Article

Forecasting Model of Silicon Content in Molten Iron Using Wavelet Decomposition and Artificial Neural Networks

Ana P. Miranda Diniz [1], Klaus Fabian Côco [1], Flávio S. Vitorino Gomes [2] and José L. Félix Salles [1,*]

[1] Electrical Engineering Department, Campus de Goiabeiras, Federal University of Espírito Santo, Vitória 29075-910, ES, Brazil; anapmdiniz@gmail.com (A.P.M.D.); klaus.coco@ufes.br (K.F.C.)
[2] Renewable Energy Engineering Department, Federal University of Paraíba, João Pessoa 58051-900, PB, Brazil; flavio@cear.ufpb.br
* Correspondence: jleandro@ele.ufes.br

Abstract: Silicon content forecasting models have been requested by the operational team to anticipate necessary actions during the blast furnace operation when producing molten iron, to control the quality of the product and reduce costs. This paper proposed a new algorithm to perform the silicon content time series up to 8 h ahead, immediately after the molten iron chemical analysis is delivered by the laboratory. Due to the delay of the laboratory when delivering the silicon content measurement, the proposed algorithm considers a minimum useful forecasting horizon of 3 h ahead. In a first step, it decomposes the silicon content time series into different subseries using the Maximal Overlap Discrete Wavelet Packet Transform (MODWPT). Next, all subseries forecasts were determined through Nonlinear Autoregressive (NAR) networks, and finally, these forecasts were summed to furnish the long-term forecast of silicon content. Using data from a real industry, we showed that the prediction error was within an acceptable range according to the blast furnace technical team.

Keywords: blast furnaces; silicon content; maximal overlap discrete wavelet packet; artificial neural network; forecasting; time series analysis

1. Introduction

Due to the higher competitiveness in the globalized market, industries are always seeking to reduce their operational costs without affecting the quality of the final product through control techniques of the relevant variables involved in the manufacturing process. The iron-making industry has played a significant role in the world economy, boosting the process of industrialization. This industry produces molten iron (or hot metal or pig iron) using blast furnaces and next directs it to be refined in the steel plant, resulting in steel. The stable and efficient operation of the blast furnace depends on the acquisition of a large amount of data to support operators in decision making through big data platforms and data-driven models [1].

The quality of the produced molten iron affects the cost of the production of the steel and defines for which ultimate purpose it can be sold [2]. However, the adverse conditions presented inside a blast furnace (high pressure, temperature, and erosive environment) may impair the direct measurements of some essential variables for quality control of the molten iron [3]. Thus, the laboratory analysis of some chemical variables of the molten iron samples is fundamental for monitoring the blast furnace operating state [4]. Notably, one of the most critical variables is the silicon content of the molten iron, due to its strong correlation with the thermal state of a blast furnace and the quality of the produced iron [2]. High values of the silicon content indicate an excess of coke in the blast furnace (i.e., a large amount of thermal energy), while lower values indicate the lack thereof. Thus, since the coke cost is prevalent in molten iron production, obvious economic benefits arise from more rigorous control of the silicon content [4–12]. In addition, better silicon control in

molten iron can reduce the CO_2 emissions to the atmosphere due to the steelmaking time decrease in the pig iron refining stage and other steps of the production process.

Over the past two decades, several papers have proposed neural network models with different inputs to predict and even control the silicon content of molten iron and other parameters related to the quality of the blast furnace process. Due to the complexity and a large number of exogenous variables involved in the molten iron production process, References [4–12] stressed the difficulty when choosing an appropriate algorithm that can establish effective the input–output relationships of neural networks. Consequently, we can observe in the literature a significant quantitative and qualitative discrepancy among the chosen inputs. All these literature works performed forecasts of silicon content with forecasting horizons lower than 3 h ahead and also disregarded the delay between the collection and laboratory analysis results of the silicon content, which is typical in any blast furnace [11]. The more we anticipate future information on the silicon content, the more effective the actions of the operators will be, aiming at increasing the quality of the final product, as well as reducing the CO_2 emissions and costs associated with the production of iron.

Initially, Reference [4] used a Multilayer Perceptron (MLP) without performing a careful analysis of the most relevant input variables and their respective time delays. Next, Reference [5] chose the inputs of a Nonlinear Input-Output (NIO) model through the blast furnace operator experience, and References [4–8] applied a pruning algorithm [13]. The work [9] considered a Nonlinear Autoregressive (NARX) neural network, including the silicon content as one of the network inputs. The authors of [10] eliminated the exogenous inputs, by decomposing the silicon content time series into different subseries from Empirical Mode Decomposition (EMD). After, these authors added the subseries forecasts through NAR neural network models. However, Reference [10] did not provide a numerical analysis of the forecasting quality.

Recently, References [11,12] used Extreme Learning Machine (ELM)-based networks with up to thee hidden layers containing about one-thousand neurons. Unlike backpropagation-based algorithms, ELM chooses the input and hidden layer weights randomly, while the output layer weights are determined analytically. Thus, ELM performs faster training than backpropagation [14]. However, due to the absence of weight adjustment in the ELM neural network, it requires many more neurons, which may increase its computational cost [15]. Therefore, because of the hardware limitation of the existing supervisory distributed computer control systems in the current steel industry, the ELM network implementation in this hardware is much more complicated than the neural network trained with backpropagation.

In this paper, we proposed a long-term forecasting algorithm with horizons from 3 h ahead, because the silicon content measurement in the molten iron is available for operators through a chemical laboratory analysis updated every three hours approximately after being collected. We considered a parsimonious neural network model trained by backpropagation, so that with a smaller number of parameters, it was possible to model and even forecast the silicon content with superior performance to those NIO and NARX neural networks. The proposed forecasting algorithm decomposes the silicon content time series into different subseries using Maximum Overlap Discrete Wavelet Packet Transformation (MODWPT) [16]. We modeled each subseries decomposed by MODWPT through a NAR neural network, and we performed the silicon content forecast by adding each subseries' forecasts.

MODWPT is the third generation of wavelet transform, and just as the Discrete Wavelet Transform (DWT) and Discrete Wavelet Packet Transform (DWPT), MODWPT also applies lowpass and highpass filters to the input signal at each level. However, MODWPT overcomes the disadvantages inherent to DWT and DWPT as it not only presents a uniform frequency bandwidth, but is also time-invariant. Besides, using MODWPT avoids edge effects at the end of the time series and allows the addition of new samples without the need to recalculate the entire transform again [16]. These properties are worthwhile

when dealing with real-time series applications focused on performing nonstationary time series forecasting. MODWPT decomposes the time series into stationary and nonlinear subseries through highpass and lowpass filters [17], characterized by frequency bands with different resolutions.

The work [18] applied the wavelet transform of the first generation in a neural model to reduce data noise and forecast the permeability index for blast furnaces. However, the analysis presented in [18] was entirely based on static data, i.e., the data were not processed sample-by-sample, which is inherent to real-time data processing. This inhibits the application of this model in industrial dynamic systems, whose forecasts must be processed for each sample, or a small set of samples collected and processed in real time, as we considered in this article.

