y_{l+g})\} \quad (22)$$

即是检验  $y_{l+g}$  是  $x_{l+g}$  的真实输出值的概率有多高, 其中  $y_{l+g}$  在是式 (22) 的位置值。为式 (22) 每个数据对  $(x_i, y_i)$  分配一个  $\alpha_i^{y_{l+g}}$  值用来指示其在此数据集的一致性。更具体的来说, 其用来指示真实值  $y_{l+g}$  和预测值  $\hat{y}_{l+g}$  的一致性。为了有效的计算  $\alpha_i^{y_{l+g}}$ , ICP 算法首先将集合 ( $l$  对个数据) 分成训练集合 ( $m < l$  个数据对) 和校验集合 ( $q := l - m$  个数据对)。接着, 利用训练集合来训练回归模型 (即软测量模型) 和利用校验集合来计算每个输出值  $y_{l+g}$  的  $p$  值。因此, 每次只需计算新的数据点  $x_{l+g}$  的可能分配到  $y_{l+g}$  的  $\alpha_i^{y_{l+g}}$  值, 将作为  $p$  值处理。

$$p(y_{l+g}) = p((x_1, y_1), \dots, (x_l, y_l), (x_{l+g}, y_{l+g})) = \frac{\#\{i = m + 1, \dots, l, l + g : \alpha_i \geq \alpha_{l+g}^{y_{l+g}}\}}{q + 1} \quad (23)$$

其中  $\#A$  代表了集合  $A$  中元素个数。假定可以计算校验集合中每个可能输出的  $p$  值, 所有输出的  $p$  值在置信度以内则其至少有概率发生故障。因此, 给定置信度  $1 - \delta$ , ICP 算法的输出为:

$$\{y_{l+g} : p(y_{l+g}) > \delta\} \quad (24)$$

公式在文献 [13] 中得到了有效的证明。

为了使 ICP 算法和 RPLS 算法能够有效的结合, 一致性测量指标是必不可少的。因为一致性测量指标是用来衡量真实输出值和预测值相对于采样点的重要指标, 因此, 其被定义为真实值和预测值的绝对误差:

$$\alpha_i = |y_i - \hat{y}_i| \quad (25)$$

ICP 算法所产生的置信区间可为 RPLS 算法提供有效的置信空间。如果在线分析仪输出超出置信区间, 现场操作人员需要多次确认从而确保仪表的可靠性。因此, RPLS 软测量模型输出提供的不再是单调的预测值输出, 而是包含了输出的校验的输出值。在软测量模型建模的过程中, 数据选择算法选出来的数据被放入数据空间  $S$  中, 无论是 RPLS 算法或 ICP 算法每次建模都是以这个数据空间的数

据为基础的。具体 RPLS 算法和 ICP 算法的计算如表 2 所示。

表 2 RPLS 算法和 ICP 算法的结合  
Table 2 RPLS algorithm coupled with ICP Algorithm

RPLS 算法和 ICP 算法的结合

- 1) 将训练数据存于 S 空间中；
- 2) 将训练集合  $\{(x_1, y_1), \dots, (x_l, y_l)\}$  分为两个集合：适配集合  $\{(x_1, y_1), \dots, (x_m, y_m)\}$  和校验集合： $\{(x_{m+1}, y_{m+1}), \dots, (x_l, y_l)\}$ ；
- 3) 利用适配集合训练 RPLS 模型；
- 4) 对于在校验集合中的每个数据对  $(x_{m+i}, y_{m+i})$ ,  $i = 1, \dots, q$ : 根据输入模式  $x_{m+i}$  训练 RPLS 模型，同时利用式 (25) 计算  $\alpha_{m+i}$ ；
- 5) 以降序的方式排列计算所得  $\alpha_{m+1}, \dots, \alpha_{m+q}$ ；
- 6) 当数据新数据  $x_{new}$  更新后，利用式 (24) 计算新数据的置信区间。

2.2.5 校验测量值 鉴于现场复杂和恶劣的条件，在线分析仪表有极大的可能发生极端的或故障的情况。同时，即使仪表工作正常，仪表定期检测也是不可避免的。因此，有必要为仪表提供定时在线测量。在本文的软测量的建模过程中，在线分析仪表的输出和模型的更新是独立的。同时，由于软测量模型输出带有不确定上下限检验，因此，当其在线分析仪表发生故障时，能为其提供可替代的信息。

