

DESENVOLVIMENTO DE UM SENSOR VIRTUAL PARA MONITORAMENTO DE ÍNDICE DE WOBBE NA PLANTA DO LTQ DA ARCELORMITTAL TUBARÃO*

Rudson Silva Venturim¹
Lucianderson Marques Ferreira²
Flávio Garcia Pereira³
Cassius Zanetti Resende⁴

Resumo

Este trabalho estuda o desenvolvimento de um sensor virtual para monitoramento de índice de wobbe do gás misto (MG) utilizados em fornos de aquecimento em indústrias siderúrgicas. Para isso foram testados quatro modelos de regressão. Dois deles sendo regressões por mínimos quadrados parciais (*PartialLeastSquares*, PLS), um utilizando o algoritmo NIPALS (*NonlinearIterativePartialLeastSquares*) e outro utilizando o algoritmo SIMPLS (*simple partial least squares*). As demais regressões foram por componentes principais (PCR) e regressão por vetor de suporte (SVR). As variáveis do processo foram selecionadas inicialmente com a ajuda de um especialista do processo e posteriormente pelo método de árvore de decisão. Os valores estimados são comparados com os valores reais obtidos na planta real. Os resultados mostram que foi possível prever o índice de wobbe com erro médio quadrático com valor de 14,43%.

Palavras-chave: Sensor Virtual; Índice de Wobbe; Modelo de Regressão; Erro Médio Quadrático.

DEVELOPMENT OF A VIRTUAL SENSOR FOR MONITORING WOBBE INDEX ON THE LTQ PLANT OF ARCELORMITTAL TUBARÃO

Abstract

This paper studies the development of a virtual sensor for wobbe index monitoring of mixed gas (MG) used in reheating furnaces at the steel industry. For this, four regression models were tested. Two of these regressions were developed using Partial Least Squares (PLS) methods, the NIPALS algorithm (Nonlinear Iterative Partial Least Squares) and the SIMPLS algorithm (simple partial least squares). For the development of the others regressions, it was used the principal components regression (PCR) and support vector regression (SVR). The process variables were initially selected with the help of experts of the process and thereafter refined using the decision tree method. The estimated values are compared with the actual values obtained in the real plant. The results show that it was possible to predict the wobbe index with a mean square error with a value of 14.43%.

Keywords: Virtual Sensor; Wobbe Index; Regression Model; Mean Square Error.

¹ Bacharel em Engenharia de Controle e Automação, Técnico de Instrumentação, Manutenção Industrial, SindusAndritz/Instituto Federal do Espírito Santo, Serra, Espírito Santo, Brasil.

² Engenheiro Eletricista, Especialista em Controle e Instrumentação, Engenheiro de Confiabilidade de Equipamentos, IEC-IECC-Gerencia de Manutenção de Ativos Centralizados, ArcelorMittal Tubarão, Serra, Espírito Santo, Brasil.

³ Engenheiro Eletricista, Doutor, Professor EBTT, Programa de Pós-Graduação em Engenharia de Controle e Automação, Instituto Federal do Espírito Santo, Serra, Espírito Santo, Brasil.

⁴ Engenheiro Eletricista, Doutor, Professor EBTT, Programa de Pós-Graduação em Engenharia de Controle e Automação, Instituto Federal do Espírito Santo, Serra, Espírito Santo, Brasil.

1 INTRODUÇÃO

Nas usinas siderúrgicas integradas o ferro gusa líquido é produzido na planta do Alto Forno e, em seguida, é transformada em aço líquido na planta da Aciaria. Este aço líquido passará por transformações físicas adquirindo a forma sólida de placas de aço que, sequencialmente, será transformada em bobina de aço na planta do Laminador de Tiras a Quente (LTQ) [1]. Para serem processadas no LTQ as placas de aços devem estar a uma temperatura conhecida como temperatura de processamento. Para atingir a temperatura do processamento, as placas de aço são carregadas para um Forno de Reaquecimento (FR).

