

# 复杂高炉炼铁过程的数据驱动建模及预测算法

郜传厚<sup>1</sup> 渐 令<sup>2</sup> 陈积明<sup>3</sup> 孙优贤<sup>3</sup>

**摘 要** 高炉炼铁过程的控制意味着控制高炉铁水温度及成份在指定的范围. 本文以高炉炉内热状态的重要指示剂——高炉铁水硅含量为研究对象, 针对机理建模难以准确预测、控制高炉铁水硅含量的发展变化, 利用数据驱动建模的思想, 建立了基于多元时间序列的高炉铁水硅含量数据驱动预测模型. 实例分析表明, 建立的数据驱动预测模型能够很好地预测高炉铁水硅含量, 连续预测 167 炉高炉铁水硅含量, 命中率高达 83.23%, 预测均方根误差为 0.07260. 这些指标均优于基于单一硅时间序列所建立的数据驱动模型, 对实际生产具有很好的指导作用.

**关键词** 高炉炼铁过程, 数据驱动, 预测模型, 铁水硅含量, 时间序列  
**中图分类号** TP27

## Data-driven Modeling and Predictive Algorithm for Complex Blast Furnace Ironmaking Process

GAO Chuan-Hou<sup>1</sup> JIAN Ling<sup>2</sup> CHEN Ji-Ming<sup>3</sup> SUN You-Xian<sup>3</sup>

**Abstract** Control of blast furnace ironmaking process means to control the temperature and compositions of blast furnace hot metal within specific bounds. For the silicon content in blast furnace hot metal as a targeted indicator of in-furnace thermal state, a data-driven based predictive model is constructed in this paper utilizing the information of multivariate time series measured from the blast finance system. Through an example, it is indicated that the constructed data-driven based predictive model has a good performance in predicting the silicon content in hot metal, with 83.23% percentage of target hitting and 0.07260 of the root mean square error of prediction when the size of predicted sample set is 167. These criteria are better than those achieved by data-driven predictive model based on univariate time series, implying that the constructed data-driven predictive model based on multivariate time series can act as an important tool to serve the production process.

**Key words** Blast furnace ironmaking process, data-driven, predictive model, silicon content in hot metal, time series

高炉炼铁过程是一个高度复杂的非线性过程, 其实质是将铁从铁矿石等含铁化合物中还原出来. 冶炼期间, 炉内将发生复杂的气-固-固-固-固-液相反应, 并伴随有高温、高压、多相共存、化学反应与传递现象同时发生等特点. 高炉炼铁过程的复杂性造成了对其进行有效控制十分困难, 时至今日, 炼铁过程仍要求高炉工长凭经验进行调节, 不能实现自动控制. 目前, 高炉炼铁过程的建模与控制研究已引起控制理论界和工业界的高度重视, 成为当今

冶金科技发展的前沿课题<sup>[1]</sup>.

长期以来, 国内外研究人员根据高炉炼铁内部所发生的化学反应和传递现象建立了多种机理数学模型<sup>[2-4]</sup>. 这些模型在理论上对于揭示高炉内部现象和反映高炉炼铁机理起了一定的积极作用, 但亦存在着准确性低及计算耗时多等缺点. 鉴于此, 数据驱动建模技术正被广泛地研究, 以试图实现对高炉炼铁过程的实时模拟和控制. 数据驱动建模技术是以描述样本数据的特征作为建模的主要准则, 在“数据为自身说话”的信念之下分析系统变量间的相互关系, 其实质是一种“黑箱”建模技术. 高炉在炼铁过程中, 具有丰富的在线和离线测量数据, 如原料参数, 包括铁矿石成份、焦炭负荷、喷煤速率、焦炭比等; 鼓风参数, 包括风量、风温、风压、富氧程度、鼓风湿度等; 铁水成份参数, 包括铁水硅含量、铁水硫含量等; 炉渣成份、炉渣碱度以及透气性指数等, 这为实现高炉炼铁过程数据驱动建模提供了可能. 目前, 利用数据驱动的思想所建立的高炉预测及控制模型主要有自回归模型<sup>[5]</sup>、神经网络模型<sup>[6-7]</sup>、非线性时间序列分析模型<sup>[8]</sup>、模糊模型<sup>[9]</sup>、贝叶斯网络模型<sup>[10]</sup>、偏最小二乘模型<sup>[11]</sup>、支持向量机模型<sup>[12]</sup>等.