There are some differences between the EMD decomposition technique used by [10] and MODWPT. EMD is an iterative procedure that extracts oscillatory-like features from the data in such a way that it uses adaptive bases that are directly generated by the data, i.e., EMD is a wavelet approach. EMD is an adaptive technique, and MODWPT is a math transform that lends itself to robust statistical analysis [19].

In Section 2, we briefly explain the molten iron production process; in Section 3, we apply the pruning algorithm to determine the neural network models with and without the MODWPT decomposition; in Section 4, using real data from a blast furnace located in Brazil, we compare the proposed algorithm, denoted by MODWPT-NAR, with the NAR, NARX, and NIO models without considering time series decomposition; finally, we conclude this article in Section 5.

2. The Blast Furnace Process

2.1. Background

The blast furnace is a big thermo-chemical reactor used for molten iron production, which is the main raw material for steel production. The operation of the blast furnace occurs following the countercurrent principle [2]. During the production process, synthetic ferrous materials (sinter and iron ore pellets) or natural raw materials (granulated iron ore) enter the upper part of the blast furnace, with one type of fuel (coke or charcoal). In addition, there may also be auxiliary injections of fuel oil, tar, pulverized coal, or natural gas. The objective is to produce an alloy, denominated molten iron, at a temperature of approximately 1500 °C, in the liquid state, composed of iron (between 92% and 95%), carbon (between 3.5% and 5%), and other elements such as silicon and manganese. In addition, slag and blast furnace gas are also generated as byproducts [2].

Molten iron is formed from the reduction of iron ore, while slag comes from the agglomeration of minerals and ashes resulting from burning fuels. Generally, it takes about six to eight hours for the raw materials to be processed in the form of molten iron and slag, which are accumulated at the base of the blast furnace (called blast furnace hearth). The molten iron remains on the bottom of the furnace hearth, while the slag remains on the upper part of it [2]. Ores and coke may contain many different kinds of impurities, and most of them are reduced in the blast furnace and eliminated in the form of slag. Some elements are fully incorporated into the molten iron (such as phosphorus and copper), while others are partially incorporated (such as sulfur, manganese, carbon, and silicon). From the point of view of thermal control, silicon is one of the most important elements [2].

The silicon in the blast furnace comes mainly from ores and coke ashes. The content of silicon depends not only on the quantity, but also on the types of raw materials used and on the method applied for producing the molten iron. The transfer of silicon to molten iron can occur either directly from slag to molten iron (through metal–slag reactions) or indirectly. For the latter, the silicon is transferred from slag (or coke ash) to the gaseous phase (SiO for example), which then reacts with molten iron–carbon, incorporating silicon [2].

The silicon content of the molten iron is an essential variable for the blast furnace process because it expresses not only the thermal state, but also the quality of it. The fluctuations of the silicon content reflect the thermal variations within the blast furnace.

This way, a low silicon content indicates that the furnace is cooling, which can compromise the operation, while high values indicate the waste of coke, shown by excessive heat generation. Thus, since the cost of coke is prevalent (usually more than 30% of the total amount) in molten iron production, the stricter the control of silicon content, the more significant the reduction in production costs is [6,8].

After passing through the blast furnace, the liquid molten iron is taken to the steelworks to transform it into steel. The thermal input of the molten iron is mainly responsible for the process heating supply. Besides, this material must show adequate physicochemical characteristics; otherwise, it is necessary to perform adjustments in the process, which can decrease productivity. Therefore, the high content of silicon can hinder the process of converting iron into steel and, consequently, generate even higher expenses. However, the presence of silicon in the steelworks is desirable, since it contributes to removing the oxygen present in the steel, through the oxidation of SiO_2 [2].

One of the major problems during the quality control of molten iron production is the long time spent from charging the raw material at the top of the blast furnace until the molten iron removal from the bottom of the blast furnace hearth [2]. Furthermore, in the real production process at hand, the laboratory analysis of the molten iron chemical composition takes about three hours to perform, as well as the silicon content amount to be available for the operators. Another problem is the nonregularity of the data collection periods of the silicon content, since the opening and closing times of the pig iron tapholes may vary according to the blast furnace production rate [20]. All these factors make it difficult to control the silicon content inside the blast furnace to reach adequate levels that guarantee the quality of the final product and the low cost of production.

2.2. The Dataset

In this study, we used data extracted from a blast furnace located in Brazil. This blast furnace has 4 tapholes, 34 tuyeres, 3617 m^3 of internal volume, and a nominal annual production capacity of 2.8 Mt. Typically, one taphole is opened at a time, although in some cases, it is necessary to open a second taphole to accelerate material extraction [3].

Regarding the blast furnace considered in this paper, twenty-eight variables may influence the silicon content. Such variables are described in Table 1.

It is noted that for each molten iron casting (time elapsed between an opening and a closing of the blast furnace taphole), three torpedo cars are filled, on average. A torpedo car is made with steel and coated with refractory, and its primary function is to transport molten iron from the blast furnace to the steelworks. Usually, the blast furnace operators collect a molten iron sample from each torpedo car loading for laboratory analysis. The result of such analysis, regarding the percentage of silicon, is then updated in the system. Because the acquisition time of the molten iron samples is not synchronous, presenting variations of around 50 min, it is crucial to resample it at regular times so that the neural network inputs can receive the past values of the silicon content time series. The molten iron temperature and electromotive force (input variables x_{27} and x_{28}) are also measured at each torpedo car loading and, therefore, also present an asynchronous sampling period, similar to the silicon content. The other input variables described in Table 1 have a sample period of 30 min. Thus, due to the interest of blast furnace operators in specifying the data at each 60 min interval, all input and output variables are resampled at this sample time.

Due to the harsh conditions of the blast furnace operation, process and measurement noises can contaminate the dataset. In this sense, noise reduction is required in data preprocessing [11]. All input and output data collected were processed to correct outliers and measurement errors, as well as to exclude significant periods of blast furnace shutdowns. In general, an abrupt decrease in the blast furnace production rate indicates a blast furnace shutdown (i.e., a period of strong process instability), and often, these moments are followed by outliers in the silicon content time series. Figure 1 shows an example of a plant shutdown identified during the analysis of the corresponding data. As the objective of this

work was to forecast the silicon content during regular operational periods, the samples corresponding to the abnormalities must be detected [21] and excluded from the dataset.

Table 1. Input variables that may influence the silicon content.