### 3 仿真实验

污水处理是一个复杂的生化处理过程，特别是污水厂的输入流量随着晴天雨天发生剧烈的变化，同时，其生物种群更是实时的发生变化，且工艺知识也不尽完善，加之现场条件恶劣，仪表发生故障的概率也随之增加。BOD<sub>5</sub> 是测量有机物重要的出水指标，也是重要的检测参数。目前 BOD<sub>5</sub> 的测量无法进行实时在线检测，而市场上现有的实时在线 5 d BOD 测量仪表造价很高，实用性和通用性差。因此，实现 BOD<sub>5</sub> 的实时在线测量显得至关重要。校验数据来源于巴塞罗那的一个污水厂，将近 2 年 400 组数据。每组数据有 38 个变量。本文中利用 Kohomen 的自组织神经网络 (SOM) 检测和选择相关向量<sup>[15]</sup>，因此，经过选择辅助变量如表 3 所示。其中 200 组数据用来训练模型，另外 200 组数据用来校验模型，而输出是 BOD<sub>5</sub>，输入变量是 19 个辅助变量。

表 3 辅助变量表  
Table 3 Auxiliary variables

位置	变量
整个污水厂	可降解固体浓度 S <sub>total</sub> [RD-SED-G]，悬浮固体浓度 SS <sub>total</sub> [RD-SS-G]，生物需氧量 BOD <sub>total</sub> [RD-DBO-G]，化学需氧量 COD <sub>total</sub> [RD-DQO-G]
初沉池	生物需氧量 BOD <sub>pr</sub> [RD-DBO-P]，悬浮固体浓度 SS <sub>pr</sub> [RD-SS-P]
二沉池	生物需氧量 BOD <sub>sec</sub> [RD-DBO-S]，化学需氧量 COD <sub>sec</sub> [RD-DQO-S]
入水	生物需氧量 BOD <sub>in</sub> [DBO] and 化学需氧量 COD <sub>in</sub> [DQO]
二级处理	COD <sub>secpro</sub> [DQO]，生物需氧量 BOD <sub>tsecpro</sub> [DBO]，悬浮固体浓度 SS <sub>secpro</sub> [SS]，PH 和可降解固体浓度 S <sub>secpro</sub> [SED]
出水	化学需氧量 COD <sub>out</sub> [DQO-S]，可降解固体浓度 S <sub>out</sub> [SED-S]，悬浮固体浓度 SS <sub>out</sub> [SS-S] 和 PH-S

#### 3.1 自确认软测量的输入传感器的校验

为了得到最佳的主元个数，通过相关性进行彼此检测 URV 的计算，URV 达到了最小值，PC 为 9 时被用来作为 PCA 模型的主元个数。传感器仪表都处于正常状况下，200 组数据被用来训练 PCA 模型，而其他 200 组数据用来校验 PCA 模型，从而得到了 PCA 模型和正常状态的阈值。为了确认 PCA 模型对软测量的校验作用，我们假定二沉池的 pH 值发生阶跃性的传感器故障，如图 2 (a) 所示从第 120 天到 170 天有 8 左右的偏差。而从图 2 (b) 中可以看出 SPE 和 SVI 能很轻易的检测出传感器的故障，SPE 值明显超出了控制限，经过数据重构以后，SPE 能很快的回复正常，但是仍然发生误报警现象，如图 2 (d) 而 SVI 比较稳定，很快定位出了是 pH 传感器发生故障，其 SVI 值几乎接近于 0 (其他传感器的 SVI 值则都近乎于 1，图中没显示)，并且没有误报警现象。同时，传感器在数据重构以后，很快的转入了正常状态。另外，图 2 (a) 显示了重构值和实际值能够很好的吻合，说明了重构的精确性。同时，SEVA 软测量不仅能辨识和重构故障，而且能进一步显示输入传感器的状态，正如图 2 (c) 所示，如果二沉池的 pH 值发