收稿日期 2008-12-08 收修改稿日期 2009-03-04  
Received December 8, 2008; in revised form March 4, 2009  
国家自然科学基金 (10826100), 高等学校博士点科研基金 (20070335161), 浙江省自然科学基金 (Y107110), 浙江大学工业控制技术国家重点实验室开放课题 (ICT0904) 资助  
Supported by National Natural Science Foundation of China (10826100), Research Fund for Doctoral Program of Higher Education of China (20070335161), Natural Science Foundation of Zhejiang Provincial (Y107110), and Open Project of State Key Laboratory of Industrial Control Technology, Zhejiang University (ICT0904)  
1. 浙江大学数学系 杭州 310027 2. 中国石油大学 (华东) 数学与计算科学学院 东营 257061 3. 浙江大学控制系 杭州 310027  
1. Department of Mathematics, Zhejiang University, Hangzhou 310027 2. School of Mathematica and Computational Science, China University of Petroleum, Dongying 257061 3. Department of Control, Zhejiang University, Hangzhou 310027  
DOI: 10.3724/SP.J.1004.2009.00725

上述模型在建模时,多是利用冶炼机理、专家经验和一些统计方法,如相关系数分析以及主元分析等,确定输入、输出变量;然后再利用神经网络、贝叶斯网络、偏最小二乘法、支持向量机等工具的非线性逼近能力,通过不断训练学习的办法获取输入与输出之间的逼近函数关系式,并在此基础上建立相应的预测、控制模型。此外,高炉炼铁过程的混沌预测模型<sup>[13]</sup>也是一种基于数据驱动模型,其建立的依据是通过分析高炉生产数据,计算表征系统具有混沌性的特征指标,如 Lyapunov 指数<sup>[13]</sup>、预测熵<sup>[14]</sup>、关联维数<sup>[15]</sup>等,确定高炉冶炼过程的动态学具有混沌特性,然后利用混沌理论和方法对高炉炼铁过程进行建模和控制。这些数据驱动模型在不同时期及不同生产条件下都曾起到一定的积极作用,但也有局限性,如自回归模型<sup>[5]</sup>只适用于平稳的炉况;神经网络、贝叶斯网络等模型<sup>[6-12]</sup>不具有普适性,难以与高炉冶炼机理相结合;混沌模型尽管能综合考虑冶金工艺机理及生产数据特性,但所建立的高炉铁水硅含量预测模型<sup>[13,16]</sup>仅利用硅含量的历史数据,没有充分利用高炉生产过程丰富的数据资源。

针对现有高炉炼铁过程数据驱动模型在实际使用中所受到的种种限制、混沌模型在高炉建模时所表现的能综合考虑冶金工艺机理及生产数据特性的双重优点以及国内外文献尚无报道过同时考虑高炉炼铁过程多个变量的混沌模型,本文在现有对高炉炼铁过程混沌辨识<sup>[13-15]</sup>、混沌建模(利用单一硅序列)<sup>[13,16]</sup>研究的基础上,继续深入探测高炉炼铁过程所具有的混沌规律,特别是考虑高炉炼铁过程的其他变量,如铁水硫含量对硅含量变化的影响。利用多元时间序列相空间重构的方法<sup>[17]</sup>,建立基于数据驱动的高炉铁水硅含量预测模型。实例分析表明,建立的多元时间序列数据驱动预测模型,相比较于单元时间序列预测模型<sup>[13,16]</sup>,具有形式简单、预测精度高、计算复杂度低等优点,可为高炉工长控制高炉提供有效的指导。

## 1 基于数据驱动的预测算法

### 1.1 相空间重构

相空间重构技术又称嵌入技术,它是上世纪 80 年代初由 Packard 等<sup>[18]</sup>首次提出,并经 Takens<sup>[19]</sup>从理论上加以证明的,其实质是用系统的一个或多个变量的时间序列来重构相空间,所得的新系统在拓扑结构和概率特性等本质特征上与原系统保持一致。因此,可通过分析研究新系统的特征来确定原系统特征。

如某一动力系统  $\Phi: \mathbf{R}^n \rightarrow \mathbf{R}^m$ , 设  $h_j: \mathbf{R}^n \rightarrow \mathbf{R}$  为二阶可微的测量函数,则可得测量序列  $\{y_i^{(j)} =$

$h_j(\mathbf{X}_i)\}_{i=1}^N$ , 其中  $y_i^{(j)} \in \mathbf{R}$  是输出变量,  $\mathbf{X} \in \mathbf{R}^n$  是状态矢量,  $j = 1, 2, \dots, m$ ,  $i$  表示离散时间。利用相空间重构技术<sup>[18]</sup>,将测量序列嵌入到  $d$  维的相空间中去,其状态矢量可表示为

$$\mathbf{Y}_k = (y_k^{(1)}, y_{k-\tau_1}^{(1)}, \dots, y_{k-(d_1-1)\tau_1}^{(1)}, y_k^{(2)}, y_{k-\tau_2}^{(2)}, \dots, y_{k-(d_2-1)\tau_2}^{(2)}, \dots, y_k^{(m)}, y_{k-\tau_m}^{(m)}, \dots, y_{k-(d_m-1)\tau_m}^{(m)})^T \quad (1)$$