Variable	Description
x_1	flame temperature (°C)
x_2	rate of CO/CO_2
x_3	H_2 gas efficiency (%)
x_4	molten iron production (tons/min)
x_5	coke rate (Kg/ton of molten iron)
x_6	pulverized coal injection rate (Kg/ton of molten iron)
x_7	direct reduction (%)
x_8	ore/coke ratio
x_9	thermal index of H0 (10^3 Kcal/ton of molten iron)
x_{10}	north central gaseous flow index
x_{11}	north peripheral gaseous flow index
x_{12}	south central gaseous flow index
x_{13}	south peripheral gaseous flow index
x_{14}	blowing air volume (Nm^3/min)
x_{15}	blowing air moisture (g/Nm^3)
x_{16}	blowing air temperature (°C)
x_{17}	blowing air pressure (Kg/cm^2)
x_{18}	top pressure (Kg/cm^2)
x_{19}	top temperature (°C)
x_{20}	composition of H_2 on the top gas (%)
x_{21}	composition of N_2 on the top gas (%)
x_{22}	composition of CO on the top gas (%)
x_{23}	composition of CO_2 on the top gas (%)
x_{24}	O_2 enrichment rate (%)
x_{25}	CO efficiency (%)
x_{26}	staves' thermal losses (10^4 Kcal/h)
x_{27}	molten iron temperature
x_{28}	electromotive force (%)
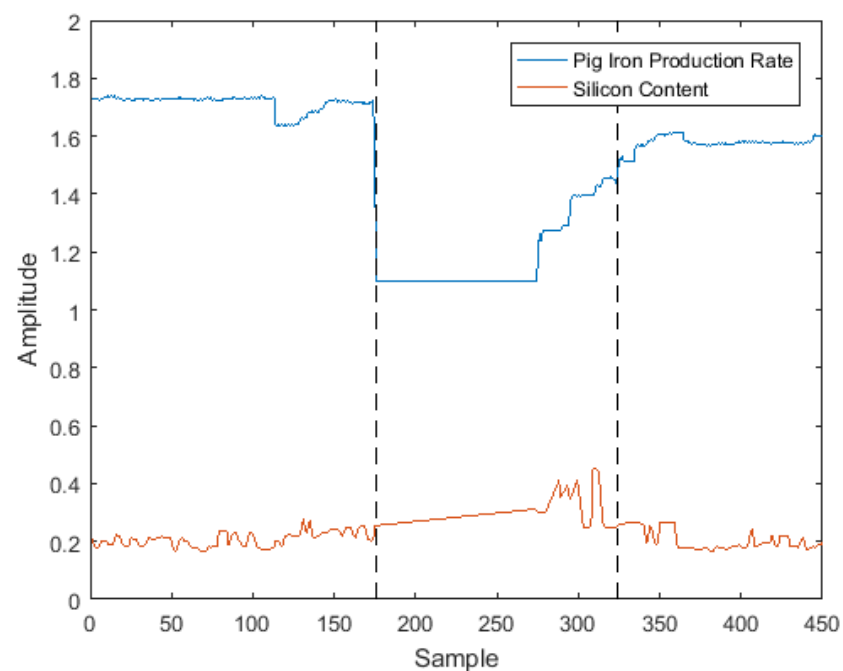
Figure 1. Abnormality period in the blast furnace.

After removing the data corresponding to instability periods in the blast furnace, we performed the stationarity analysis of the silicon content time series. First, we detected the nonstationary characteristic of this time series through the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test [22]. Then, we used the Lilliefors [23] and Kruskal–Wallis [24] tests to

confirm the existence of nonlinearities by splitting the time series into five packages of data; see Figure 2. Each one of the packages was composed of a 10-day measurement, to show that these samples were not from the same distribution.
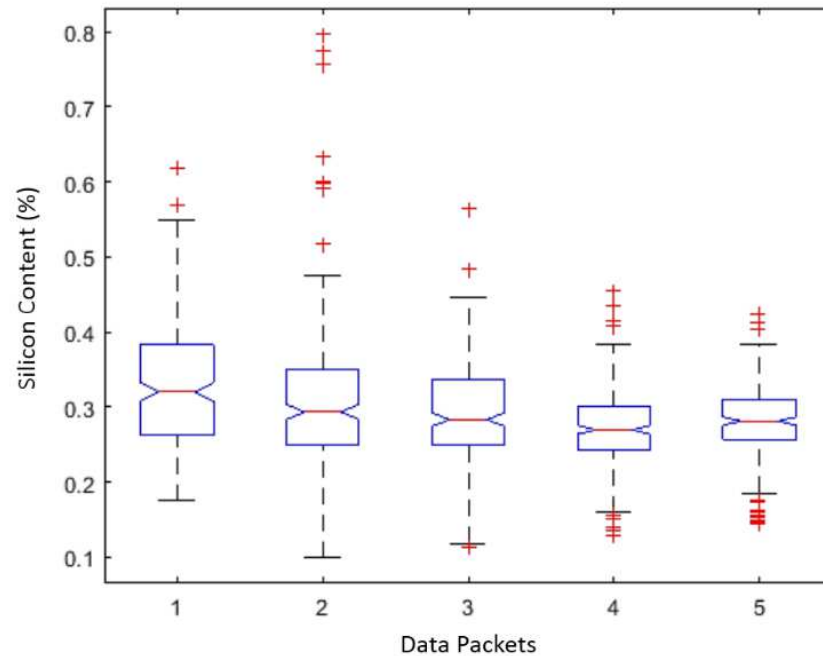

Figure 2. Box plot test of silicon content time series.

Based on these analyses, we concluded that the time series had nonlinear components due to the asymmetric dispersion concerning the normal distribution, as shown in Figure 3.
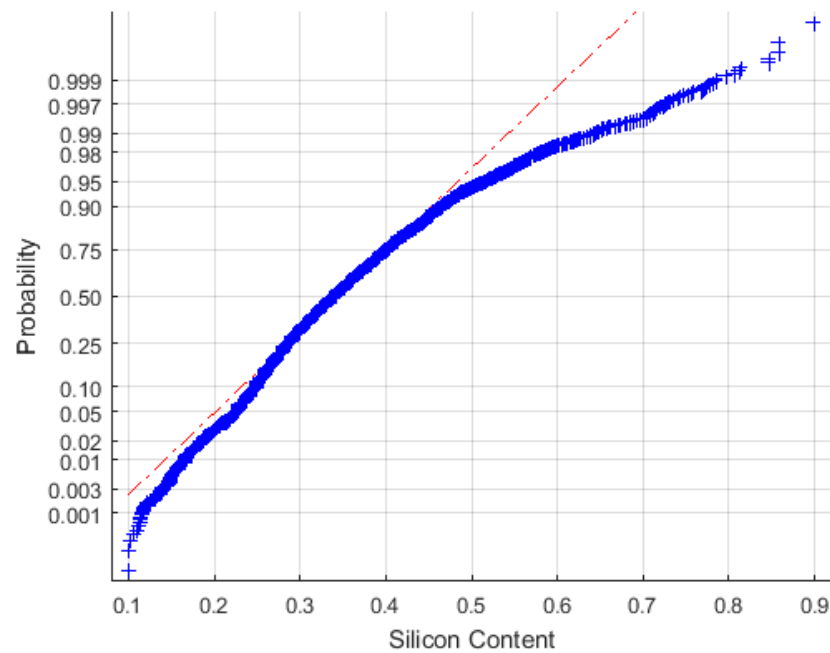

Figure 3. Q-Q plot of silicon content time series.

3. Silicon Content Time Series Modeling

Firstly, we present the ANN models without considering the decomposition of silicon content using MODWPT.

