MDPI

Article

Prediction Models for Nitrogen Content in Metal at Various Stages of the Basic Oxygen Furnace Steelmaking Process

Jaroslav Demeter *[ID], Branislav Buľko [ID], Peter Demeter [ID] and Martina Hrubovčáková [ID]

Faculty of Materials, Metallurgy and Recycling, Institute of Metallurgical Technologies and Digital Transformation, Technical University of Košice, Letná 1/9, 042 00 Košice, Slovakia; branislav.bulko@tuke.sk (B.B.); peter.demeter@tuke.sk (P.D.); martina.hrubovcakova@tuke.sk (M.H.)
* Correspondence: jaroslav.demeter@tuke.sk; Tel.: +421-55-602-2407

Abstract

Controlling dissolved nitrogen is critical to meeting increasingly stringent steel quality targets, yet the variable kinetics of gas absorption and removal across production stages complicate real-time decision-making. Leveraging a total of 291 metal samples, the research applied ordinary least squares (OLS) regression, enhanced by cointegration diagnostics, to develop four stage-specific models covering pig iron after desulfurization, crude steel in the basic oxygen furnace (BOF) before tapping, steel at the beginning and end of secondary metallurgy processing. Predictor selection combined thermodynamic reasoning and correlation analysis to produce prediction equations that passed heteroscedasticity, normality, autocorrelation, collinearity, and graphical residual distribution tests. The k-fold cross-validation method was also used to evaluate models' performance. The models achieved an adequate accuracy of 77.23–83.46% for their respective stages. These findings demonstrate that statistically robust and physically interpretable regressions can capture the complex interplay between kinetics and the various processes that govern nitrogen pick-up and removal. All data are from U. S. Steel Košice, Slovakia; thus, the models capture specific setup, raw materials, and production practices. After adaptation within the knowledge transfer, implementing these models in process control systems could enable proactive parameter optimization and reduce laboratory delays, ultimately minimizing excessive nitrogenation in finished steel.

Keywords: nitrogen; prediction equations; production modeling; process optimization; digitalization

1. Introduction

In recent years, controlling the amount of nitrogen in the steelmaking process has become increasingly important. Controlling nitrogen content is important because high nitrogen content in metal has an overall adverse effect on the properties of the steel produced [1,2]. An exception is nitrogen in austenitic stainless steel, where the presence of nitrogen is desirable [3–8]. Increased nitrogen content leads to a deterioration in deep drawability and resistance to ageing, reduces the degree of recrystallization, and also impairs the mechanical properties of steel, which include formability and strength [9–11]. The increased presence of nitrogen in general steel also negatively affects the weldability of steel [12,13] (over 0.4 wt.% Nitrogen), as well as the electrical parameters [14]. In general, the nitrogen content of steel produced in a basic oxygen furnace (BOF) ranges from 20 to 60 ppm. Steel can contain nitrogen in an unbound form or as a chemical compound, such

as a nitride [15]. A large number of factors influence the nitrogen content in steel, and their interaction during the steelmaking process determines the final nitrogen value. The solubility of gases in molten metals is generally very low, which leads to the consideration of gases in metals as infinitely diluted solutions [16]. The process of nitrogen dissolving in molten metal is dictated by Equation (1), and the Gibbs free energy associated with this process is expressed by Equation (2) [15–19]. An explanation of all symbols listed in this paper can be found in the "List of symbols" at the end of the article.

$$\tfrac{1}{2} N_{2(g)} = [N] \tag{1}$$

$$\Delta G° = -3590 - 23.89\ T\ [J.mol^{-1}] \tag{2}$$

The equilibrium constant of Equation (1) has the form of Equation (3) [16,20].

$$K_N = \frac{a_N}{(p_{N_2})^{\frac{1}{2}}} = \frac{f_N [\%N]}{\sqrt{p_{N_2}}} \left[Pa^{-1/2} \right] \tag{3}$$

Given that nitrogen in molten metal is a diluted solution, it can be assumed that $f_N = 1$ [21,22]. Equation (1) being endothermic, it can be deduced that the solubility of nitrogen in the molten state increases with rising temperature. The solubility of nitrogen in liquid metal is found to be significantly dependent on gas pressure, temperature, and also on the metal itself, whose chemical composition undergoes substantial changes during the smelting process. Equation (4) expresses the relationship between the quantity of elemental nitrogen dissolved in metal and the pressure at which it is held above the molten metal at a constant temperature, which is known as Sievert's law [16,17,20,23,24].

$$[\%N] = K_N \cdot \left(p_{N_2} \right)^{\frac{1}{2}} \tag{4}$$

In the event that the dissolution of gases in metals occurs in strict accordance with Sievert's law, this serves as an indication that the gas is present in the liquid metal in its elemental form. However, in reality, the dependence of gas solubility on pressure is often more complex. This observation indicates that gas may not be present in metals exclusively in atomic form. Relevant results according to Sieverts' law are also only achieved at lower gas pressures above the surface of the liquid metal [16,25]. The temperature-dependent amount of dissolved elemental gas in molten metal can be expressed at constant pressure via the following Equation (5) [26].

$$[\%N] = C \cdot e^{\frac{-\Delta H}{2kT}} \tag{5}$$

According to the established relationship, it can be determined that the concentration of dissolved nitrogen in molten metal is directly proportional to the equilibrium constant KN, as depicted in Equation (1). Van't Hoff's reaction isobar demonstrates that the equilibrium constant is a function of temperature as showed in Equation (6).

$$\frac{d ln K_N}{dT} = \frac{\Delta H}{RT^2} \tag{6}$$

A significant increase in the solubility of nitrogen in iron is observed at a temperature of 907 °C. As the data demonstrate, with an increase in temperature, the solubility of nitrogen in γ-Fe decreases, which is related to the formation of Fe_4N and Fe_2N nitrides according to Equations (7) and (8) [27].

$$2\ N + 8\ Fe \leftrightarrow 2\ Fe_4N \tag{7}$$

$$N + 2\,Fe \leftrightarrow Fe_2N \tag{8}$$

The courses of Equations (7) and (8) are exothermic, with values that exceed the negative value of the heat of dissolution of nitrogen in iron. Therefore, as demonstrated in Equation (5), it can be concluded that the solubility of nitrogen in γ-Fe decreases with rising temperature.

In addition to iron, nitrogen in steel also reacts with other elements dissolved in the melt. This process gives rise to the formation of various chemical compounds, the existence of which depends on several factors. These include the chemical composition of the steel, the sequence of the steelmaking process, the temperature or pressure within the system, and the subsequent heat treatment of the steel. Figure 1a illustrates the impact of specific elements on the solubility of nitrogen in molten iron [28,29].