式中  $\tau_j$  和  $d_j$  分别是时间序列  $\{y_i^{(j)}\}_{i=1}^N$  的嵌入时滞和嵌入维数,且满足  $\sum_{j=1}^m d_j = d$ ,  $k = \max\{(d_1 - 1)\tau_1, (d_2 - 1)\tau_2, \dots, (d_m - 1)\tau_m\} + 1, \dots, N$ 。相空间的状态按照下述函数关系进行演化

$$\mathbf{Y}_{k+1} = \mathbf{F}(\mathbf{Y}_k) \quad (2)$$

Takens 定理<sup>[19]</sup>认为,当嵌入维数  $d$  足够大时,映射  $\mathbf{F}: \mathbf{R}^d \rightarrow \mathbf{R}^d$  与  $\Phi: \mathbf{R}^n \rightarrow \mathbf{R}^n$  是拓扑等价的,所以可通过研究  $\mathbf{F}$  的特性来确定  $\Phi$  的特性;同时 Takens 定理<sup>[19]</sup>也指出,当序列长度无限长且无噪声污染时,单一变量时间序列即包含系统演化的所有信息,因此只要利用单一变量时间序列就可重构出原系统的相空间,即  $m = 1$ 。

### 1.2 预测算法设计

由式 (2) 知,系统将来行为的预测可以通过寻找合适的映射  $\mathbf{F}$  来实现。重新改写式 (2),可得

$$\begin{cases} y_{k+1}^{(1)} = \mathbf{F}_1(\mathbf{Y}_k) \\ y_{k+1}^{(2)} = \mathbf{F}_2(\mathbf{Y}_k) \\ \vdots \\ y_{k+1}^{(m)} = \mathbf{F}_m(\mathbf{Y}_k) \end{cases} \quad (3)$$

预测输出将来值即寻找合适的  $\mathbf{F}_i$  ( $i = 1, 2, \dots, m$ )。根据  $\mathbf{F}_i$  的类型,可将预测模型分为全局模型、局部模型及半局部模型,其中局部模型因其计算简单、易于实现而在实际中得到广泛应用,其基本思想是根据历史上与当前状态相似的状态的演变趋势,预测当前状态的演变趋势,它所利用的只是有限个邻近点处的信息,而不是利用全部的样本点来拟合  $\mathbf{F}_i$ 。

以预测  $y_{k+1}^{(1)}$  为例,考虑局部线性模型,即

$$\hat{y}_{k+1}^{(1)} = \mathbf{F}_1(\mathbf{Y}_k) \approx a_0 + \boldsymbol{\alpha}^T \mathbf{Y}_k + \varepsilon_{k+1} \quad (4)$$

式中  $\hat{y}_{k+1}^{(1)}$  是  $y_{k+1}^{(1)}$  的预测值,  $\varepsilon_{k+1}$  是噪声(一般假定为具有正态分布的白噪声),  $a_0, \boldsymbol{\alpha} = (a_1^{(1)}, a_2^{(1)}, \dots, a_{d_1}^{(1)}, a_1^{(2)}, \dots, a_{d_1}^{(m)}, \dots, a_{d_m}^{(m)})^T$  是待定系数,可通过最小二乘拟合而得。根据 Farmer 等<sup>[20]</sup>提出的局部预测的思想,可设计预测算法为:

**步骤 1.** 确定各个变量时间序列的重构参数  $\tau_j$  和  $d_j$ ,  $j = 1, 2, \dots, m$ , 并设定  $k = N$ 、邻域半径  $\xi$ 、预测精度  $r^*$  及预测样本长度  $l$ 。

**步骤 2.** 搜寻  $\mathbf{Y}_k$  的  $\xi$  邻域点  $\mathbf{Y}_{k_1}, \mathbf{Y}_{k_2}, \dots, \mathbf{Y}_{k_{n(\xi)}}$ , 即满足

$$\|\mathbf{Y}_k - \mathbf{Y}_{k_j}\| \leq \xi, \quad k_j < k, \quad j = 1, 2, \dots, n(\xi) \quad (5)$$

式中  $\|\cdot\|$  表示无穷范数.

**步骤 3.** 确定这  $n(\xi)$  个邻域点的下一步迭代点  $\mathbf{Y}_{k_1+1}, \mathbf{Y}_{k_2+1}, \dots, \mathbf{Y}_{k_{n(\xi)+1}}$ , 并求它们的第一个分量  $y_{k_1+1}^{(1)}, y_{k_2+1}^{(1)}, \dots, y_{k_{n(\xi)+1}}^{(1)}$ .