3.1. NIO, NARX, and NAR Neural Network Models

Inspired by the architecture and learning of the human brain, ANNs exhibit self-adaptive behavior and are very favorable to model and forecast nonlinear processes that have complex characteristics, in contrast to traditional deterministic methods. They work as processors made up of simple processing units—the neurons—that can store the knowledge acquired and can also adapt to a changing environment through a learning process. The links that connect neurons to carry out signal transmission are called "synapses," over which there is a corresponding weight. Each neuron receives multiple inputs from other neurons, computes them, and generates a single output, which can be propagated to several other neurons [25].

In this paper, we addressed the Time-Delay Neural Networks (TDNN). Such networks are extensions of Multilayer Perceptron (MLP) networks, with the introduction of one unit delay in the input variable, aiming at representing the dynamics (cause and effect relationships) of the nonlinear model [26]. Examples of TDNN are the NIO, NARX, and NAR neural networks. Their outputs at k-steps ahead ($\hat{y}(t+k)$) are described by (1)–(3), respectively, where n_a and n_b are the maximum delays of the output $y(t)$ (silicon content) and inputs $x(t) = (x_1(t), x_2(t), \ldots, x_{28}(t))$ (defined in Table 1), respectively; t is the moment immediately after the molten iron chemical analysis is delivered by the laboratory. $f()$ represents a nonlinear function that is parameterized through the synaptic weights calculated during the neural network training phase [25,26].

$$\hat{y}(t+k) = f(x(t), \ldots, x(t-n_b)) \tag{1}$$

$$\hat{y}(t+k) = f(y(t), \ldots, y(t-n_a), x(t), \ldots, x(t-n_b)) \tag{2}$$

$$\hat{y}(t+k) = f(y(t), \ldots, y(t-n_a)) \tag{3}$$

The NIO, NARX, and NAR neural network models were defined by the input and output sets $\{x(t-n), \ldots, x(t)\}$ and $\{y(t-n), \ldots, y(t)\}$, respectively, where n is the total of historical data samples. We split both datasets into two blocks. Regarding the first block, each time series had 2160 samples (90 days) used for training, and as for the second block, each time series had 720 samples (30 days), used for validation. Each time series was normalized within the range $[0.1, 0.9]$. Normalization was important to ensure non-zero values in the time series, thus allowing us to calculate the performance indices used to compare the neural network models. Furthermore, the normalization technique can provide a reduction of errors in the training of the neural networks [25]. High-frequency noises present in the input time series were removed by applying wavelet filters of the symlet type with order four [27].

The criteria used to define the ANN models were the k-steps ahead Mean Squared Error (MSE(k)) and k-steps ahead Mean Absolute Percentage Error (MAPE(k)), defined according to (4) and (5), respectively:

$$\text{MSE}(k) = \frac{1}{n} \sum_{t=1}^{n} (y(t+k) - \hat{y}(t+k))^2, \tag{4}$$

$$\text{MAPE}(k) = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y(t+k) - \hat{y}(t+k)}{y(t+k)} \right|, \tag{5}$$

We considered a NAR neural network topology with six inputs because the leaks (casting) through the tapholes occurred at a frequency of between eight- and fourteen-times a day, which corresponded to a leakage period of 1.5 to 3 h. Besides, the residence time of the load inside the blast furnace was 6 h approximately; therefore, the current sample of the silicon content depended on the samples collected up to 6 h before, i.e., up to six lags considering a sample time of one hour. For that reason, the NAR neural network must have six entries, which corresponded to the current sample plus five previous samples.

Regarding the NIO and NARX neural networks, the delays of the input variables $x_1 \ldots, x_{28}$ were, initially, equal to their respective average response times, i.e., the average times the production process responded to the variations of each input. In this sense, except the coke (x_5) and ore/coke (x_8) rates, which presented a slower response time (8 h), all input variables of the NARX network had a delay of three samples (3 h).

The NIO, NARX, and NAR neural networks have a hyperbolic tangent-type activation function for the hidden layer neurons and a linear activation function for the output layer neuron. We used a single hidden layer on each of the networks, which were defined by testing structures with five to two-hundred hidden neurons, adding five neurons per test. For each of these networks, we determined the MSE(0) errors after five consecutive training rounds performed by using the Levenberg–Marquardt learning algorithm. From these experiments, we concluded that the hidden layers for the NIO, NARX, and NAR neural models with 95, 90, and 100 neurons, respectively, provided the lowest MSE(0) errors in the training phase.

Due to a large number of variables involved in the process, it was challenging to choose an appropriate algorithm that could select the most relevant inputs and, consequently, to establish input–output relations effectively [6]. To remove redundant connections between the neurons of the input and hidden layers, without harming the model performance, we applied a pruning algorithm based on Hessian matrix, using the Optimal Brain Surgeon method (OBS) [28,29]. The pruning algorithm allowed removing regressors (input variables) that were not significant for the model, thus facilitating the choice of relevant variables. We verified that after applying the pruning algorithm, the number of hidden neurons for the NIO, NARX, and NAR neural networks reduced from 95 to 3, 90 to 6, and 100 to 13, respectively. The pruning algorithm also detected the significant regressors to be considered in the input layer of each neural network, as shown in Table 2. Each of the hidden layer neurons of these final ANN models was associated with more than one regressor, although the total number of regressors decreased significantly concerning the ANN models initially proposed. In Section 4, we compare the performances of the NIO, NARX, and NAR neural networks before and after applying the pruning algorithm.

Table 2. Significant regressors in the input layers of neural networks.

ANN Model	Regressors
NIO	$x_5(t-5)$, $x_{24}(t-1)$, $x_{27}(t-2)$
NARX	$x_2(t-2)$, $x_3(t-3)$, $x_4(t-3)$, $x_5(t-k)$ for $k=1,2,3$ $x_7(t-3)$, $x_8(t-1)$, $x_{10}(t-k)$ for $k=1,2$, $x_{12}(t-2)$, $x_{14}(t-1)$, $x_{15}(t-1)$, $x_{19}(t-2)$, $x_{23}(t-2)$, $x_{24}(t-3)$, $x_{25}(t-2)$, $x_{27}(t-k)$ for $k=1,2$, $x_{28}(t-1)$, $y(t-k)$ for $k=1,2,3,4,5$.
NAR	$y(t-k)$ for $k=1,2,3,4,5$.