A temperatura de processamento pode alterar as propriedades físico-químicas do produto final. Assim, para garantir o aquecimento correto da placa dentro do FR, faz-se necessário o monitoramento online, inclusive, da qualidade dos gases que são queimados dentro da câmara de aquecimento do FR. Para isso, analisadores de índice de wobble são utilizados para verificar a qualidade do poder de queima dos gases combustíveis utilizados [1].

Este trabalho investiga o desenvolvimento de um sensor virtual para prever o valor do índice de wobble do Gás Misto (MG) em caso de falta do analisador de índice de wobble. Sensores virtuais são largamente utilizados com base na medição online de sensores reais disponíveis (como pressão, temperatura, vazão, dentre outros), para prever indicações que muitas vezes são impossíveis de medir fisicamente ou são muito caras [2].

Para o desenvolvimento do sensor virtual de índice de wobble do Gás Misto (MG) foram utilizadas quatro técnicas de regressão: parcial de mínimos quadrados não linear iterativo (NIPLAS), parcial de mínimos quadrados simples (SIMPLS), componentes principais (PCR) e máquina de vetor de suporte (SVM). Já para determinar quais grandezas dentre as disponíveis são importantes para a previsão do índice de wobble do Gás Misto (MG) utilizou-se a técnica da árvore de decisão.

2 MATERIAIS E MÉTODOS

Os processos industriais são equipados com uma grande variedade de sensores como temperatura, vazão, pressão, condutividade, que informam de forma online as grandezas medidas do processo. Contudo, existem outras variáveis do processo que devido à falta de sensores disponíveis ou até mesmo pelo alto custo para a implantação do sistema de medição *in-situ*, estão medidas através da realização de coleta manual e análise em laboratório de forma off-line [2].

A fim de medir de forma *online* as variáveis de processo que são de difícil obtenção ou aquelas que são muito importantes para o processo e não podem perder sua indicação, podem ser usados métodos de inteligência computacional para construir um sensor virtual de modo que se possa inferir o valor das mesmas. A base para se construir um sensor virtual é que a variável que se deseja monitorar tenha alguma relação funcional com outra variável de processo disponível em tempo real. De forma abrangente, um sensor virtual é um modelo computacional que utiliza variáveis fácil de medir que estão disponíveis no processo para inferir o valor de variáveis difícil de medir, que muitas vezes possuem o custo muito alto ou grande atraso de medição, feita por laboratório [2].

No trabalho realizado por Pani e Mohanta [3] foi desenvolvido um sensor virtual para medir a qualidade do clínquerem indústrias de cimento. As técnicas de rede neural

artificial *feed-forward* e inferência *fuzzy* foram a base do modelo desenvolvido para prever cada parâmetro associado à qualidade do clínquer. A partir de simulações de dados reais retirados de uma planta de produção de cimento, foi observado que o modelo *fuzzy* obteve um bom resultado para o monitoramento online da qualidade do clínquer.

Para o desenvolvimento de um sensor virtual é necessário seguir alguns passos [4], sendo eles a coleta e filtragem de dados, a seleção de dados de entrada, a escolha do modelo, treinamento e validação, que são discutidos nas seções seguintes.

2.1 COLETA E FILTRAGEM DE DADOS

Geralmente, as grandes empresas possuem sistemas para armazenamento de base de dados de todos os processos principais, no caso da indústria siderúrgica, suas principais plantas como Alto Forno e Laminador de Tiras a quente. Essa é a base para o desenvolvimento de um sensor virtual. Entretanto, este tipo de aquisição de dados nas indústrias geralmente vem com problemas indesejáveis como dados perdidos, *outliers* e tempos de amostragem elevados e distintos. Quando se fala em tempo de amostragem, significa que dentro da indústria, algumas variáveis possuem diferentes taxas de aquisição. Geralmente, quanto mais fácil de medir uma variável, menor será o tempo de amostragem. Como geralmente o sensor virtual é implementado para prever o valor de variáveis difíceis de medir, muitas vezes é preciso fazer um sincronismo entre estes tempos de amostragem, reduzindo o tempo de amostragem da variável de entrada do sistema [5].