**步骤 4.** 利用最小二乘法拟合参数  $a_0$  和  $\alpha$ , 使得下式达到最小值

$$\min S = \frac{1}{n(\xi)} \sum_{i=1}^{n(\xi)} (y_{k_i+1}^{(1)} - a_0 - \alpha^T \mathbf{Y}_{k_i})^2 \quad (6)$$

即通过求解

$$\frac{\partial S}{\partial a_0} = 0, \quad \frac{\partial S}{\partial \alpha} = \mathbf{0} \quad (7)$$

可得

$$\begin{cases} \alpha^T = \left( \sum_{i=1}^{n(\xi)} y_{k_i+1}^{(1)} \mathbf{Y}_{k_i}^T - \frac{1}{n(\xi)} \sum_{i=1}^{n(\xi)} y_{k_i+1}^{(1)} \sum_{i=1}^{n(\xi)} \mathbf{Y}_{k_i}^T \right) \times \\ \left( \sum_{i=1}^{n(\xi)} \mathbf{Y}_{k_i} \mathbf{Y}_{k_i}^T - \frac{1}{n(\xi)} \sum_{i=1}^{n(\xi)} \mathbf{Y}_{k_i} \sum_{i=1}^{n(\xi)} \mathbf{Y}_{k_i}^T \right)^{-1} \\ a_0 = \frac{1}{n(\xi)} \left( \sum_{i=1}^{n(\xi)} y_{k_i+1}^{(1)} - \alpha^T \sum_{i=1}^{n(\xi)} \mathbf{Y}_{k_i} \right) \end{cases} \quad (8)$$

**步骤 5.** 利用式 (4) 对  $y_{k_i+1}^{(1)}$  进行预测.

**步骤 6.** 更新  $k = k + 1$ , 重复步骤 2~5, 直到  $k = N + l$ .

**步骤 7.** 计算预测精度

$$r = \sqrt{\frac{1}{l} \sum_{i=1}^l (y_{N+i}^{(1)} - \hat{y}_{N+i}^{(1)})^2} \quad (9)$$

比较  $r$  及  $r^*$  的大小, 若  $r \leq r^*$ , 停止计算; 若  $r > r^*$ , 修正  $\xi$ ,  $k = N$ , 重复步骤 2~6.

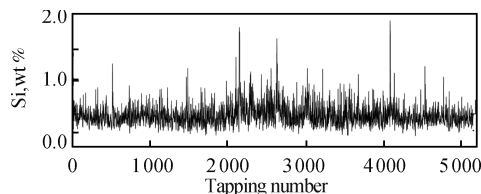
## 2 应用验证

### 2.1 生产数据

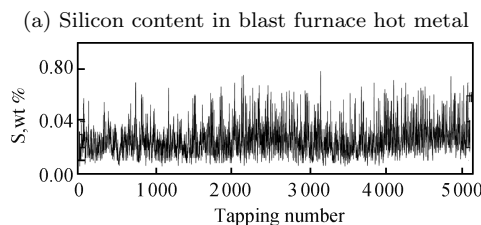
本文以国内某钢铁厂 1 号高炉采集的生产数据为研究对象, 验证上述预测模型及算法的有效性.

在高炉生产过程的各项输出数据中, 高炉炉温作为高炉炉缸热状态的指示, 是评定生铁质量和高炉炉况稳定性的重要指标. 一方面为了保证生铁质量, 高炉炉温必须处于某一指定的范围之内; 另一

方面为了保证高炉炉况的稳定和顺行, 高炉炉温的波动幅度又不宜过大. 在实际生产中, 由于测量困难, 常常用高炉铁水硅含量 (Si, wt % — 重量百分率) 来表示高炉炉温, 间接地反映高炉炉缸的热状态. 铁水硅含量的波动反映了炉温的波动, 它的持续上升和持续下降, 则反映了炉缸是向热还是向凉变化, 因此高炉铁水硅含量的预测与控制构成了高炉操作指导的基础. 鉴于此, 本文选用高炉铁水硅含量为预测目标 (记作  $y^{(1)}$ ). 文献 [13–15] 的研究表明, 高炉铁水硅含量时间序列在其演化过程中具有正的最大 Lyapunov 指数<sup>[13]</sup>、预测熵为大于 0 的有限值<sup>[14]</sup>、关联维数存在且为分数<sup>[15]</sup>, 这说明高炉炼铁过程具有混沌性, 也为将混沌理论用于研究高炉铁水硅含量的建模和控制提供了依据. 因此, 可以用第 1.2 节建立的预测算法对高炉铁水硅含量进行预测. 虽然在理论上可通过无限长且无噪声污染的单一高炉铁水硅含量时间序列重构系统相空间<sup>[19]</sup>, 但在实际中, 过程噪声及测量噪声总是存在于实际测量序列中. 因此, 为了获取更多的系统信息, 本文同时选用铁水中另外一个重要成份 — 高炉铁水硫含量 (S, wt %, 记作  $y^{(2)}$ ) 作为输出变量, 其时间序列也参与重构系统相空间. 图 1 是这两个输出变量的时间序列. 时间序列长度为 5167, 采样周期大约为 2 h, 其中前 5000 炉数据作为训练集, 后 167 炉作为测试集. 这样选择训练集及测试集是为了与文献 [16] 中的结果进行对比.