3.2. MODWPT-NAR Neural Network Model

To increase the accuracy of prediction models using neural networks, we developed a hybrid algorithm based on MODWPT decomposition and ANN models to forecast the content of silicon in molten iron (see Figure 4). Firstly, the silicon content time series was decomposed into three levels through the MODWPT, using a four-order symlet mother wavelet. As a result, we obtain eight data subsets $\{X_{WP1}\}, \{X_{WP2}\}, \ldots, \{X_{WP8}\}$), which are illustrated in Figure 5.
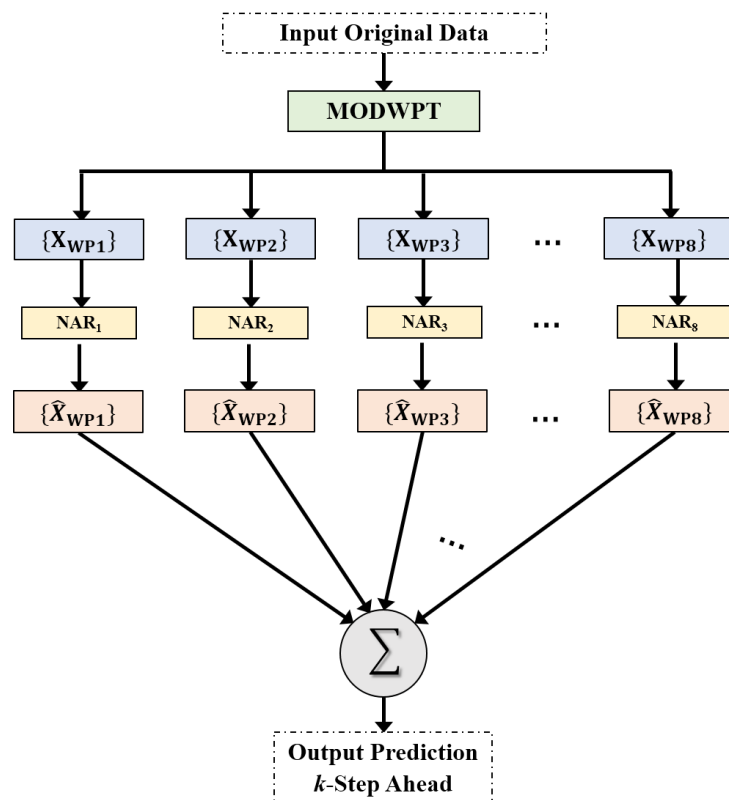
Figure 4. MODWPT-NAR neural network model.

After that, we used eight NAR networks to perform multistep forecasts of each one of the eight subseries. Finally, we added the subseries forecasts to obtain the original signal forecasting.

The Discrete Wavelet Transform (DWT) is a time and frequency multiresolution analysis technique that decomposes the original signal into several levels. According to [27], multiresolution analysis is a useful tool since many signals found in practical applications have high-frequency components for short time durations and low-frequency components for a long duration of time.

The implementation of DWT is an iterative process with successive decompositions of a signal, which uses two sets of functions called scale functions and displacement functions, associated with the lowpass and highpass filters of the wavelets. Employing convolution, each one of the lowpass and highpass filters yields signals in the approximation and detail components, respectively, of the original signal [27].

A single function $\Psi(t)$, called the mother wavelet, generated the wavelet functions. Mother wavelets can take many forms (Daubechies wavelet (Daubechies), Daubechies' least-asymmetric wavelet (symlet), and coiflet [27]) and, in general, are defined by [16]:

$$\Psi_{j,k}(t) = a_0^{-j/2}\Psi(a_0^{-j}t - kb_0) \tag{6}$$

where j and k are the expansion and translation coefficients, respectively, and a_0 and b_0 give the expansion variation and the translation step, respectively. We usually assumed binary expansion and unit translations, i.e., $a_0 = 1$ and $b_0 = 2$, resulting in the following expression:

$$\Psi_{j,k} = 2^{-j/2}\Psi(2^{-j}t - k) \tag{7}$$
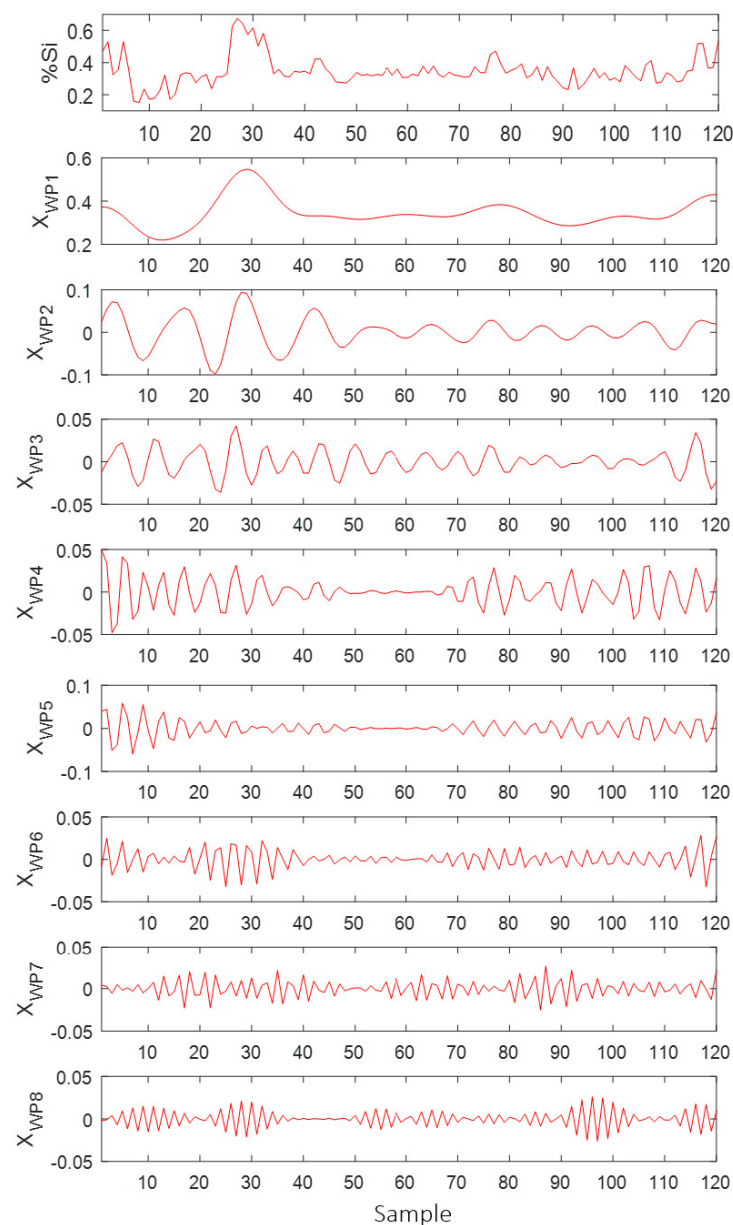
Figure 5. MODWT decomposition of silicon content time series.

However, at each DWT iteration, only the elements were divided into new approximation and detail components. In contrast to DWT, the Discrete Wavelet Packet Transform (DWPT) allows detail elements to be also divided into new approximation and detail components, resulting in uniform frequency bands. In addition, in the DWT, the approximation coefficients identify the frequency range $[0, \frac{f_s}{2^{(j+1)}}]$, while the detail coefficients at level j describe the frequency band $[\frac{f_s}{2^{(j+1)}}, \frac{f_s}{2^j}]$, where f_s represents the sampling frequency of the original signal. On the other hand, the DWPT allows that, at each level j, the frequency band is divided into 2^j partitions of equal length [16].

In the DWT and DWPT approaches, the signal length is limited to the power of two integers. Let L be the maximum decomposition level, int() the function that returns the nearest integer, and n the length of the input signal. We defined the maximum decomposition level by [27]:

$$L = \text{int}(log_2(n)) \tag{8}$$

In contrast to DWT and DWPT, the Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) is invariant after adding new samples in the time series. It has the same

number of coefficients at each level of decomposition since there is no signal decimation at each level. The MODWPT has often been used as an analytical tool to represent signals in the analysis of nonstationary time series, by dividing the variance of the data by the scale [27]. The main characteristic of such a transform is the lack of sensitivity to the time series starting point; consequently, the temporal analysis does not depend on the instant of time we take the samples [16].