No caso de dados perdidos, é comum observar que durante a amostragem de alguma variável, determinadas amostras podem vir sem valor nenhum. Nesse caso, existem duas principais formas de lidar com os dados perdidos. A primeira delas seria eliminar as amostras que contém dados perdidos. A segunda seria tentar preencher os dados perdidos com o auxílio de alguns métodos de preenchimento de valor. Caso o valor perdido seja baixo, a primeira forma pode ser aplicada, mas do contrário deve-se evitar a eliminação de amostras. No caso da segunda estratégia, uma boa aproximação seria tirar uma média das amostras sem perda de dados [6].

Os *outliers* são valores observados que divergem de forma significativa da média das amostras no geral. Eles apresentam valores inconsistentes quando comparados com o resto das amostras e isso pode afetar e muito o desempenho do sensor virtual. Estes valores podem ser causados por diversos fatores, como um sensor com mau funcionamento, um erro de comunicação ou até mesmo a degradação de um sensor. O primeiro passo para tratar um *outlier* é identificar o mesmo. Existem diversos métodos estatísticos para tal finalidade e o mais comum deles é a regra 3σ , que se baseia em uma distribuição sendo a média das amostras (μ) e desvio padrão (σ). Qualquer amostra fora do intervalo $[\mu - 3\sigma, \mu + 3\sigma]$ é considerado um *outlier* [7].

A etapa de aquisição de dados foi realizada por meio do banco de dados do processo do LTQ e da Utilidades da siderúrgica Arcelor Mittal Tubarão (AMT). Observa-se que todos os gases produzidos nas diversas áreas da usina da AMT são direcionados para três misturadoras (Mist01, Mist02 e Mist03) localizadas na planta Utilidades, que após misturados são enviados para o LTQ para o forno de reaquecimento.

Os dados de entrada do preditor proposto são variáveis de processo das três misturadoras, sendo coletados vários tipos de dados, como indicação de pressão, vazão e temperatura de gás. Já o índice de wobbe do gás MG é coletado na planta do LTQ. Inicialmente para compor o banco de dados com 11162 amostras foram

selecionados 42 atributos de entrada para a predição do índice de wobbe. A Tabela 01 mostra a descrição de cada um dos atributos de entrada e saída.

Tabela 1. Descrição das variáveis da base de dados

Número	Descrição do Atributo	Número	Descrição do Atributo
1	Controlador indicador de fluxo (%)	22	Indicador de fluxo
2	Controlador indicador de fluxo	23	Indicador de fluxo
3	Controlador indicador de fluxo (%)	24	Indicador de fluxo
4	Controlador indicador de fluxo	25	Vazão total de gás COG
5	Controlador indicador de fluxo (%)	26	Indicador de Fluxo
6	Controlador indicador de fluxo	27	Indicador de Fluxo
7	Controlador indicador de fluxo (%)	28	Indicador de Fluxo
8	Controlador indicador de fluxo	29	Indicador de Fluxo
9	Indicador de Fluxo	30	Indicador de Temperatura
10	Indicador de Fluxo	31	Indicador de Temperatura
11	Indicador de Fluxo	32	Indicador de Temperatura
12	Indicador de Fluxo	33	Indicador de Temperatura
13	Indicador de Temperatura	34	Indicador de Pressão
14	Indicador de Temperatura	35	Indicador de Pressão
15	Indicador de Temperatura	36	Indicador de Pressão
16	Indicador de Temperatura	37	Indicador de Pressão
17	Indicador de Pressão	38	Indicador de fluxo
18	Indicador de Pressão	39	Indicador de fluxo
19	Indicador de Pressão	40	Indicador de fluxo
20	Indicador de Pressão	41	Indicador de fluxo
21	Indicador de fluxo	42	Vazão total de gás BFG

2.2 SELEÇÃO DE DADOS DE ENTRADA

Durante a seleção de dados de entrada, é possível observar diversos dados reais de entrada que estão diretamente ligados à variável que se deseja inferir. Entretanto, apenas algumas variáveis de entrada são necessárias para compor o modelo do sensor virtual. A utilização de poucas variáveis de entrada vem com algumas vantagens como redução do tempo de desenvolvimento do modelo, redução do número de sensores reais, redução de custos e a possibilidade de aplicação deste modelo [2].