(a) 高炉铁水硅含量



(b) 高炉铁水硫含量

(b) Sulfur content in blast furnace hot metal

图 1 时间序列图

Fig. 1 Time series

### 2.2 重构参数确定

在利用式 (1) 对高炉炼铁过程进行相空间重构时, 有两种参数需要确定, 即时滞  $\tau_i$  和嵌入维数  $d_i$  ( $i = 1, 2$ ). 时滞的选取应使得相空间矢量点最大限

度的独立但又要保持一定的关联, 通常可采用自相关函数法<sup>[21]</sup> 确定时滞, 即计算

$$R(\tau) = \frac{\sum_{i=1}^{N-\tau} (y_i - \bar{y})(y_{i+\tau} - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (10)$$

式中  $\bar{y}$  表示时间序列  $\{y_i\}_{i=1}^N$  的平均值. 当  $R(\tau) \approx e^{-1}$  时, 对应的  $\tau$  即为所确定的时滞. 图 2 给出了通过自相关函数所确定的高炉铁水硅含量和硫含量时间序列对应的时滞. 从图 2 可看出, 当时滞为 3 时, 硅含量时间序列的自相关函数最接近  $e^{-1}$ , 所以  $\tau_1 = 3$ ; 当时滞为 2 时, 硫含量时间序列的自相关函数最接近  $e^{-1}$ , 所以  $\tau_2 = 2$ .

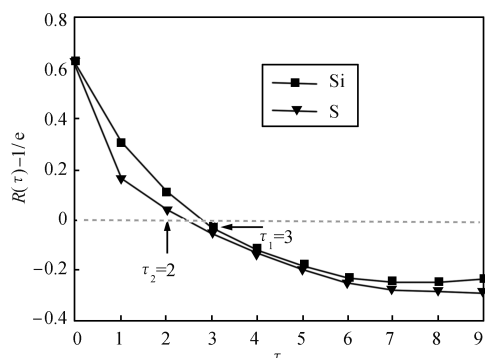


图 2 高炉铁水硅含量和硫含量时间序列  
重构参数 — 时滞的确定

Fig. 2 Determination of time delay used for  
reconstructing state space from the time series of silicon  
content and sulfur content in blast furnace hot metal

嵌入维数  $d_i$  ( $i = 1, 2$ ) 的确定可采用 Kennel 等<sup>[22]</sup> 和 Boccaletti 等<sup>[23]</sup> 提出的伪邻域点方法进行, 具体步骤如下:

**步骤 1.** 设定初始值:  $d_i = 1$ , 计数器  $M_i(d) = 0$  ( $i = 1, 2$ ), 阈值  $\sigma$  和  $\delta$ ;

**步骤 2.** 在  $d = d_1 + d_2$  维空间内, 计算  $\mathbf{Y}_k$  与其最近邻域点  $\mathbf{Y}_{k,NN}$  的欧式距离  $s(\mathbf{Y}_k, \mathbf{Y}_{k,NN}; d)$ ;

**步骤 3.** 更新  $d = d + 1$ , 可通过  $d_1 + 1$  或  $d_2 + 1$  实现, 计算此时  $d + 1$  维的距离为

$$s_i(\mathbf{Y}_k, \mathbf{Y}_{k,NN}; d + 1) = \sqrt{s^2(\mathbf{Y}_k, \mathbf{Y}_{k,NN}; d) + (y_{k-d_1\tau_1}^{(i)} - y_{(k,NN)-d_1\tau_1}^{(i)})^2}$$

其中,  $i = 1, 2$ , 若  $s_i/s \geq \sigma$ , 则认为当维数  $d \rightarrow d + 1$  时, 出现一对伪邻域点, 相应地, 计数器  $M_i(d) = M_i(d) + 1$ ;

**步骤 4.** 改变  $k$  值, 重复步骤 2 和 3, 直至  $k$  历经相空间轨道中的所有点, 若此时  $M_i(d) \geq \delta$ , 则  $d_i$

$= d_i + 1$ ;

**步骤 5.** 更新  $d = d + p$ , 其中  $p$  为  $M_i(d)$  ( $i = 1, 2$ ) 中计数器大于  $\delta$  的个数, 重复步骤 2~4, 直至所有的  $M_i(d) < \delta$  ( $i = 1, 2$ ), 便得到最优的  $d_i$  ( $i = 1, 2$ ), 因此可得最优嵌入维数  $d = d_1 + d_2$ .