The MODWPT is a function that allows the signal representation in frequency bands with different resolutions [16]. As the DWPT, it divides the frequency band into 2^j partitions of equal length for each level j [16].

Another feature of subseries resulting from the decomposition process via MODWPT is that the original signal is perfectly reconstructed since the decomposition coefficients pass through lowpass and reverse-highpass filters [16]. The MODWPT overcomes the disadvantages of the DWT and DWPT, i.e., it not only has uniform frequency bandwidth, but it is time-invariant. Furthermore, since the MODWPT components are nonorthogonal, there is a slight overlap of frequency bands enabling solving the edge effect problem found in time series forecasting. These properties allow the addition of new samples in the time series without the need to recalculate the entire transform again [16,17].

In the proposed hybrid MODWPT-NAR algorithm, the decomposed signals were modeled by the $NAR_1, \ldots, NAR_8$ neural networks. We performed two experiments to select the number of neurons in the hidden layer of each NAR model by using the learning algorithm Levenberg–Marquardt [25]. The first considered the same number of neurons in the hidden layer of all NAR_1 to NAR_8 networks. We varied the hidden neurons from one to thirty for all these networks, simultaneously, and we obtained 30 MODWPT-NAR models. The MODWPT-NAR model with seven neurons in the hidden layer of all NAR_1 to NAR_8 networks had the lowest training error, and its MSE(0) index was equal to 5.32×10^{-5}.

Regarding the second experiment, we considered different numbers of neurons in the hidden layer of each NAR_1 to NAR_8 network within the range of one to thirty neurons. This was necessary because we determined the number of neurons of the hidden layer of each neural network according to the complexity of the decomposed subseries. Next, we found the number of neurons in the hidden layer of each NAR_1 to NAR_8 network that provided the lowest MSE(0) error in the training phase, as shown in Table 3. We calculated this error from each decomposed subseries and corresponding forecast, without considering the original time series.

Table 3. Hidden layer size of networks for the second experiment.

Neural Network	Hidden Layer Size	MSE(0) ($\times 10^{-4}$)
NAR_1	27	0.0169
NAR_2	8	0.2642
NAR_3	7	0.1299
NAR_4	19	0.0487
NAR_5	21	0.0500
NAR_6	25	0.0442
NAR_7	23	0.0708
NAR_8	14	0.0047

The training error obtained from the MODWPT-NAR model after the second experiment had an MSE(0) index equal to 6.55×10^{-5}. Therefore, we chose the MODWPT-NAR model obtained from the first experiment as the final hybrid model, because it had a smaller hidden layer size for all NAR_1 to NAR_8 networks and provided a smaller MSE(0) training error.

4. Comparative Analysis of Forecasting Models

In this section, we compare the silicon content forecasting performances obtained from the NAR, NIO, and NARX neural networks and the MODWPT-NAR model. All numerical results were performed on a desktop PC i3-3110M, 2.4-GHZ CPU, and 4G RAM with the MATLAB (2018a) software, using real data obtained from a steel industry located in Brazil. We used the toolboxes provided by [29] to implement the neural networks and perform the pruning algorithm.

In addition to the MSE and MAPE indices, another important criterion used to compare the silicon forecasting models is the absolute error percentage of the k-steps ahead forecasting lower than ϵ ($\text{AEP}_\epsilon(k)$), given by [9]:

$$\text{AEP}_\epsilon(k) = \frac{1}{n} \sum_{t=1}^{n} N_E \times 100 \tag{9}$$

where

$$N_E = \begin{cases} 1 & \text{if} \quad |y(t+k) - \hat{y}(t+k)| < \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

The $\text{AEP}_\epsilon(k)$ index is defined according to the blast furnace specialists' need. In this paper, we considered $\epsilon = 0.05$, i.e., the forecasting absolute errors at k-steps ahead were acceptable when they were lower than this range. Therefore, this index is better when the resulting value is higher.

To better understand the forecasting results presented in this section, let us define by t the time when the last molten iron chemical analysis is delivered by the laboratory. Since silicon content measurements were updated every 3 h and the molten iron samples were collected at intervals of 1 h, the collection time of the current molten iron sample was $t+3$. Therefore, the silicon content forecast at this time was $\hat{y}(t+3)$, which was obtained through neural models using the data stored until time t. In addition, the future silicon content samples would be collected at times $t+k, k = 4, 5, \dots$, and the respective forecasts would be $\hat{y}(t+k)$. In fact, to control the silicon content in the blast furnace and steelmaking plant, only forecasting horizons longer than or equal to 3 h ahead starting from instant t were adequate; due to the great inertia of the blast furnace, and the time spent in the steelmaking shop floor operations, the longer the forecasting horizon was, the better actions could be obtained in correcting the silicon content.

Tables 4 and 5 show the forecasting performances obtained by the NIO, NARX, and NAR networks before and after implementing the pruning algorithm.

From Tables 4 and 5, one can observe a significant improvement of the networks with exogenous inputs (NIO and NARX), in addition to a considerable decrease in the size of the hidden layer. For example, for the NARX network before applying the pruning algorithm, the number of hidden neurons was 90, the MSE(3) value 13.42×10^{-3} the MAPE(3) value 32.15%, and the $\text{AEP}_\epsilon(3)$ value 33.80%, for a three-hour-ahead forecasting horizon. In turn, after applying the pruning algorithm, the $\text{AEP}_\epsilon(3)$ value became 56.33% and the MAPE(3) value 17.98%, by considering only six neurons for the hidden layer.

Table 4. Performance of NIO, NARX, and NAR models before implementing the pruning algorithm.

Model	Horizon (*k*)	MSE(*k*) (10^{-3})	MAPE(*k*) (%)	AEP$_\epsilon$(*k*) (%)
NIO	3	21.15	38.85	27.32
	4	20.54	38.28	30.60
	5	20.29	38.05	30.90
	6	20.48	38.25	28.14
NARX	3	13.42	32.15	33.80
	4	11.65	29.55	36.53
	5	12.37	30.42	34.74
	6	11.87	29.04	36.63
NAR	3	3.05	12.88	73.73
	4	4.40	15.07	66.94
	5	9.07	21.76	53.80
	6	15.17	29.50	40.48

Table 5. Performance of NIO, NARX, and NAR models after implementing the pruning algorithm.