Uma das principais técnicas utilizadas para o processo de seleção de dados de entrada é denominado seleção supervisionada de variável. Nesta técnica, as variáveis de entrada são selecionadas tendo como finalidade buscar sempre a maior precisão nos resultados para o modelo do sensor virtual. O critério para a seleção de variáveis de entrada durante processo de busca pode ser classificado em três tipos: métodos de filtragem, método de algoritmos de envoltório e métodos algoritmos incorporados [8].

O método da Árvore de Decisão foi utilizado para classificar os 42 atributos em nível de correlação com a saída desejada. O resultado do método da Árvore de Decisão é apresentado na Figura 1. A partir da correlação de cada atributo, foram eliminados os atributos de menor correlação com a variável de saída, sendo eles os atributos

22, 23, 27, 29, 38, 39, 41 e 42. Assim, foram utilizados os 34 atributos de maior correlação como atributo de entrada nos modelo de regressão apresentados na próxima seção.

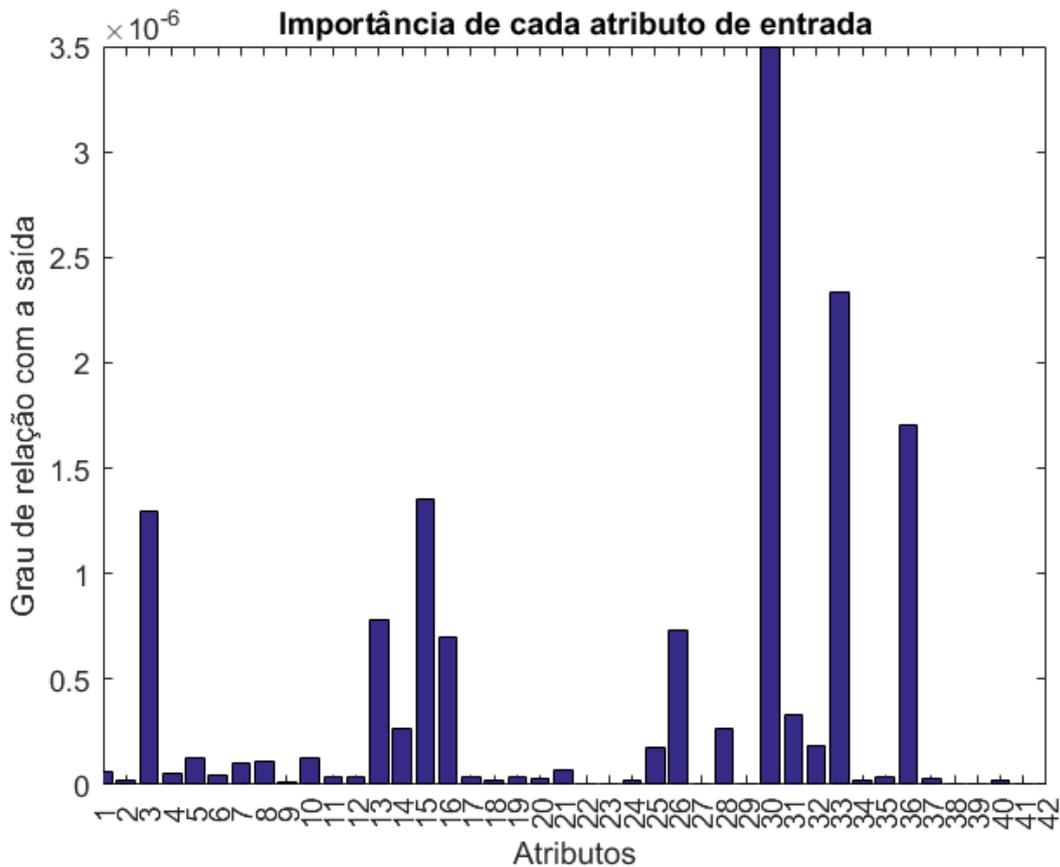


Figura1.Relação de importância de cada atributo com a saída

2.3 ESCOLHA DO MODELO E TREINAMENTO

Após coletar e selecionar os dados de entrada iniciou-se a escolha do modelo e treinamento. Foram testados quatro modelos de regressão. Dois deles sendo regressões por mínimos quadrados parciais (*PartialLeastSquares*, PLS), um utilizando o algoritmo NIPALS (*NonlinearIterativePartialLeastSquares*) e outro utilizando o algoritmo SIMPLS. As demais regressões serão por componentes principais (PCR) e regressão por de vetor de suporte (SVR).

Para o desenvolvimento da modelagem fez-se a divisão da base original em 70% dos dados para treinamento e validação; e 30% para teste. O principal objetivo é comparar estes modelos e avaliar qual pode apresentar o melhor resultado.