(a)   (b)
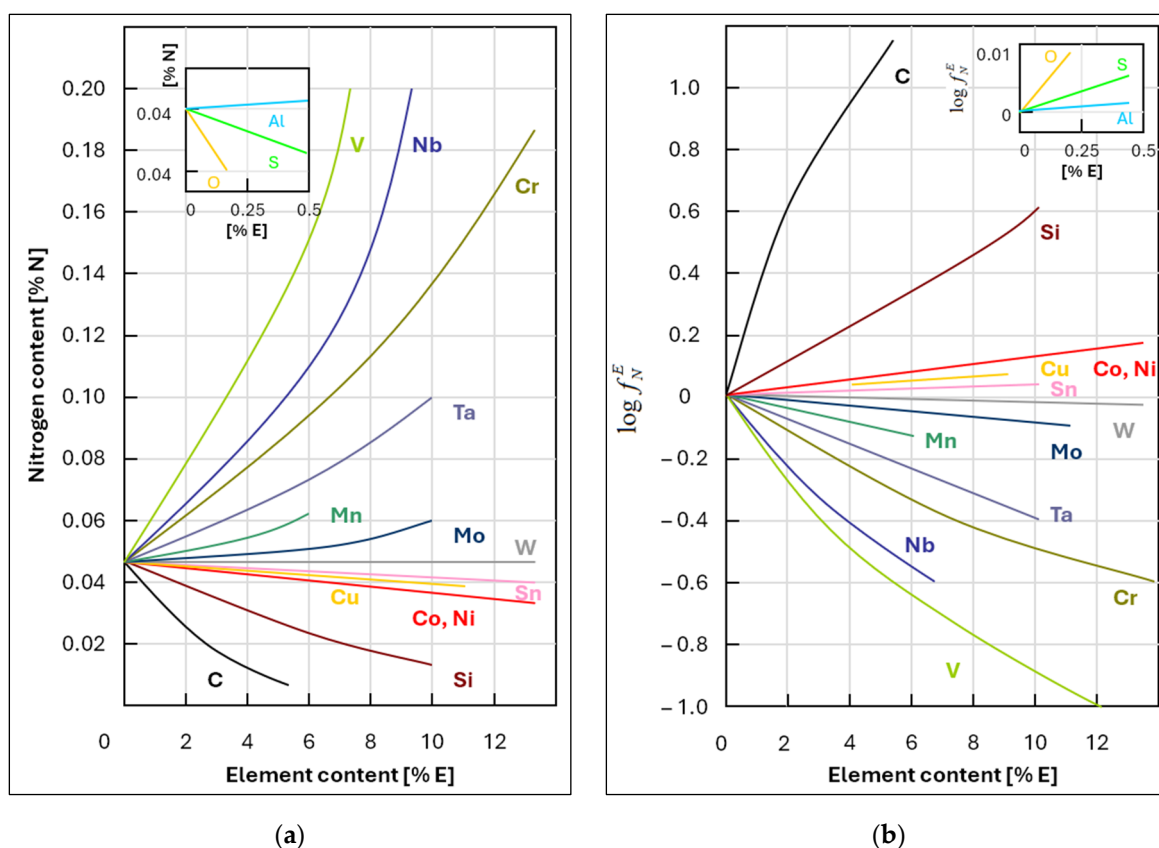
Figure 1. (**a**) The influence of individual elements on the solubility of nitrogen in liquid iron at 1600 °C and a nitrogen pressure of 101,325 Pa (picture recreated based on [28,29]); (**b**) the influence of individual elements on the nitrogen activity coefficient in Fe–N...E melts at 1600 °C (picture recreated based on [29]).

The presence of nitrogen in molten metal solutions has a significant influence, primarily due to the amount of carbon present. In this regard, the status of carbon (Figure 1b) is determined by the fact that its atoms occupy analogous positions to iron atoms when the crystal lattice forms. At the same time, carbon reacts with oxygen to form CO bubbles, which carry nitrogen atoms as they travel through the metal towards the surface [30].

In the event that the steel under consideration contains multiple dissolved elements, the solubility of nitrogen in the poly-component system can be expressed based on the transformed Equation (3) into the following Equation (9), which can be modified to form Equation (10).

$$a_{[N]} = f_{[N]} \cdot [N]_{steel} = K_N \cdot \sqrt{p_{N_2}} \tag{9}$$

$$log[N]_{steel} = logK_N + 0.5log.p_{N_2} - logf_{[N]} \tag{10}$$

The application of Sieverts law allows for the determination of the dependence of solubility on the heat, pressure, and chemical composition of steel, as expressed by Equation (11). The activity coefficient of nitrogen, $f_{[N]}$, is expressed as a function of elemental content, e_N^E, according to the following Equation (12), while the activity coefficient of nitrogen in the multicomponent system (Fe–N...E) is determined by the Wagner equation according to Equation (13).

$$log[N]_{steel} = -375T - 1.154 + 0.5log.p_{N_2} - logf_{[N]} \tag{11}$$

$$logf_N^E = e_N^E.[\%E] \tag{12}$$

$$f_N^{Fe...E} = f_N^*.f_N^C.f_N^D \cdots f_N^E \tag{13}$$

The actual nitrogen content in the metal melt is determined by the dissolution rate of nitrogen in the melt $\left(v_{N_{diss}}\right)$ and the removal rate of nitrogen from the melt $\left(v_{N_{rmvl}}\right)$. The temporal variation in the nitrogen content [N] of the melt, as a function of the blowing time (τ), is calculated by Equation (14).

$$\frac{d[N]}{d\tau} = v_{N_{diss}} - v_{N_{rmvl}} \tag{14}$$

The dissolution of nitrogen into molten metal can be described by the sequence of the following individual steps [29,31,32]: (1) transfer of nitrogen molecules from the gas volume to the phase interface; (2) adsorption of nitrogen molecules on the interfacial layer; (3) Reaction at the phase interface; (4) desorption of nitrogen atoms from the phase interface; (5) diffusion of nitrogen atoms from the gas-melt phase interface by melt flow into its volume.

Desorption followed by diffusion of nitrogen atoms into the metal volume probably has the greatest effect on the dissolution rate of nitrogen in liquid metal, as these are the slowest processes in the sequence of steps described above. As the rate of dissolution is determined by the slowest action, it is possible to formulate Equation (15) for the kinetics of nitrogen dissolution in molten metal [17].

$$v_{N_{diss}} = \frac{D}{\delta} \cdot S \cdot \left([N]_{srfc} - [N]_{vol}\right) \left[mol.s^{-1}\right] \tag{15}$$

The dissolution rate of nitrogen in molten metal is defined by Equations (16) and (17), as stated in the literature [17]. Equation (16) is used for the calculation at low concentrations of surfactants such as sulfur and oxygen, while Equation (17) is applied to find the dissolution kinetics of nitrogen at high concentrations of these elements.

$$\frac{dC_{[N]}}{dt} = \frac{A}{V} \cdot k \cdot \left(C_{[N]eq.} - C_{[N]}\right) \tag{16}$$

$$\frac{dC_{[N]}}{dt} = \frac{A}{V} \cdot k \cdot \left(C_{[N]eq.}^2 - C_{[N]}^2\right) \tag{17}$$

The determination of the kinetics of nitrogen dissolution in metal can be achieved through the utilization of Equation (18), which is dependent on the composition of the metal solution, its concentration, and the temperature [33].

$$k = \frac{D}{\delta} \cdot S = \beta \cdot S \tag{18}$$

Figure 2a,b shows a graphical representation of the nitrogen dissolution rate constant β in Fe–C and Fe–O alloys as a function of temperature and the concentration of the given element.