Model	Horizon (*k*)	MSE(*k*) (10^{-3})	MAPE(*k*) (%)	AEP$_\epsilon$(*k*) (%)
NIO	3	5.24	18.76	58.30
	4	5.28	18.76	58.53
	5	5.38	18.74	60.02
	6	5.45	18.76	59.83
NARX	3	5.31	17.98	56.33
	4	5.67	18.82	54.58
	5	6.04	19.55	51.41
	6	6.40	20.13	49.50
NAR	3	3.36	12.29	76.26
	4	5.88	16.49	66.24
	5	15.31	26.26	52.53
	6	21.52	31.76	40.76

We observed the same effect for the NIO network, which previously had 95 hidden neurons and presented an MSE(3) of 21.15×10^{-3}, a MAPE(3) of 38.85%, and an AEP$_\epsilon$(3) of 27.323%, for a forecasting horizon of 3 h ahead. After applying the pruning algorithm, on the other hand, this network could perform forecasts with superior performance (MSE(3) of 5.24×10^{-3}, MAPE(3) of 18.76%, and AEP$_\epsilon$(3) of 58.309%) using only three hidden neurons. Figure 6 illustrates the six-hour-ahead forecasts of silicon content obtained from the NIO and NARX models.
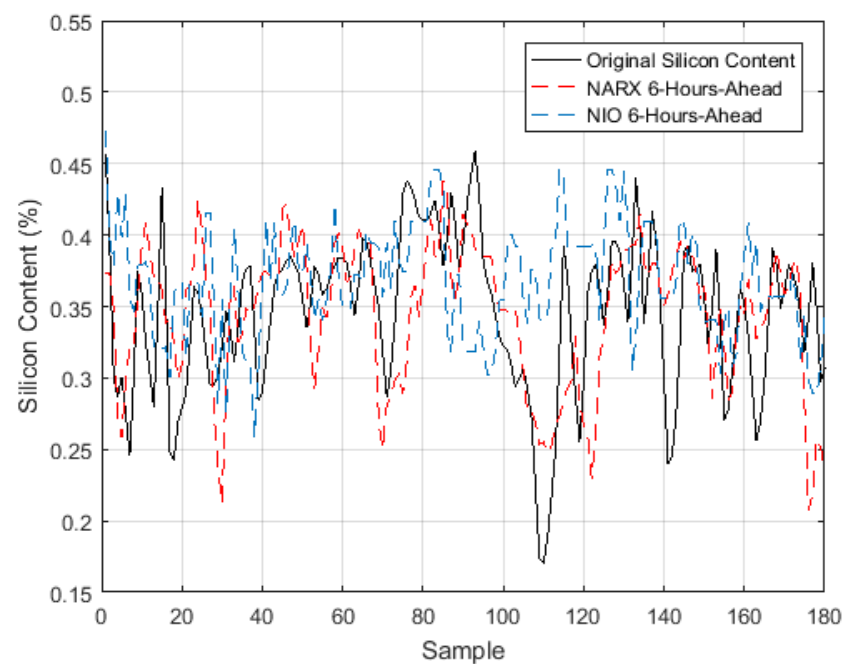
Figure 6. Silicon content and its six-hour-ahead forecasts from the NIO and NARX models.

The most significant effect brought by the pruning algorithm for the NAR network was that the size of the hidden layer was reduced from 100 neurons to 13 hidden neurons, without harming the performance. Notice in Table 5 the improvements in NIO and NARX networks due to the use of the pruning algorithm. Besides, observe in Table 5 that the NAR network still performed the best in forecasting for horizons until four steps ahead, and for longer horizons, the performance of NAR was worse than NIO and NARX. Thus, it was relevant to use the previous values of silicon content as the ANN input. However, the nonstationary behavior of this time series could impair the learning ability of the NAR and NARX networks for a longer forecasting horizon.

In Table 6, observe the performance of the MODWPT-NAR forecasts for the same data considered in Table 5.

Table 6. Performance of the MODWPT-NAR model.

Horizon (k)	MSE(k) (10^{-3})	MAPE(k) (%)	AEP_ϵ (%)
3	0.043	1.58	100
4	0.14	2.82	99.72
5	0.22	3.54	99.02
6	0.30	4.21	98.18
7	0.33	4.54	97.90
8	0.83	6.57	93.70

We noticed that this hybrid algorithm presented superior performance to the NAR model. As an example, the 3-h-ahead forecasting of the MODWPT-NAR model reached a MAPE(3) of 1.58%, while for the NAR model, this index was 12.29%. Furthermore, when the forecasting horizon increased, notice from Table 6 that the performance of the hybrid model did not decrease as much as in the model that used only the NAR neural network without considering MOPWPT decomposition. However, the proposed forecasting model became inaccurate for horizons longer than 8 h ahead, because the AEP index remained below 90%.

We also verified from Tables 5 and 6 that 40.76% of the six-hour-ahead forecasts obtained from the NAR model were inside the tolerance range, against the 98.18% obtained by the MODWPT-NAR hybrid model. We justified this result by the fact that the operators

frequently take actions on the blast furnace input variables, aiming to correct instabilities to keep the silicon content within an acceptable range. These actions may increase the nonstationary behavior of the silicon content time series and hamper the learning ability of the NAR network when compared to the proposed MOPWT-NAR hybrid model. This is because MODWPT allows the decomposition of the nonstationary signal into stationary subseries. Figure 7 illustrates the six-hour-ahead forecasts of silicon content obtained from the NAR and MODWPT-NAR models.
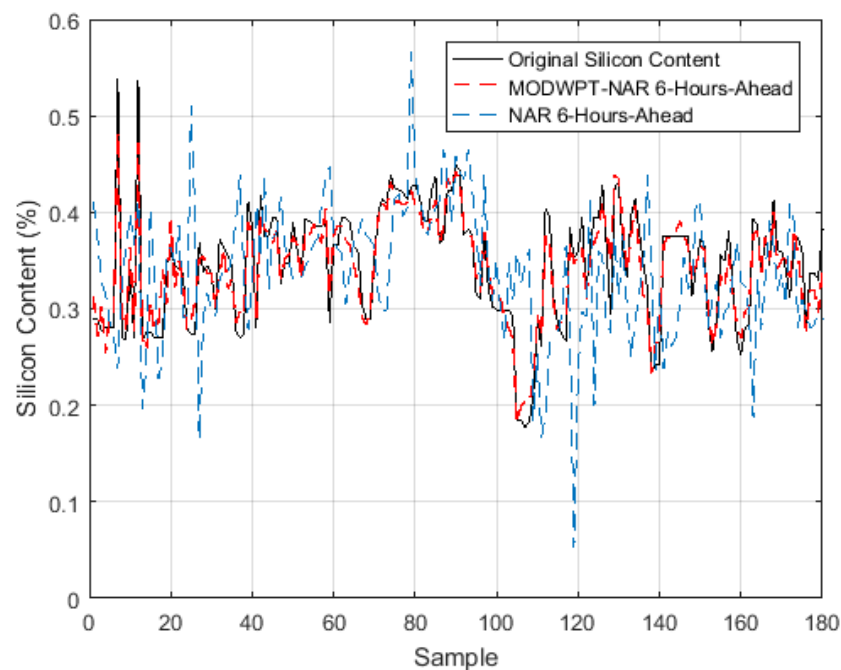
Figure 7. Silicon content and its six-hour-ahead forecasts from the NAR and MODWT-NAR models.