Uma técnica de regressão utilizada foi o *PartialLeastSquares*(PLS), sendo esta uma técnica utilizada para reduzir os atributos iniciais a um conjunto menor de componentes não correlacionados e efetua regressão de mínimos quadrados para esses componentes no lugar dos dados originais. A regressão PLS é muito útil quando se observa uma alta relação de colinearidade entre os atributos ou ainda quando o número de atributos é maior do que de amostras e a regressão de mínimos quadrados ordinários gera coeficientes com erros padrão alto [9].

Outro método utilizado foi a regressão por componentes principais (PCR) que tem como principal objetivo eliminar a colinearidade etambém reduzir a quantidade de ruído inserida no modelo matemático final. Dessa forma, além de apresentar variáveis independentes que são ortogonais, em cada componente é possível descrever uma fração da variação total contida nos dados. Assim se torna mais fácil determinar quantas variáveis independentes serão incluídas na regressão [10].

Por fim desenvolveu-se uma regressão por vetor suporte. As máquinas de vetores suporte para regressão buscam encontrar um atributo que aproxime bem os dados e são construídas para encontrar uma função $h : X \rightarrow Y$ vai que aproximar bem os dados de amostra [11].

3 RESULTADOS

Do total de 11162 amostras disponíveis, para a modelagem fez-se a divisão da base original em 70% das amostras para treinamento e validação; e 30% para teste. As amostras destinadas ao treinamento e validação aplicaram-se a validação cruzada *k-fold* com 10 partições, método amplamente empregado para assegurar a capacidade de generalização dos modelos [12]. O método de erro médio quadrático foi utilizado para avaliar o desempenho de cada modelo durante a fase de treinamento e validação. Os resultados são apresentados na Tabela 2. Observa-se que os melhores resultados foram obtidos com a utilização dos métodos SIMPLS e PCR, ambos com erro médio quadrático de 10,16%. O método NIPALS obteve o pior resultado com RMSE de 15,58%.