(**a**)

(**b**)
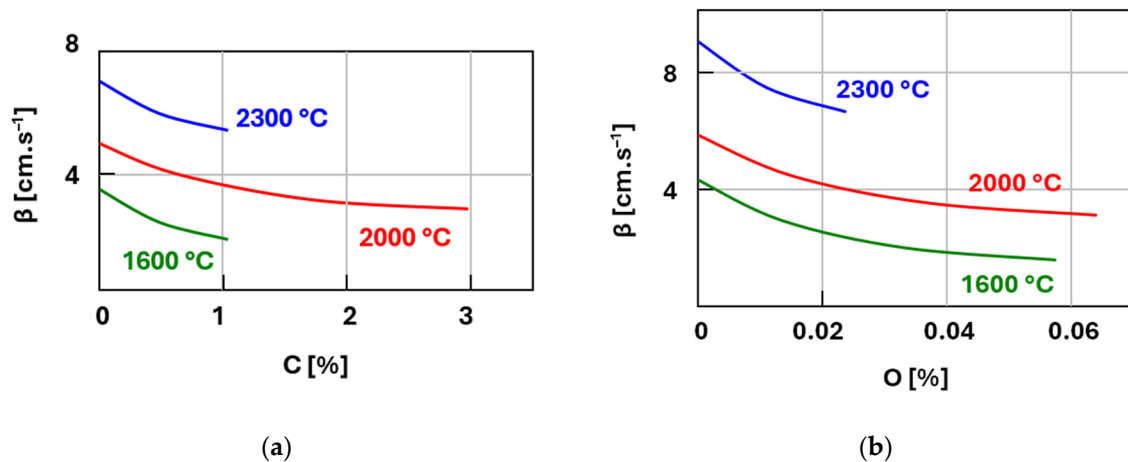
Figure 2. Dependence of the dissolution rate constant of nitrogen in alloying elements on carbon concentration and temperature (picture recreated based on [33]): (**a**) Fe–C alloy; (**b**) Fe–O alloy.

Improper control of nitrogen content during steel production can lead to critical quality issues that significantly compromise the mechanical properties, surface quality, and overall performance of steel products. These issues occur at various stages of production. One of the most significant quality issues associated with high nitrogen content is the formation of gas bubbles and porosity in steel. Nitrogen has limited solubility in molten steel, and its solubility decreases dramatically as temperature drops during solidification. Nitrogen bubbles form within the steel if nitrogen exceeds the solubility limit [34,35]. Research indicates that chromium increases nitrogen solubility most significantly, followed by manganese and molybdenum, while nickel decreases solubility. This compositional sensitivity requires careful alloy design to control nitrogen effects [34]. The formation of these bubbles follows Sieverts' law, which describes the dependence of gas concentration on pressure [35]. The porosity formation is particularly problematic in welding operations [36]. Steel can become brittle when exposed to high nitrogen concentrations, particularly when combined with other interstitial elements [37]. Improper nitrogen control can increase susceptibility to cold cracking, particularly in high-strength steels. The interstitial nitrogen atoms create stress concentrations that can initiate and propagate cracks [38]. Excessive nitrogen pickup reduces the steel's ability to deform plastically, leading to premature failure under load. This is particularly critical in structural applications, where ductility is essential for safety reasons [39].

Linear regression effectively predicts metallurgical outputs (production, temperature, strength, corrosion, nitrogen content) when relationships are approximately linear. With proper variable selection, data cleaning, and validation, it achieves up to 8% measurement accuracy [40]. Its advantages include simple implementation, high interpretability, and fast calculations. Combined with kinetic/thermodynamic models, it provides fundamental process understanding [41]. Random Forest algorithms consistently achieve 88–92% success across parameters, offering noise resistance and variable importance identification [42,43]. Support Vector Regression (SVR) and Extreme Learning Machine (ELM) excel in specific applications, with ELM providing extremely fast training [44]. Neural networks and Deep Learning demonstrate the highest accuracy (R^2 = 0.79–0.92) for various parameters [42,45,46], forming promising real-time solutions for comprehensive nitrogen content prediction in steelmaking processes.

The models presented in the available scientific literature focus on predicting the nitrogen content in metal in only one specific phase of BOF steel production. The novel aspect of this study is that it provides a comprehensive perspective on predicting the nitrogen content in metal throughout the entire BOF production cycle. This approach to the matter has not yet been published in the available studies.

The main aim was to develop predictive models of nitrogen content in liquid pig iron and liquid steel. The proposed approach to the model and the scope of research applied to the entire steel production cycle have not yet been developed in such a comprehensive form and published in professional literature. Models were created to predict the nitrogen content of metal at four production stages, namely, the nitrogen content of pig iron after desulphurization, the nitrogen content of crude steel before tapping from basic oxygen furnace (BOF) to the ladle, the nitrogen content of steel at the beginning of secondary metallurgy processing, and the nitrogen content of steel at the end of secondary metallurgy processing. This facilitates the monitoring of the development of nitrogen content in metal at each crucial production stage of a specific heat. This provides a tool for appropriate corrective intervention in the production process and its optimization to achieve the required quality parameters.

It is notable that all data originate from U. S. Steel Košice, Slovakia; as such, the models under consideration are representative of specific technological settings, raw materials, and production practices. Following adaptation within the knowledge transfer, implementation of these models in process control systems has the potential to enable proactive parameter optimization and reduce laboratory delays, minimizing excessive nitrogenation of finished steel.

2. Materials and Methods

The production process of the material used to generate the data for this article can be summarized as follows. Pig iron produced in a blast furnace was pretreated (desulfurized) using a vertical refractory lance (Scandinavian lance) with a CaO–Mg-based mixture for desulfurization. This mixture was blown into the pig iron using nitrogen as the carrier gas. The pig iron was then charged into a top-blown basic oxygen furnace (BOF/LD process), where a water-cooled oxygen lance was used to blow high-purity oxygen at supersonic velocity onto the surface of the metal to oxidize the impurities. The BOF's capacity was 170 tons. After approximately 17 min, the crude steel was tapped into a steel ladle. During tapping from the BOF, the steel in the ladle was deoxidized by adding aluminum and alloyed by adding ferroalloys. The steel chemical composition and thermal composition were also finished and stirred in the ladle by blowing/bubbling argon through a porous plug at the bottom. The chemical composition of the steel was finally refined, and the temperature was finally optimized by blowing/bubbling argon through a porous refractory block located at the bottom of the ladle. The analyzed heats were not vacuum-treated in the RH vacuum degasser. At the end of secondary metallurgy, the steel was prepared for casting on a continuous casting machine. The flow charts (Figure A1a–d) that illustrate the overall research framework and model implementation within given phases can be found in Appendix A.

A total of 291 metal samples from 76 heats were collected from 17 May 2025 to 22 May 2025 and analyzed for their nitrogen content. The sampling design aimed to obtain samples from all four production phases within a single heat. This made it possible to monitor the amount of nitrogen in a given heat and how it varied throughout the production process. A collection of 76 samples (standard "lollypop-shaped" samples) was obtained from pig iron after desulfurization from the ladle, 68 samples were obtained from crude steel before tapping from BOF, 75 samples were obtained from molten steel in the ladle at the beginning

of secondary metallurgy, and 72 samples were obtained from molten steel in the ladle at the end of secondary metallurgy. The reason why it was not possible to obtain exactly 76 samples from each phase was most likely due to a failure in sampling, in which it was not possible to adequately and reliably evaluate the nitrogen content in the sample itself.

The samples taken consisted of two grades of steel. The first was a structural steel grade with a manganese content above 0.80% and a guaranteed Al content with the following prescribed composition: 0.07–0.21% C; 0.8–1.6% Mn; 0.03–0.6% Si; min 0.02% Al; max. 0.025% P; max. 0.020% S. The second grade was deep-drawing Al-killed steel with the following prescribed composition: 0.02–0.1% C; 0.1–0.55% Mn; max. 0.08% Si; 0.02–0.07% Al; 0.01–0.07% P; max. 0.020% S; 0.004–0.0075% Nb.