5. Conclusions

In this paper, we proposed a long-term forecasting model of silicon content, which performed the decomposition of the time series into additive components using MODWPT. We used NAR neural networks to model each decomposed signal, and the long-term silicon content forecast was performed by adding each subseries forecasting. In this case study, ninety-three-point-seven percent of the 8-h-ahead forecasting errors determined by MOPDWT-NAR model were within the acceptable range of 0.05%, while at most 60% of the 6-h-ahead forecasting errors furnished by the ARX, NIO, and NAR neural network models without decomposition were within this acceptable range. Therefore, with this hybrid model, we obtained very accurate forecasts with a horizon of up to 8 h ahead from the moment the last molten iron chemical analysis was delivered by the laboratory.

The proposed model proved to be a promising tool for the prediction of pig iron silicon content, and we can extend it to other blast furnaces, with appropriate changes. To implement this forecasting system on industrial plants, an operating station accessing the process database to collect input data and send the forecasts in real time to the supervisory system is necessary. In the control room, the operator may observe the long-term silicon content forecasts while performing blast furnace operational changes. Thus, such a forecasting system can optimize the blast furnace raw materials, reduce the operational costs, and increase the quality of the pig iron. It also allows optimizing the steelmaking logistics since, currently, the technical team of the steelmaking plant receives the silicon measurements 3 h after the molten iron sample's collection. In this way, the anticipation of silicon content information can decrease the steelmaking time and, consequently, reduce the production costs and the CO_2 emissions to the atmosphere.

Conflicts of Interest: The authors declare no conflict of interest.The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Li, H.; Bu, X.; Liu, X.; Li, X.; Li, H.; Liu, F.; Lyu, Q. Evaluation and Prediction of Blast Furnace Status Based on Big Data Platform of Ironmaking and Data Mining. *ISIJ Int.* **2021**, *61*, 108–118. [CrossRef]
2. Geerdes, M.D.; Toxopeus, H.; Vliet, C.; Young, G.O. *Modern Blast Furnace Ironmaking: An Introduction*; IOS Press BV: Amsterdam, The Netherlands, 2009.
3. Gomes, F.S.V.; Coco, K.F.; Salles, J.L.F. Multistep forecasting models of the liquid level in a blast furnace hearth. *IEEE Trans. Autom. Sci.* **2017**, *14*, 1286–1296. [CrossRef]
4. Singh, H.; Sridhar, N.V.; Deo, B. Artificial neural nets for prediction of silicon content of blast furnace hot metal.*Steel Res. Int.* **1996**, *67*, 521–527. [CrossRef]
5. Chen, J. A predictive system for blast furnaces by integrating a neural network with qualitative analysis. *Eng. Appl. Artif. Intell.* **2001**, *14*, 77–85. [CrossRef]
6. Saxén, H.; Pettersson, F. Nonlinear prediction of the hot metal silicon content in the blast furnace. *ISIJ Int.* **2007**, *47*, 1732–1737. [CrossRef]
7. Nurkkala, A.; Pettersson, F.; Saxén, H. Non linear modeling method applied to prediction of hot metal silicon in the ironmaking blast furnace. *Ind. Eng. Chem. Res.* **2011**, *50*, 9236–9248. [CrossRef]
8. Saxén, H.; Gao, C.; Gao, Z. Data-Driven Time Discrete Models for Dynamic Prediction of the Hot Metal Silicon Content in the Blast Furnace—A Review. *IEEE Trans. Ind. Inform.* **2013**, *9*, 2213–2225. [CrossRef]
9. Chen, W.; Wang, B.X.; Han, H.L. Prediction and control for silicon content in pig iron of blast furnace by integrating artificial neural network with genetic algorithm. *Ironmak. Steelmak.* **2010**, *37*, 458–463. [CrossRef]
10. Wang, H.; Yang, G.; Pan, C.; Gong, Q. Prediction of hot metal silicon content in blast furnace based on EMD and DNN. In Proceedings of the 34th Chinese Control Conference, Hangzhou, China, 28–30 July 2015; pp. 8214–8218.
11. Zhang, H.; Zhang, S.; Yin, Y.; Chen, X. Prediction of the hot metal silicon content in blast furnace based on extreme learning machine. *Int. J. Mach. Learn. Cybern.* **2017**,*9*, 697–1706. [CrossRef]
12. Su, X.; Zhang, S.; Yin, Y.; Hui, Y.; Xiao, W. Prediction of hot metal silicon content for blast furnace based on multi-layer online sequential extreme learning machine. In Proceedings of the 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; pp. 8025–8030.
13. Saxén, H.; Pettersson, F. Method for the selection of inputs and structure of feedforward neural networks. *Comput. Chem. Eng.* **2006**, *30*, 1038–1045. [CrossRef]
14. Cao, J.; Lin, Z.; Huang, G.B. Self-adaptive evolutionary extreme learning machine. *Neural Process. Lett.* **2012**, *36*, 285–305. [CrossRef]
15. Chazal, P.; Tapson, J.; Schaik, A.V. A comparison of extreme learning machines and backpropagation trained feed-forward networks processing the mnist database. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 19–24 April 2015; pp. 2165–2168.
16. Percival, D.B.; Walden, A.T. *Wavelet Methods for Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2000.
17. Walden, A.T.; Cristan, A.C. The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events. *Proc. R. Soc. Lond.* **1998**, *454*, 2243–2266. [CrossRef]
18. Su, X.; Zhen, S.; Yin, Y.; Xiao, W. Prediction model of permeability index for blast furnace based on the improved multi-layer extreme learn machine and wavelet transform. *J. Frankl. Inst.* **2018**, *355*, 1663–1691. [CrossRef]
19. Labate, D.; Foresta, F.L.; Occhiuto, G.; Morabito, F.C.; Lay-Ekuakille, A.; Vergallo, P. EMD vs. DWT for the extraction of respiratory signal from single-channel ECG: A comparison. *IEEE Sens. J.* **2013**, *13*, 2666–2674. [CrossRef]

20. Saxén, H. Short-term prediction of silicon content in pig iron. *Can. Metall. Q.* **1994**, *33*, 319–326. [CrossRef]
21. Zhou, B.; Ye, H.; Zhang, H.; Li, M. Process monotoring of iron-making process with PCA-based methods. *Control Eng. Pract.* **2016**, *47*, 1–14. [CrossRef]
22. Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P.; Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *J. Econom.* **1982**, *54*, 159–178. [CrossRef]
23. Lilliefors, H.W. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* **1967**, *62*, 399–402. [CrossRef]
24. Kruskal, W.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621. [CrossRef]
25. Haykin, S. *Neural Networks and Learning Machines*; Pearson Education, Inc.: Upper Saddle River, NJ, USA, 2009.
26. Hu, Y.H.; Hwang, J. *Handbook of Neural Network Signal Processing*; CRC Press: Boca Raton, FL, USA, 2002.
27. Mallat, S. *A Wavelet Tour of Signal Processing*; Academic Press: Burlington, MA, USA, 2008.
28. Hassibi, B.; Stork, D.G.; Wolff, G. Optimal Brain Surgeon and general network pruning. In Proceedings of the IEEE International Conference on Neural Networks, San Francisco, CA, USA, 28 March–1 April 1993; pp. 293–299.
29. The NNSYSID Toolbox-a MATLAB(R) Toolbox for System Identification with Neural Networks. Available online: https://www.mathworks.com/matlabcentral/fileexchange/87-nnsysid (accessed on 18 May 2021).