在运用上述算法确定嵌入维数时, 不同的阈值  $\sigma$  和  $\delta$  将产生不同的结果.  $\sigma$  用于判别相空间中最邻近的矢量点是否为伪邻域点, 理论上其值应大于 1, 并且嵌入维数  $d$  的值随  $\sigma$  的增大将趋于饱和<sup>[23]</sup>;  $\delta$  表示伪邻域点的百分比上限, 超出此上限重构过程即停止, 其值的选取应越小越好<sup>[23]</sup>. 对于干净序列  $\delta = 0$ , 即表示在相空间中伪邻域点个数为 0, 但在实际中, 由于序列内包含噪声, 因此  $\delta$  应选为大于 0 的正数<sup>[23]</sup>. 由于  $\sigma$  和  $\delta$  不是嵌入维数  $d$  的连续函数, 所以它们的值不能经严格的数学推导得出. 在实际计算时, 常常从模型精度与计算速度两方面考虑来确定  $\sigma$  和  $\delta$ , 因此所得结果未必是全局最优的, 但应能满足实际要求. 考虑到  $\sigma$ 、 $\delta$  和  $d$  的关系, 本文采用正交实验的方法确定它们的值, 即选取使得  $d$  饱和的最小  $\sigma$ , 并选取此时对应的最小  $\delta$ . 通过正交实验表明, 本文选取的硅序列和硫序列在  $\sigma = 10$ 、 $\delta = 15$  时嵌入维数达到饱和, 此时  $d_1 = 8$ 、 $d_2 = 3$ , 即相空间的维数为  $d_1 + d_2 = 11$ . 这一结果低于单一硅时间序列重构需要的维数  $d = 18$ <sup>[16]</sup>, 因此所建立的模型的预测速度也会快于基于单一硅时间序列的预测模型的计算速度.

### 2.3 预测结果与讨论

利用上述所得的时滞及嵌入维数结果, 即可通过预测算法对图 1 中的测试集进行预测, 设定  $r^* = 0.08538$ , 该数值是文献 [16] 的最好结果. 考虑到铁水硅含量与铁水硫含量不在一个数量级, 在利用式 (5) 寻找  $\mathbf{Y}_k$  的最近邻域点时, 对硅序列和硫序列做了归一化处理. 计算表明, 当  $\xi = 1.0$  时,  $r = 0.07260 < r^*$ . 图 3 (见下页) 是相应的预测结果图. 从图 3 可以看出, 预测值能很好地跟踪实际值的变化, 特别是在一些炉次波动比较大的情况, 如 18、19 炉, 58、59 炉, 65、66 炉, 88、89 炉, 113、114 炉和 146、147 炉等, 预测值也能及时地反映实际值的变化, 这在一定程度上改善了一般时间序列预测模型强惯性的缺点<sup>[8]</sup>, 即预测值难以及时跟踪波动比较大的硅含量变化. 这一现象也可通过将预测结果与一般时间序列模型的预测结果对比加以说明, 为此构造最为简单的时间序列模型

$$\hat{y}_{k+1}^{(1)} \approx y_k^{(1)} \quad (11)$$

图 4 (见下页) 给出了该模型的预测结果. 观察发现, 在上述硅值波动比较大的炉次, 硅预测值均难

以及跟踪实际值的变化. 进一步考虑模型的预测命中率  $J$ , 其定义为

$$J = \frac{1}{l} \sum_{i=1}^l H_i \times 100\% \quad (12)$$

其中,  $l$  表示预测样本长度,  $\hat{y}_i^{(1)}$  表示  $y_i^{(1)}$  的预测值,  $H(\cdot)$  是 Heaviside 函数, 定义为

$$H_i = \begin{cases} 1, & \|\hat{y}_i^{(1)} - y_i^{(1)}\| < 0.1 \\ 0, & \text{否则} \end{cases} \quad (13)$$

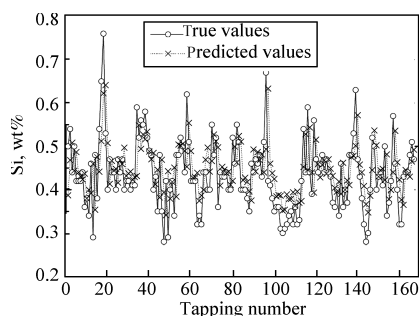


图3 基于多元时间序列的高炉铁水硅含量的混沌局部线性预测结果图

Fig. 3 The chaotic local linear prediction results of silicon content in hot metal based on multivariate time series

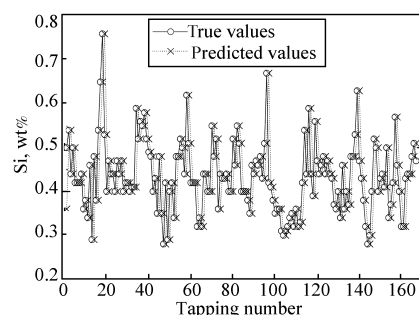


图4 基于单一硅序列的高炉铁水硅含量预测结果图

Fig. 4 The prediction results of silicon content in hot metal based on single silicon time series

预测命中率是冶金生产中用来衡量模型优劣的重要指标. 通过计算得, 模型 (4) 的预测命中率为  $J = 83.23\%$ , 而模型 (11) 的预测命中率仅为  $74.85\%$ , 上述硅值波动比较大的炉次模型 (11) 均未命中. 这一结果对于连续预测 167 炉铁水硅含量而言, 是十分令人鼓舞的, 并且与模型 (11) 相比, 模型 (4) 还有很大的改进空间. 此外, 这个结果也较文献 [16] 所建立的基于单一硅序列的二阶预测模型的结果  $J = 79.04\%$  有了较大改进, 对高炉工长控制高炉操作具有很好的指导意义.