The samples were evaluated for nitrogen content in pig iron and steel at the certified Quantometric laboratory of U. S. Steel Košice (Labortest, s.r.o.) by using a combustion analyzer ELTRA ON 900 (ELTRA GmbH, Haan, Germany) that determines nitrogen using a thermal conductivity detection method (based on ASTM E-1019 standard [47]). The equipment is calibrated by the manufacturer and checked once a year by the manufacturer's service technician. Every four hours, a laboratory technician checks the device by using a sample standard. The analyzer has a measurement range of 0.0001–0.03% N with an error range of ±0.1 ppm or ±1% of nitrogen content. Two measurements, main and control, are evaluated.

The nitrogen content of the metal was determined using heat and sample identification numbers. These data were synchronized with relevant databases containing records of the metal's chemical composition, temperature, weight, and other parameters attributable to that stage of processing. The synchronized parameter database was initially compiled in Microsoft Excel 365 with the Lumivero XLSTAT 2019 statistical add-in. Subsequently, the correlation matrix, generated using STATISTICA 7.0 from StatSoft Inc. (Tulsa, OK, USA), enabled the determination of the order of factors affecting the final nitrogen content at a given stage of production. The data file was processed using Gretl 2025a (build 2025-03-20), a powerful and sophisticated statistical tool. The k-fold cross-validation was performed using the open-source program Orange Data Mining 3.39.0 from the Bioinformatics Lab at the University of Ljubljana, Slovenia.

Determining the factors and assessing their impact on the amount of nitrogen in metal is a key attribute in optimizing technological and operational measures aimed at minimizing or preventing over-nitrogenation of the final steel. It is crucial to identify and quantify the influence of individual factors on the nitrogen content in metal to effectively address the issue of predicting its quantity during the production and processing stages of steelmaking. A correlation matrix was created that included all analyzed factors, which made it possible to determine the quantitative dependence in which changes in one variable lead to changes in another.

This dependence is expressed by a function called a regression function. Depending on its shape, the regression function can indicate a positive or negative correlation. A positive correlation occurs when the regression function is increasing, and a negative correlation occurs when the regression function is decreasing. Each indicator was subjected to classical and modern regression analysis, while elements of cointegration analysis of non-stationary variables were also applied [48].

The existence of linear correlation dependence in a two-dimensional statistical set is indicated by the presence of a covariance in Equation (19).

$$k = S_{xy} = \frac{1}{n} \cdot \sum_{i=1}^{n}(x_i - \overline{x}) \cdot (y_i - \overline{y}) \tag{19}$$

where (x_i, y_i) are individual pairs of observations, $\overline{x}$ and $\overline{y}$ are the corresponding arithmetic means. The covariance depends on the choice of scale for the values of random variables ξ and η. Therefore, Pearson's correlation coefficient is used to measure the degree of linear correlation [49,50]. The correlation coefficient of a statistical set ranges within the closed interval $\langle -1,1 \rangle$. In absolute terms, the value approaches 1 as the linear correlation between the variables ξ and η increases. The calculation of the correlation coefficient of a two-dimensional basic set is performed in accordance with the following Equation (20) [51].

$$\rho(\xi, \eta) = \frac{cov(\xi, \eta)}{\sqrt{D(\xi) \cdot D(\eta)}} \tag{20}$$

The correlation coefficient R measures the strength of statistical dependence between two quantitative variables. Unlike regression, correlation analysis does not express a cause-and-effect relationship $Y = f(X)$. Variable Y does not depend on variable X, but the two random variables X and Y change together. Regression analysis assumes that variable Y is random and variable X is fixed. The term correlation coefficient most often refers to Pearson's correlation coefficient in Equation (21) [52].

$$R_{xy} = \frac{\sum x_i y_i - n\overline{xy}}{(n-1)s_x s_y} \tag{21}$$

As demonstrated in Equation (21), it is feasible to identify the order of relevance of the factors that exert the most significant influence on the amount of nitrogen in the metal within a specific phase of steel production in a closed cycle. The objective of this study was to identify the most significant factors for each phase that was analyzed.

The interpretation of the correlation coefficient R depends on the context and nature of the data. A value of 0.8 is very low for verifying a physical law, but very high for the social sciences [53,54]. Jacob Cohen developed a simplified instrument for the interpretation of correlation coefficients in research [55], as illustrated in Table 1.

Table 1. Cohan's interpretation of correlation coefficients [55].

Correlation coefficient	0–0.1	0.1–0.3	0.3–0.5	0.5–0.7	0.7–0.9	0.9–1
Correlation interpretation	trivial, very small	little, low	moderate	big, high	very big, very high	perfect, clear

The R^2 value (22) is referred to as the coefficient of determination, which quantifies the proportion of common variability between two variables. The R^2 value achieved is contingent upon the nature of the data being processed. The interpretation of the coefficient of determination values achieved is significantly influenced by the nature of the analyzed data.

$$R_{xy}^2 = 1 - \frac{s_{y|x}^2}{s_y^2} \quad \text{or} \quad R_{xy}^2 = -\frac{s_{x|y}^2}{s_x^2} \tag{22}$$

Non-stationarity is a property of time series or datasets in which the statistical properties (e.g., mean, variance) are shown to change over time. Econometrics is characterized by the utilization of non-stationary datasets, a principle that aligns with the methodologies employed in metallurgy [56,57]. In the field of economic sciences, it has been established that coefficient of determination values above $R^2 = 0.15$ are not achieved [58]. Non-stationarity is also observed in metallurgical data, particularly in processes where conditions vary over

time, such as temperatures during melting, cooling, or changes in chemical composition during metal processing [59].

The combination of modern linear regression analysis and cointegration analysis enables the design, analysis, and verification of predictive models that can be applied in operating conditions. Non-stationary variables are a characteristic of metallurgy. When such time-dependent variables are used, there is a risk of spurious regression. This occurs when there is no cointegration effect between non-stationary variables, i.e., no long-term, stable relationship. If spurious regression were not detected, this would negatively impact the identification of factors, the interpretation of results, and the application of the model in real technological practice. In this study, cointegration analysis was employed to validate the model derived from modern linear regression analysis.

The Ordinary Least Squares (OLS) method was used for data processing. Compared to other estimation techniques, the OLS method provides optimal estimates even for small samples of observations, and the algorithm for calculating these estimates is relatively simple [60,61]. Furthermore, this method forms the basis of a wide range of more sophisticated estimation tests and procedures.

For statistical processing purposes, the following general model was proposed for calculating the predicted amount of nitrogen in metal at individual stages of steel production (23).

$$N_n = z_0 + z_1 \cdot Z_1 + z_2 \cdot Z_2 + \cdots + z_n \cdot Z_n \tag{23}$$

Adequate coefficients were calculated for the analyzed factors influencing the amount of nitrogen present in metal during the various stages of steel production. The corresponding standard error, t-ratio, and *p*-value were also calculated for each coefficient. The t-ratio parameter expresses the value of the standard deviation [62,63] and assesses whether two groups (dependent and independent) differ statistically from each other [64]. The *p*-value is a random variable based on the measured values of the monitored quantity. It expresses the probability of obtaining a value at least as extreme as the one actually observed, assuming the null hypothesis is true. It ranges [0, 1] [65].