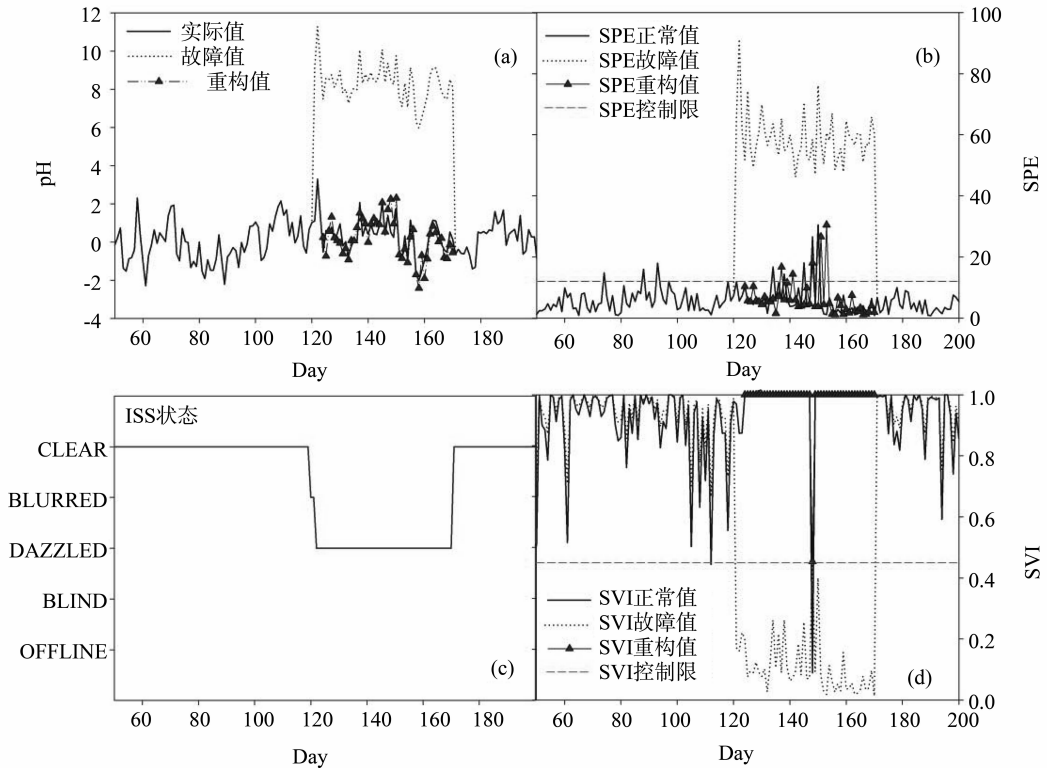


图 2 故障, 正常和重构的 pH

Fig. 2 Fault, normal and reconstructed pH

生 8 左右的偏差从第 120 天到 170 天, 由于故障的发生, ISS CLEAR 状态转变到了 BLURRED 状态, 同时没有显示是不是永久性故障。但是由于 PCA 模型对故障的检测和重构, pH 传感器跳到了 DAZZLED 状态, 并保持这个状态直到仪表恢复, 并使得 ISS 在第 170 后跳到了 CLEAR 状态。

### 3.2 自确认软测量的预测模型的校验和不确定值 (UV) 的输出

为了显示预测模型的预测误差, 文中引入的均方差 (RMSE) 和相关系数是准确的评估软测量模型的重要指标。因为污水厂的复杂性, 不仅受到季节变化、晴雨变化以及污水中微生物变化的影响, 单纯的预测值输出的软测量模型不能很好的确定其模型以及外部的不确定性, 因此, 我们利用 ICP 算法来计量预测输出值的置信区间, 从而来描述模型的不确定性。从图 3 显示, 由于在过渡过程中过程变量偏离原本的稳定状态, UV 会相对应的放大以及对相应的不确定性作出指示。

### 3.3 校验值 (VM), 输出传感器状态 (OSS) 以及其故障诊断的应用

在工业现场中, 在仪表校验或运行的时候, 遭遇噪声的概率极高, 因此, 对软测量输出或在线仪

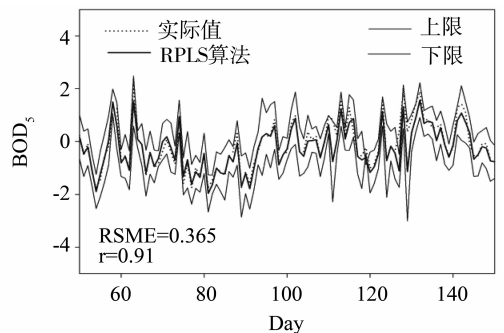


图 3 带 90% 置信区间的自确认软测量的预测效果

Fig. 3 Prediction result of self-validating soft sensors with 90% confidence

表的输出的校验是必不可少的。由于 RPLS 软测量模型的输出是独立于在线仪表输出的, 因此, 软测量的输出及其置信区间能用来校验在线仪表。正如图 4 (a) 和 (b) 分别分析仪表发生突变和输出值漂移两种故障, 在 90% 的置信度下, 都能够准确的检测出来。此外, SEVA 软测量能够很好的校验软测量在线仪表能否发生故障并且指示出在线仪表当前的状态, 从而进一步确认 UV 的有效性和其应用于获取 VM (此处即软测量的输出值)。