从以上结果可看出, 基于多元时间序列的混沌局部预测模型比单元时间序列模型具有更高的命中

率及预测精度, 同时也较单元序列模型具有更强的跟踪硅含量变化的能力. 进一步对比模型发现, 多元时间序列预测模型重构维数为 11, 而基于单一硅序列预测模型的重构维数为 18<sup>[16]</sup>, 前者的模型参数比后者少, 因此多元序列模型在具有更高精度的前提下, 还具有模型形式简单、计算复杂度低、运算速度快等优点. 同时, 与文献 [16] 所建立的二阶模型相比, 本文所采用的仅仅是基于多元时间序列重构的混沌局部线性预测模型, 而且建立模型时只是增加了硫序列的信息, 因此模型还具有很大的改进空间. 可以预见, 随着模型阶数 (非线性) 的提高或更多高炉变量信息 (如输入变量信息) 的利用, 预测模型的精度有望进一步提高. 不过也应注意到, 基于多元时间序列的预测模型比基于单元时间序列的预测模型需要更多的高炉信息, 因此在检测设备以及数据分析上具有更高的要求. 此外, 在重构相空间确定阈值  $\sigma$  和  $\delta$  时, 所得结果并不是全局最优的, 如何获取  $\sigma$  和  $\delta$  的全局最优值十分困难, 值得进一步对其进行研究.

### 3 结论

1) 基于多元时间序列的高炉铁水硅含量数据驱动预测模型能很好地反映高炉铁水硅含量的变化, 取得了较高的命中率  $J = 83.23\%$  和较低的预测均方根误差  $r = 0.07260$ , 并且在一些硅含量波动比较大的炉次, 多元序列模型也能及时跟踪硅值的变化, 这在一定程度上改善了一般时间序列模型强惯性的缺点.

2) 与基于单一硅序列的高炉铁水硅含量预测模型<sup>[16]</sup> 相比较, 基于多元时间序列的数据驱动模型具有精度高、形式简单、计算复杂度低、运算速度快等优点, 但亦存在着要求更多的高炉数据检测设备及更为充分的数据分析等条件. 同时随着更多高炉变量的引入, 最优阈值  $\sigma$  和  $\delta$  的确定也会变得更为困难.

3) 目前, 对基于多元时间序列的高炉铁水硅含量数据驱动模型的研究尚处于探索阶段, 模型的建立仅仅利用高炉铁水硅含量和铁水硫含量的信息, 并且是线性的. 从应用验证的结果可以看出, 模型在描述高炉铁水硅含量的发展变化上具有很大的潜力, 并有望通过提高模型阶数 (非线性)、引入更多高炉变量信息、寻求阈值  $\sigma$  和  $\delta$  的全局最优值等手段进一步提高模型精度, 最终实现高炉铁水硅含量的准确控制.

### References

- 1 State Metallurgical Industry Bureau, Chinese Society of Metal, Center of Metallurgical Science and Technology. Guide for Development of Metallurgical Science and Tech-