In statistical hypothesis testing, the null hypothesis formally evaluates specific aspects of the statistical behavior of datasets. This evaluation is considered valid if the actual behavior of the datasets contradicts the assumption. Consequently, the null hypothesis can be accepted or rejected [66]. The proposed models were analyzed using several verification methods, such as the heteroscedasticity test, normality test, autocorrelation test, collinearity test, and graphical examination of residual distribution. To confirm the accuracy of the conclusions of the modern regression analysis, an econometric cointegration analysis was also performed.

The estimation of reliable performance was accomplished through the utilization of k-fold cross-validation. This statistical technique is used to evaluate the performance of statistical models. It operates on the principle of dividing a dataset into k equal parts (folds). In this study, the setting k = 10 (10-fold cross-validation) was employed due to the smaller dataset. This configuration engenders reduced bias, with 90% of the data being utilized for training in each iteration. In an analogous manner, when k = 10, a sufficient number of iterations are performed to obtain stable results, which also represents computational efficiency.

The following were used to determine the accuracy of the prediction models: Mean Absolute Error (MAE), which expresses the average absolute deviation of the actual values from the estimated (or predicted) values. MAE is determined using Equation (24).

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t| \tag{24}$$

The Mean Percentage Error (MPE), which expresses the degree of distortion, was also applied to determine the accuracy of the model. Equation (25) can be used to calculate whether the model underestimates or overestimates reality.

$$MPE = \frac{1}{n} \sum_{t=1}^{n} \frac{e_t}{y_t} \cdot 100 [\%] \tag{25}$$

Mean Absolute Percentage Error (MAPE), according to Equation (26), was used to express the average size of forecast errors as a percentage of the actual measured values over the entire forecasting period [67].

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|e_t|}{y_t} \cdot 100 [\%] \tag{26}$$

The accuracy of the entire model is determined according to Equation (27).

$$\text{Model accuracy: } N_{stage} = 100 - MAPE\ [\%] \tag{27}$$

3. Results

Determining the factors and assessing their impact on the nitrogen content of metal is key to optimizing technological and operational measures to minimize or prevent excessive nitrogenation of the final steel. The identification and quantification of the influence of individual factors on the nitrogen content in metal is imperative for the effective prediction of its quantity during the production and particular steelmaking processing stages. Through the synthesis of correlation analysis findings, physical and empirical evidence, the most significant agents were identified, thereby establishing a foundation for subsequent, more intricate statistical procedures. This study utilized a comprehensive analytical approach encompassing classical and modern regression analysis, complemented by elements of cointegration analysis of non-stationary variables employed in econometric time series analyses.

3.1. Parameters Affecting the Amount of Nitrogen in Molten Desulfurized Pig Iron

As shown in Table 2, the following factors have been identified as the most significant in determining the amount of nitrogen present in desulphurized pig iron: the results of the regression analysis were obtained using a correlation matrix. As demonstrated in Table 2, the correlation indices could be interpreted as low to trivial according to Table 1. However, it should be noted that these are non-stationary integrated variables, where high correlation coefficients are not achieved even though these parameters have a significant effect on the dependent variable, which in this case is nitrogen.

Figure 3 shows a graphical representation of the effect of the amount of nitrogen added as a carrier gas for the desulphurization mixture on the amount of sulfur removed and the resulting nitrogen content of the pig iron. Figure 4a,b illustrates the impact of certain parameters listed in Table 2 on the final nitrogen content in metal during the desulphurization phase of pig iron processing. As demonstrated in Figure 4a, an increase in the amount of sulfur removed results in an increase in the amount of nitrogen present in the pig iron. Similarly, Figure 4b illustrates that an increase in the amount of nitrogen blown results in an increase in the amount of nitrogen dissolved in the pig iron A more thorough examination of the results is presented in Section 4.1.

Table 2. Ranking of factors affecting the nitrogen content in desulphurized pig iron.

| Rank	Coefficient	Factor	Correlation Coefficient R	Coefficient of Determination R^2
1.	a_1	Amount of sulfur removed [%]	0.2226	0.0495
2.	a_2	Amount of nitrogen added as a carrier gas [l]	0.1379	0.0190
3.	a_3	Weight of pig iron after desulfurization [kg]	0.1123	0.0126
4.	a_4	Amount of desulphurization mixture [kg]	0.0965	0.0093
5.	a_5	Temp. difference before and after desulfurization [°C]	0.0964	0.0093
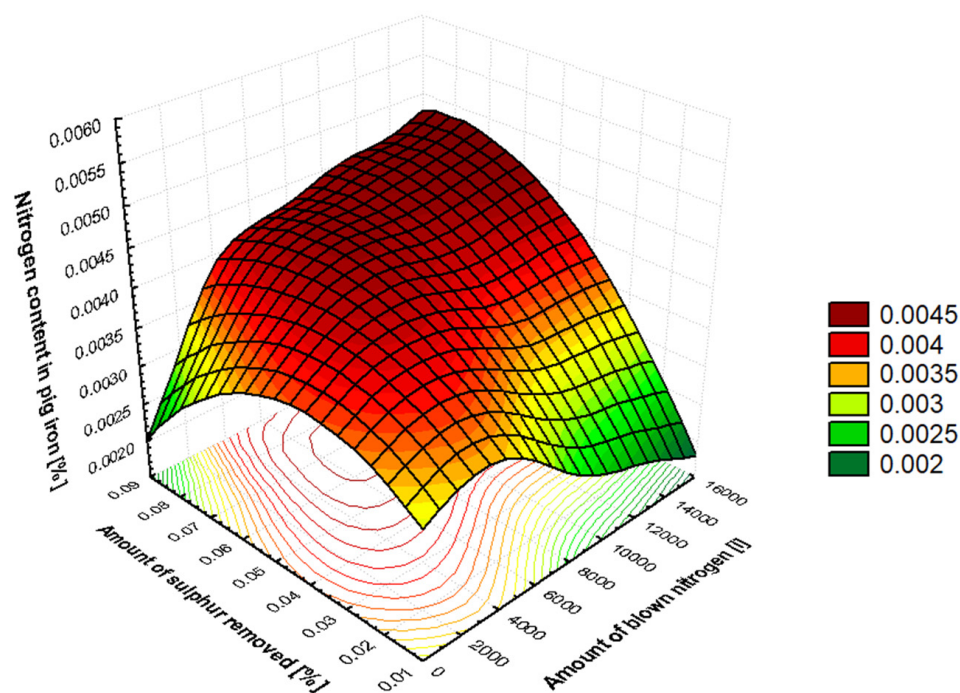
Figure 3. The effect of the amount of nitrogen added as a carrier gas during the desulphurization of pig iron and the amount of sulfur removed during the desulfurization process on the nitrogen content in pig iron.