Tabela 2. Média dos RMSEs - Fase de Treinamento e Validação

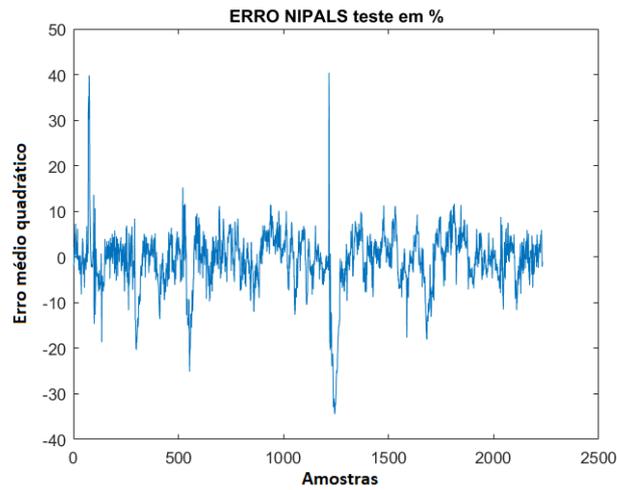
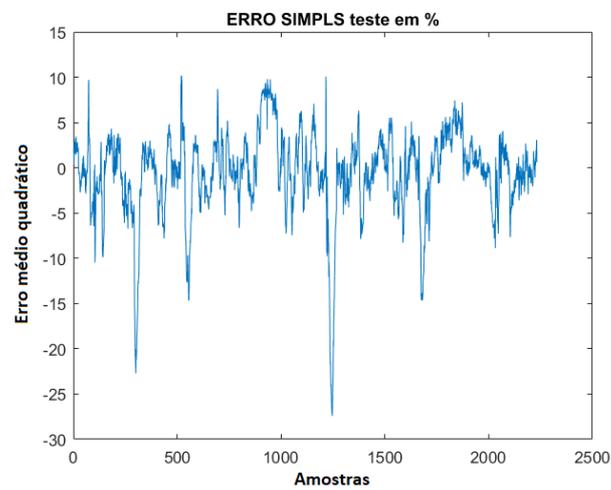
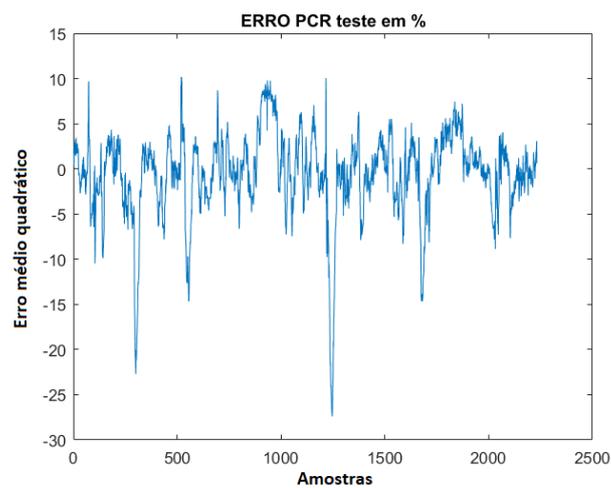
Modelo de Regressão	Erro médio Quadrático (%)
NIPALS	15,58
SIMPLS	10,16
PCR	10,16
SVR	10,35

Os resultados apresentados na Tabela 3 referem-se ao total de 30% das amostras destinadas ao teste dos modelos. Tais amostras não tiveram contato algum com a etapa de modelagem. Nota-se que também para os dois métodos SIMPLS e PCR foram obtidos os melhores resultados, com erro médio quadrático igual a 14,43. Também na fase de testes, o método NIPALS obteve o pior resultado com RMSE de 18,5%.

Tabela 3. Média dos RMSEs - Fase de Validação

Modelo de Regressão	Erro médio Quadrático (%)
NIPALS	18,5
SIMPLS	14,43
PCR	14,43
SVR	15,23

As Figuras 2, 3, 4 e 5 mostram, respectivamente, o gráfico do Erro Médio Quadrático dos modelos NIPLAS, SIMPLS, PCR e SVM durante a fase de teste.