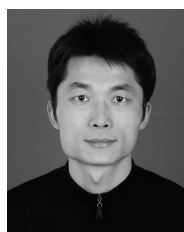
- nology (2000–2005), 1999  
(国家冶金工业局规划发展司, 中国金属学会, 冶金科技发展中心合编. 冶金科技发展指南 (2000–2005 年), 1999)
- 2 Takahashi H, Kawai H, Kobayashi M, Fukui T. Two dimensional cold model study on unstable solid descending motion and control in blast furnace operation with low reducing agent rate. *The Iron and Steel Institute of Japan International*, 2005, **45**(10): 1386–1395
  - 3 Nogami H, Chu M S, Yagi J. Multi-dimensional transient mathematical simulator of blast furnace process based on multi-fluid and kinetic theories. *Computer and Chemical Engineering*, 2005, **29**(11-12): 2438–2448
  - 4 Chu M S, Yang X F, Shen F M, Yagi J, Nogami H. Numerical simulation of innovative operation of blast furnace based on multi-fluid model. *Journal of Iron and Steel Research*, 2006, **13**(6): 8–15
  - 5 Saxén H. Short-term prediction of silicon content in pig iron. *Canadian Metallurgical Quarterly*, 1994, **33**(4): 319–326
  - 6 Radhakrishnan V R, Mohamed A R. Neural networks for the identification and control of blast furnace hot metal quality. *Journal of Process Control*, 2000, **10**(6): 509–524
  - 7 Chen J. A predictive system for blast furnaces by integrating a neural network with qualitative analysis. *Engineering Applications of Artificial Intelligence*, 2001, **14**(1): 77–85
  - 8 Saxén H, Pettersson F. Nonlinear prediction of the hot metal silicon content in the blast furnace. *ISIJ International*, 2007, **47**(12): 1732–1737
  - 9 Martín R D, Obeso F, Mochón J, Barea R, Jiménez J. Hot metal temperature prediction in blast furnace using advanced model based on fuzzy logic tools. *Ironmaking and Steelmaking*, 2007, **34**(3): 241–247
  - 10 Liu Xue-Yi, Liu Xiang-Guan, Wang Wen-Hui. Application of Bayesian network to predicting silicon content in hot metal. *Iron and Steel*, 2005, **40**(3): 17–20  
(刘学艺, 刘祥官, 王文慧. 贝叶斯网络在高炉铁水硅含量预测中的应用. 钢铁, 2005, **40**(3): 17–20)
  - 11 Bhattacharya T. Prediction of silicon content in blast furnace hot metal using partial least squares (PLS). *ISIJ International*, 2005, **45**(12): 1943–1945
  - 12 Jian Ling, Liu Xiang-Guan. Application of SVM to prediction of silicon content in hot metal. *Metallurgical Industry Automation*, 2005, **29**(3): 33–36  
(渐令, 刘祥官. 支持向量机在铁水硅含量预报中的应用. 冶金自动化, 2005, **29**(3): 33–36)
  - 13 Gao C H, Zhou Z M, Qian J X. Chaotic identification and prediction of silicon content in hot metal. *Journal of Iron and Steel Research International*, 2005, **12**(5): 35–46
  - 14 Gao C H, Zhou Z M, Chen J M. Assessing the predictability for blast furnace system through nonlinear time series analysis. *Industrial and Engineering Chemistry Research*, 2008, **47**(9): 3037–3045
  - 15 Gao Chuan-Hou, Liu Xiang-Guan. Chaotic identification of BF ironmaking process I: the calculation of saturated correlative dimension. *Acta Metallurgica Sinica*, 2004, **40**(4): 347–350  
(郜传厚, 刘祥官. 高炉冶炼过程的混沌性辨识 I: 饱和关联维数的确定. 金属学报, 2004, **40**(4): 347–350)
  - 16 Gao C H, Chen J M, Zeng J S, Liu X Y, Sun Y X. A chaos-based iterated multistep predictor for blast furnace ironmaking process. *AIChE Journal*, 2009, **55**(4): 947–962
  - 17 Cao L Y, Mees A, Judd K. Dynamics from multivariate time series. *Physica D*, 1998, **121**(1-2): 75–88
  - 18 Packard N H, Grutchfield J P, Farmer J D, Shaw R S. Geometry from a time series. *Physical Review Letters*, 1980, **45**(9): 712–716
  - 19 Takens F. Detecting strange attractors in turbulence. In: *Proceedings of a Symposium Held at the University of Warwick*. Warwick, UK: Springer, 1981. 366–381
  - 20 Farmer J D, Sidorowich J J. Predicting chaotic time series. *Physical Review Letters*, 1987, **59**(8): 845–848
  - 21 Lai Y C, Ye N. Recent developments in chaotic time series analysis. *International Journal of Bifurcation and Chaos*, 2003, **13**(6): 1383–1422
  - 22 Kennel M B, Brown R, Abarbanel H D. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 1992, **45**(6): 3403–3411
  - 23 Boccaletti S, Valladares D L, Pecora L M, Geffert H P, Carroll T. Reconstructing embedding spaces of coupled dynamical systems from multivariate data. *Physical Review E*, 2002, **65**(3): 52–55



**郜传厚** 浙江大学数学系副教授. 主要研究方向为复杂系统的信号处理与数据挖掘, 复杂工业过程的模型化与控制, 复杂工程系统的应用理论. 本文通信作者. E-mail: gaochou@zju.edu.cn

(GAO Chuan-Hou Associate professor in the Department of Mathematics, Zhejiang University. His research

interest covers signal processing and data mining in complex systems, modeling and control of complex industrial processes, and theory application of complex engineering systems. Corresponding author of this paper.)



**渐令** 中国石油大学(华东)数学与计算科学学院讲师. 主要研究方向为冶金生产过程预测和建模及控制.

E-mail: jianling\_jam@163.com

(JIAN Ling Lecturer at the School of Mathematica and Computational Science, China University of Petroleum. His research interest covers prediction,

modeling, and control of metallurgical production process.)



**陈积明** 浙江大学控制系副研究员. 主要研究方向为无线传感器网络和网络化系统的控制与优化.

E-mail: jmchen@ieee.org

(CHEN Ji-Ming Associate professor in the Department of Control, Zhejiang University. His research interest covers wireless sensor networks, and op-

timization and control of networked systems.)



**孙优贤** 中国工程院院士, 浙江大学控制系教授. 主要研究方向为控制理论与应用, 复杂工业过程的建模和优化及控制. E-mail: yxsun@iipc.zju.edu.cn

(SUN You-Xian Academician of Chinese Academy of Engineering, professor in the Department of Control, Zhejiang University. His research inter-

est covers control theory and application, modeling, and optimization and control of complex industrial processes.)