(**a**)   (**b**)
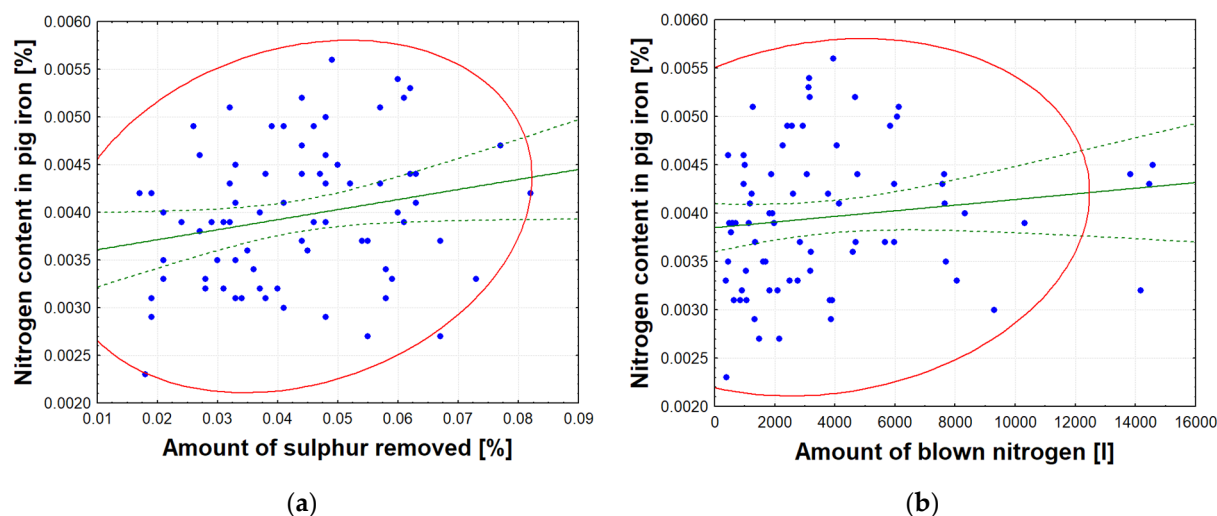
Figure 4. (**a**) The effect of the amount of sulfur removed on the nitrogen content in a metal sample after desulfurization of pig iron; (**b**) the effect of the amount of blown nitrogen (carrier gas) on the nitrogen content in the metal after desulfurization of pig iron.

Table 3 shows the conditions for statistical testing of nitrogen presence in pig iron before the desulphurization process.

Table 3. Conditions for statistical testing of nitrogen presence in pig iron before desulphurization.

Entry Requirement	Value
Significance level, α	0.05 *
Number of observations, n	76
Number of parameters, m	5

* Unless otherwise stated.

Tables 4 and 5 show the results of modern regression analysis for the individual factors (variables) listed in Table 2.

Table 4. Results of a modern regression analysis of the factors influencing the nitrogen content in metal during the desulfurization phase of pig iron.

Coefficient	Value	Standard Error	t-Ratio	p-Value
a_0	0.000105058	0.0081269	0.0129	0.98972
a_1	0.0220528	0.00959012	2.2995	0.02446
a_2	2.5678×10^{-8}	2.69278×10^{-8}	0.9536	0.34357
a_3	2.50746×10^{-8}	5.74314×10^{-8}	0.4366	0.66374
a_4	-2.35159×10^{-6}	1.38981×10^{-6}	-1.6920	0.09509
a_5	4.96884×10^{-6}	1.0213×10^{-5}	0.4865	0.62812

Where: a_0 constant/intercept, a_1–a_5 coefficients of variables (from Table 2).

Table 5. Results of individual tests using modern regression analysis to identify the factors influencing nitrogen content in metal during the desulphurization phase of pig iron production.

Parameter	Value
Mean value of the dependent variable	0.003958
Sum of squared residuals	0.000036
R^2—Coefficient of Multiple Determination	0.097373
$F_{(5, 70)}$—F-test	1.510277
Standard deviation of the dependent variable	0.000730
Durbin–Watson test	1.641496

3.2. Parameters Affecting the Amount of Nitrogen in Molten Crude Steel Before Tapping from BOF

Table 6 lists the most significant factors affecting the amount of nitrogen in crude steel before its tapping from BOF.

Table 6. Ranking of factors affecting the nitrogen content in crude steel before tapping from BOF.

Rank	Coefficient	Factor	Correlation Coefficient R	Coefficient of Determination R^2
1.	b_1	Oxygen reblow time [s]	0.4291	0.1842
2.	b_2	Manganese content in crude steel [%]	-0.3017	0.0910
3.	b_3	Phosphorus content in crude steel [%]	-0.2339	0.0547
4.	b_4	Carbon content in crude steel [%]	-0.2055	0.0422
5.	b_5	Briquettes [kg]	-0.1594	0.0254
6.	b_6	Temperature of tapping steel [°C]	0.1457	0.0212
7.	b_7	Oxygen blowing time [s]	-0.1314	0.0173

Based on Cohen's distribution, the correlation coefficients can be interpreted as ranging from small to medium. However, it is important to note the non-stationary nature of the

data from integrated systems, and that achieving medium correlation coefficient values with operational data is a great success.

Figure 5 shows the effect of the levels of phosphorus and manganese in crude steel on the final nitrogen content in crude steel before its tapping from BOF. Figure 6a–d illustrates the impact of certain parameters listed in Table 6 on the final nitrogen content in crude steel produced in BOF prior to tapping into the ladle. As illustrated in Figure 6a, an increase in the amount of manganese in crude steel is accompanied by a decrease in the amount of nitrogen dissolved in crude steel. In a similar manner, an increase in phosphorus (Figure 6b) and carbon (Figure 6c) results in a decrease in the amount of nitrogen dissolved in crude steel. Conversely, an increase in the temperature of crude steel (Figure 6d) has been shown to result in an increase in the amount of nitrogen dissolved in the crude steel. Further analysis of the results is provided in Section 4.2.
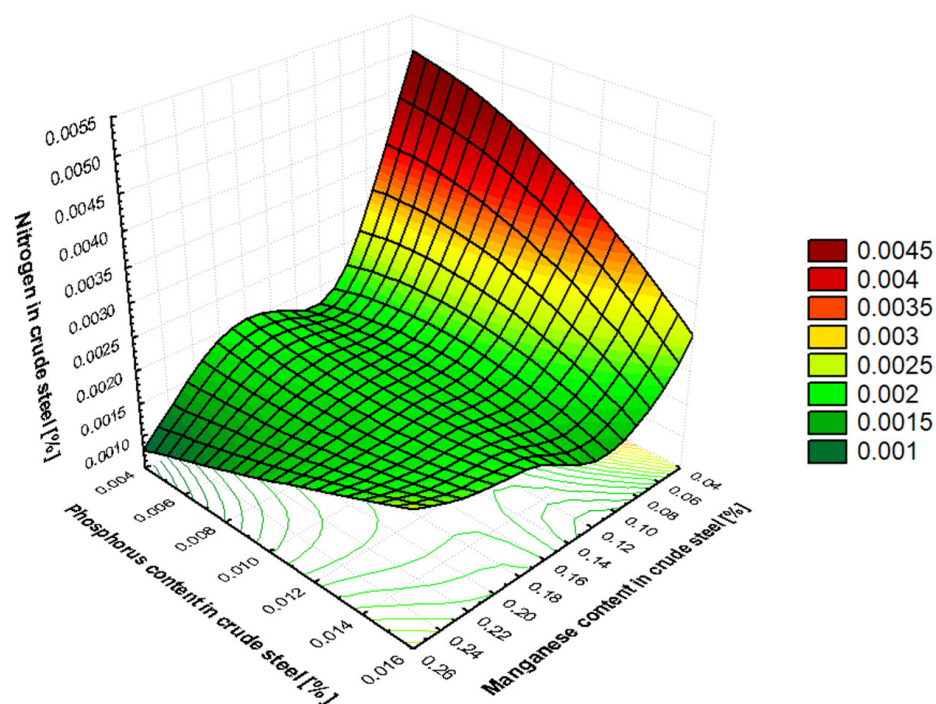
Figure 5. The effect of the phosphorus and manganese content in crude steel on the nitrogen content in crude steel produced in BOF prior to tapping.

The conditions for the statistical testing of nitrogen presence in crude steel prior to tapping from the BOF are provided in Table 7.