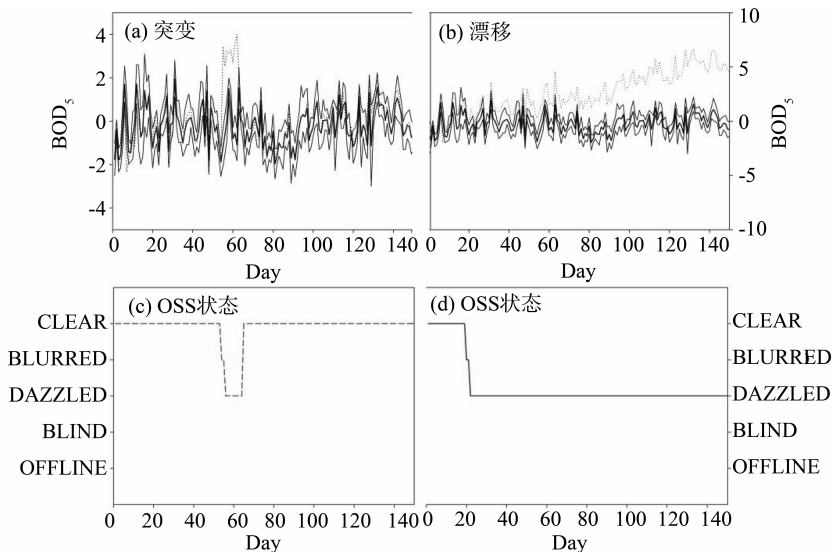


图4 突变和漂移两种故障状态下的故障诊断以及状态预警

Fig. 4 Fault detection, diagnosis and status alarm under the bias and drift scenario

## 4 结论

本文解决了传统软测量的不可靠性问题,并提出了SEVA软测量模型的概念。SEVA软测量不仅能实现模型的自诊断和数据的自重构,而且能够提供多种输出,包括预测值(PV),软测量输入传感器状态(ISS),输出传感器状态(OSS),校验输出(VM)和能描述软测量输出可信度的不确定性值(UV)。算法在污水厂的 $BOD_5$ 预测中得到了广泛的应用。基于PCA模型的SPE和SVI指标能很好的辨识和重构软测量某个输入传感器发生故障时候的数据,而且RPLS算法能在出现故障的情况下预测污水厂的出水指标。此外,SEVA软测量还能用于校验在线仪表,这些在污水厂的出水指标预测中得到了进一步的验证。所提出方法能推广到其他类似的大延迟生化处理过程。

### 参考文献:

[1] 黄道平,刘乙奇,李艳. 软测量在污水处理过程中的研究与应用[J]. 化工学报,2012,62(1):1-9.

[2] 李战明,王云峰,袁占亭. 基于流量特性的调节阀阀芯曲线智能拟合[J]. 中山大学学报:自然科学版,2009,48(1):22-25.

[3] 汤议平,金福江. 低浓度混合染料间歇式染色过程色泽软测量[J]. 仪器仪表学报,2011,32(3):690-694.

[4] YAN W, SHAO H, Wang X. Soft sensing modeling based on support vector machine and Bayesian model selection [J]. Computers & Chemical Engineering, 2004, 28(8): 1489 - 1498.

[5] KADLEC P, GABRYS B, STRANDT S. Data-driven Soft Sensors in the process industry [J]. Computers and

Chemical Engineering, 2009, 33(1): 795 - 814.

[6] YUE H H, QIN S J. Reconstruction-Based Fault Identification Using a Combined Index [J]. Industrial & Engineering Chemistry Research, 2001, 40(20): 4403 - 4414.

[7] MARTIN E B, MORRIS A J. Non-parametric confidence bounds for process performance monitoring charts [J]. Journal of Process Control, 1996, 6(6): 349 - 358.

[8] 罗中良,麦宜佳,余剑峰. 基于多传感器的神经网络模式识别方法[J]. 中山大学学报:自然科学版, 2001,40(5):21-24.

[9] QIN S J, YUE H, DUNIA R. Self-Validating Inferential Sensors with Application to Air Emission Monitoring [J]. Industrial & Engineering Chemistry Research, 1997, 36(5): 1675 - 1685.

[10] HENRY M P, CLARKE D W. The self-validating sensor: rationale, definitions and examples [J]. Control Engineering Practice, 1993, 1(4): 585 - 610.

[11] LIU X, KRUGER U, LITTLER T, et al. Moving window kernel PCA for adaptive monitoring of nonlinear processes [J]. Chemometrics and Intelligent Laboratory Systems, 2009, 96(2): 132 - 143.

[12] JOE Q S. Recursive PLS algorithms for adaptive data modeling [J]. Computers & Chemical Engineering, 1998. 22(4/5): 503 - 514.

[13] DUNIA R, QIN S J, EDGAR T F, et al. Identification of faulty sensors using principal component analysis [J]. Aiche Journal, 1996, 42(10): 2797 - 2812.

[14] PAPAPOPOULOS H, PROEDROU K, VOVK V, et al. Inductive confidence machines for regression [C] // ECML 2002 in Machine Learning, Berlin: Springer, 2002:185 - 194.

[15] RALLO R, GIRALT F. Best feature selection and data completion for the design of soft neural sensors [C] // Proceedings of AIChE 2003, 2nd Topical Conference on Sensors, San Francisco: ACS, 2003.