**Figura 2.** Erro do modelo NIPALS**Figura 3.** Erro do modelo SIMPLS**Figura 4.** Erro do modelo PCR

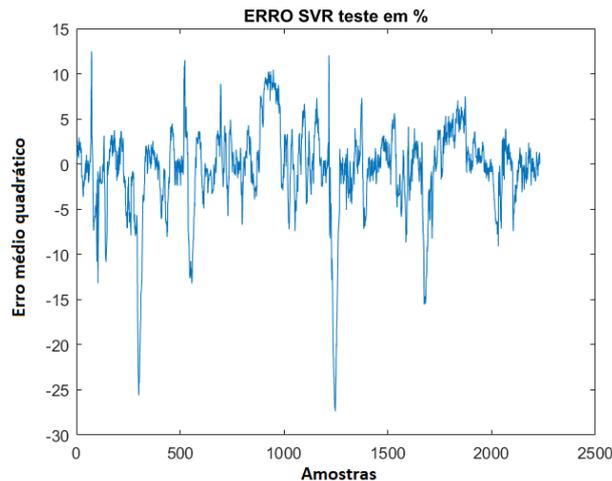


Figura 5. Erro do modelo SVR

A Figura 6 mostra a comparação do valor de índice de wobble estimado e real para o modelo de regressão SIMPLS. Observando as Tabelas 2 e 3 e as Figuras 2, 3, 4 e 5, pode-se verificar que os dois modelos de regressão SIMPLS, PCR tiveram os menores erros médios quadráticos tanto na fase de treinamento quanto na fase de teste. Podemos observar que todos os modelos mostraram um grande erro para amostras com índice de wobble acima de 3,4 e abaixo de 3,0, que são faixas onde o banco atual possui poucas amostras. Tal fato indica a necessidade de se criar um banco de dados mais balanceado.

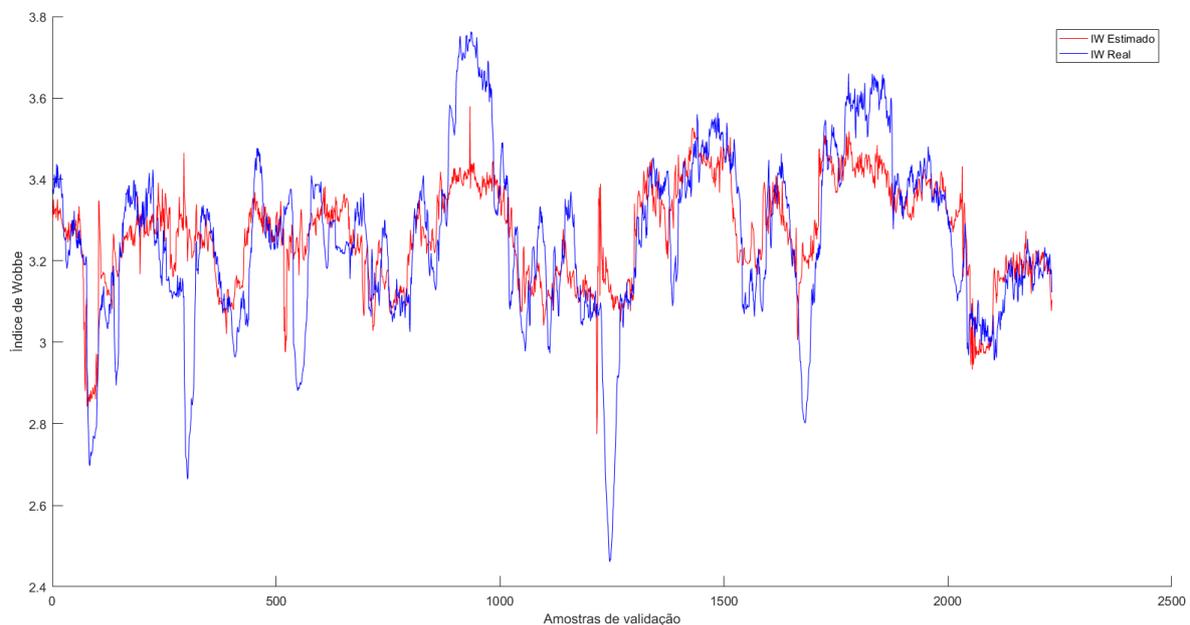


Figura 6. Comparação de Índice de Wobble real e estimado

4 CONCLUSÃO

Com o desenvolvimento deste trabalho ficou evidente a importância das primeiras etapas para o desenvolvimento de um sensor virtual. A etapa de coleta e filtragem de dados e também a seleção de variáveis se mostraram de suma importância para

garantir uma boa correlação do estimador com os dados reais. Para minimizar o erro médio quadrático se torna necessário uma boa análise inicial dos dados principalmente com a ajuda de especialistas do processo que possam mostrar as principais variáveis que estão relacionadas com o valor de saída desejado.