Table 7. Conditions for statistical testing of nitrogen presence in steel before tapping from BOF.

Entry Requirement	Value
Significance level, α	0.05 *
Number of observations, n	63
Number of parameters, m	7

* Unless otherwise stated.

Tables 8 and 9 present the findings of modern regression analysis for the individual factors (variables) enumerated in Table 6. It can be observed that correct results were achieved using the OLS method.
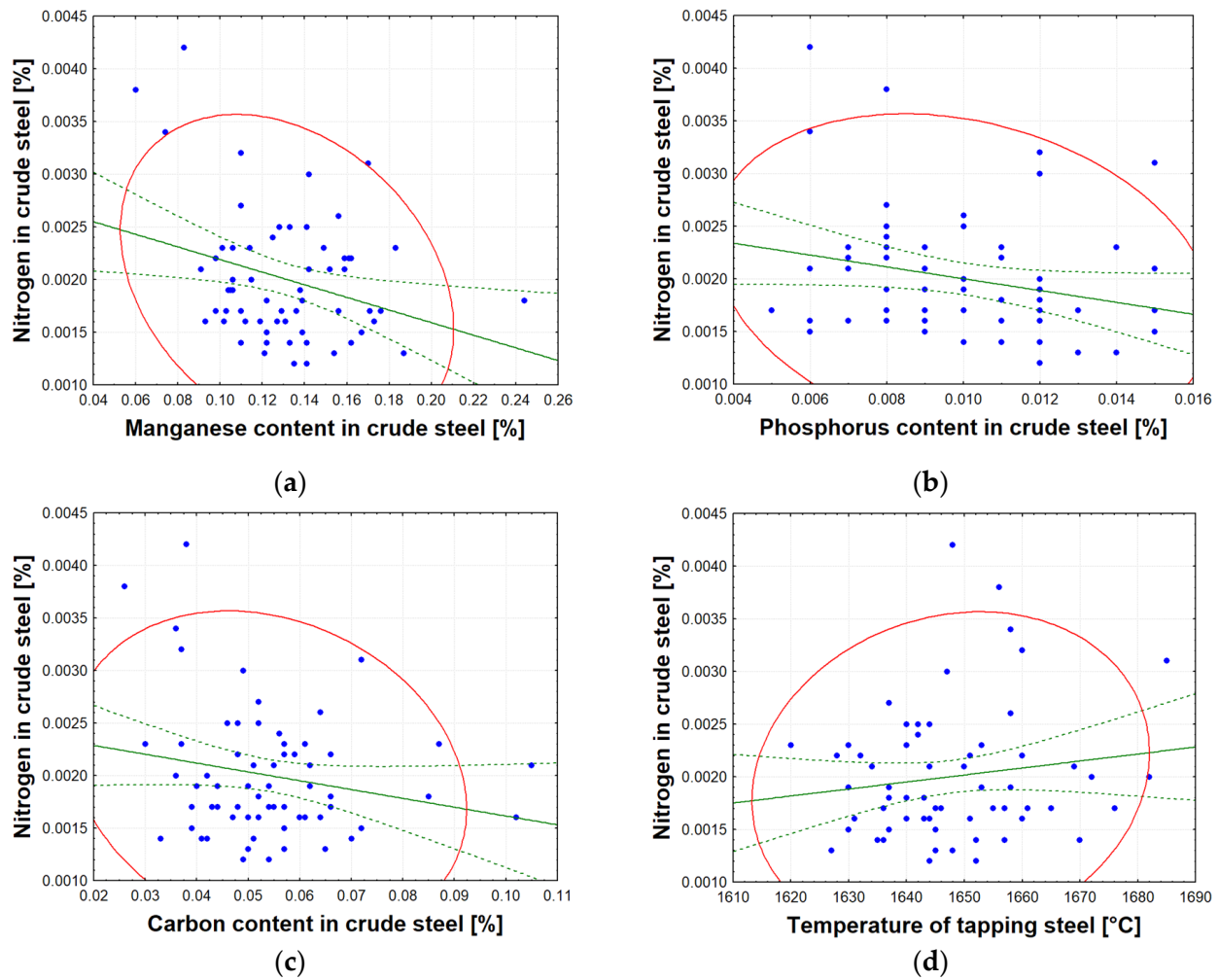
Figure 6. (**a**) The effect of manganese content in molten steel on nitrogen content in crude steel before tapping; (**b**) the effect of phosphorus content in molten steel on nitrogen content in crude steel before tapping; (**c**) the effect of carbon content in molten steel on nitrogen content in crude steel before tapping; (**d**) the effect of temperature of tapping molten steel on nitrogen content in crude steel before tapping.

Table 8. Results of a modern regression analysis (OLS) of the factors influencing the nitrogen content in crude steel prior to tapping from BOF.

Coefficient	Value	Standard Error	*t*-Ratio	*p*-Value
b_0	-0.0137962	0.0113016	-1.2207	0.22740
b_1	7.06124×10^{-6}	2.94964×10^{-6}	2.3939	0.02011
b_2	-0.000306035	0.00286768	-0.1067	0.91540
b_3	-0.0469186	0.0353764	-1.3263	0.19023
b_4	-0.00499063	0.00482186	-1.0350	0.30520
b_5	-3.44051×10^{-8}	7.63705×10^{-8}	-0.4505	0.65412
b_6	1.04457×10^{-5}	6.87455×10^{-6}	1.5195	0.13437
b_7	-3.96352×10^{-7}	4.22613×10^{-7}	-0.9379	0.35242

Where: b_0 constant/intercept, b_1–b_7 coefficients of variables (from Table 6).

Table 9. Results of individual tests using modern regression analysis to identify the factors influencing nitrogen content in crude steel before its tapping from BOF.

Parameter	Value
Mean value of the dependent variable	0.002002
Sum of squared residuals	0.000017
R^2—Coefficient of Multiple Determination	0.292051
$F_{(7, 55)}$—F-test	2.326593
Standard deviation of the dependent variable	0.000614
Durbin–Watson test	2.000830

3.3. Parameters Affecting the Amount of Nitrogen in Molten Steel at the Beginning of Secondary Metallurgy

Table 10 lists the strongest factors influencing the amount of nitrogen in steel at the beginning of secondary metallurgy (SM).