Apesar de apresentar um bom resultado quando analisado de forma ampla, para que este modelo possa ser aplicado no setor industrial, substituindo o medidor físico, busca-se um erro médio quadrático menor ou igual a 5%. Para se alcançar este resultado duas soluções podem ser propostas. A primeira delas é voltar na primeira etapa deste trabalho e com a ajuda do especialista encontrar novos atributos de entrada relacionados com o índice de wobbe. Outra opção é a construção de um banco de dados mais balanceado e representativo dentro da faixa de possíveis valores para o índice de wobbe do processo.

Por fim, ressalta-se que com a utilização de um preditor será possível informar com maior precisão a estimativa do índice de wobbe na indisponibilidade ou em caso de falha do instrumento analítico físico. Atualmente a prática é considerar que o Gás Misto (MG) possui um índice de wobbe de valor constante na indisponibilidade da medição desta variável.

REFERÊNCIAS

- 1 ABRAHÃO, I. N. Modelagem e otimização em sistemas de produção siderúrgica. 2000. 71 f. Dissertação de Mestrado – Universidade Federal do Espírito Santo, Vitória, 2000.
- 2 SOUZA, F. A. A.; ARAÚJO, R.; MENDES, J. Review of soft sensor methods for regression applications. *Chemometrics and Intelligent Laboratory Systems*, v. 152, p. 69-79, 2016.
- 3 PANI, A. K.; MOHANTA, H. K. Online monitoring of cement clinker quality using multivariate statistics and Takagi-Sugeno fuzzy-inference technique. *Control Engineering Practice*, v. 57, p. 1-17, 2016.
- 4 KADLEC, P.; GABRYS, B.; STRANDT, S. Data-driven Soft Sensors in the process industry. *Computers and Chemical Engineering*, v. 33, p. 795-814, 2009.
- 5 WU, Y.; LUO, X. A novel calibration approach of soft sensor based on multirate data fusion technology. *Journal of Process Control*, v. 20, p. 1252-1260, 2010.
- 6 HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning*, Springer Series in Statistics. Nova York, Springer, 2001.
- 7 PEARSON, R. K. Outliers in Process Modeling and Identification. *IEEE Transactions on Control Systems Technology*, v. 10, n. 1, p. 55-63, 2002.
- 8 SILVA, A. M.; LEONG, P. H. W. *Grammar-Based Feature Generation for Time-Series Prediction*. Sydney, Springer, 2015.
- 9 OLIVEIRA, C.L.; RESENDE, Z.C.; CAVALIERI, C.D. Prediction of phosphorus concentration in primary steelmaking. *Cilamce*, Santa Catarina, p. 5-8, 2017.
- 10 PERREIRA, T.F. *Utilização de Métodos Quimiométricos em dados de Natureza Multivariada*. 2003. Dissertação (Mestrado), Universidade Estadual de Campinas – UNICAMP. Campinas, São Paulo.
- 11 VERDÉRIO, A. *Sobre O Uso De Regressão Por Vetores Suporte Para A Construção De Modelos Em Um Método De Região De Confiança Sem Derivadas*. Universidade Federal do Paraná, Curitiba, p. 16-17, 2015.
- 12 Liang, T. and von Davier, A. A. (2014). Cross-validation: An alternative bandwidth selection method in kernel equating, *Applied Psychological Measurement* 38(4): 281-295.