Based on Table 1, the results in Table 10 can be interpreted as indicating small to medium levels of dependency. However, higher correlation coefficient values can be observed when compared to the results in Table 6. Figures 7–9 illustrate the impact of certain parameters from Table 10 on the nitrogen content of molten steel at the beginning of the SM processing. It is evident from Figure 9a,b that an increase in the amount of carbon and manganese in the molten steel prior to argon bubbling results in an increase in the amount of nitrogen dissolved in the molten steel. A more detailed analysis of the results can be found in Section 4.3.
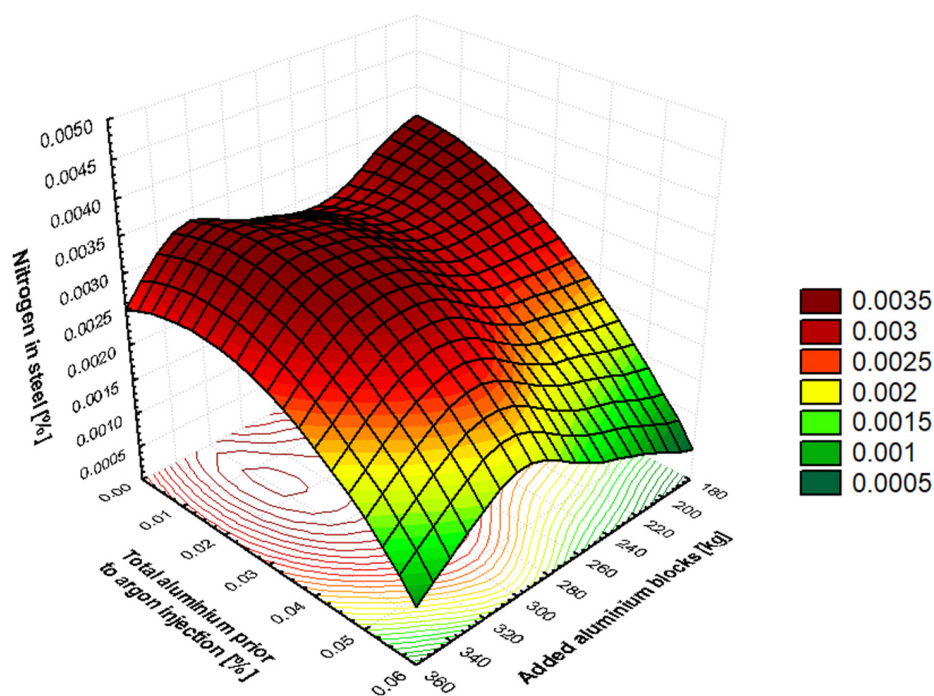
Figure 7. The influence of the total amount of aluminum in steel before argon bubbling and the amount of added aluminum blocks into steel on the nitrogen content in steel at the beginning of the SM process.

Table 10. Ranking of factors affecting the nitrogen content in steel at the beginning of secondary metallurgy.

| Rank	Coefficient	Factor	Correlation Coefficient R	Coefficient of Determination R^2
1.	c_1	Tapping angle [°]	−0.4470	0.1998
2.	c_2	Silicon in molten steel prior to argon bubbling [%]	0.3711	0.1377
3.	c_3	Total aluminum prior to argon bubbling [%]	−0.2749	0.0756
4.	c_4	Carbon in molten steel prior to argon bubbling [%]	0.2285	0.0522
5.	c_5	Manganese in molten steel prior to argon bubbling [%]	0.2251	0.0507
6.	c_6	Tapping time [s]	0.2203	0.0485
7.	c_7	Added aluminum blocks [kg]	0.2040	0.0416
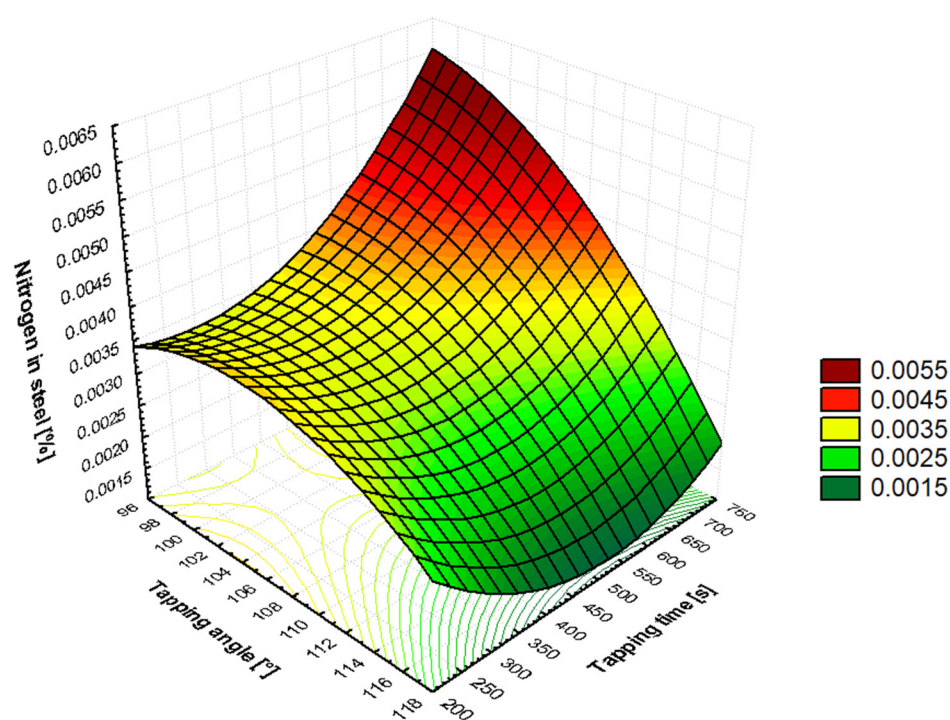
Figure 8. The influence of the tapping angle of BOF and the overall tapping time of steel from BOF on the nitrogen content in steel at the beginning of the SM process.

Table 11 shows the conditions for statistical testing of nitrogen presence in steel at the beginning of secondary metallurgy processing.

Table 11. Conditions for statistical testing of nitrogen presence in steel at the beginning of secondary metallurgy.

Entry Requirement	Value
Significance level, α	0.05 *
Number of observations, n	75
Number of parameters, m	7

* Unless otherwise stated.
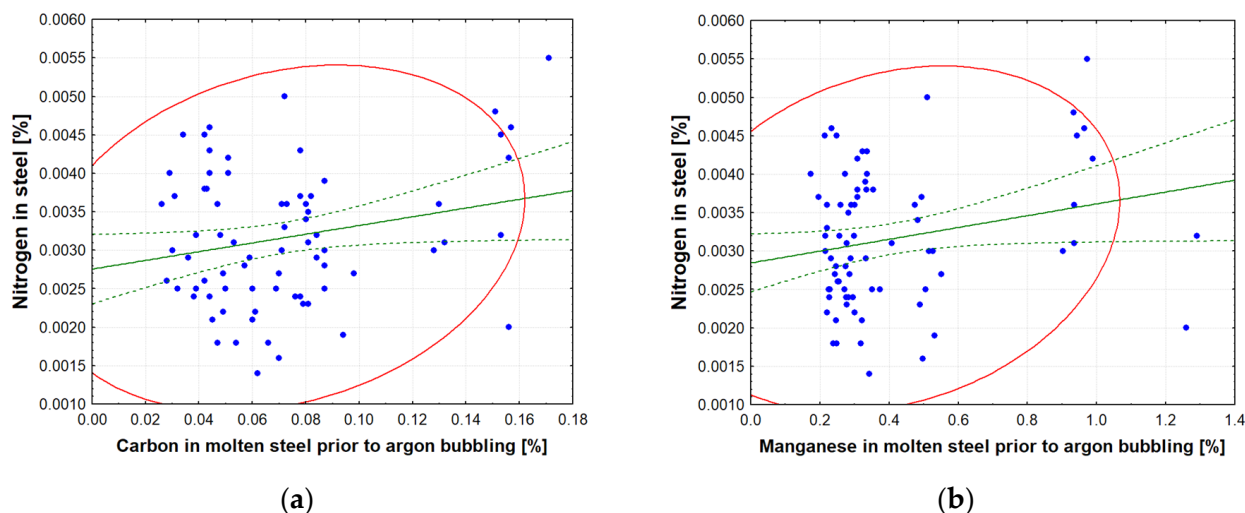
Figure 9. (**a**) The effect of the carbon content in molten steel prior to argon gas bubbling on the nitrogen content in a metal sample before SM; (**b**) the influence of the manganese content in molten steel prior to argon gas bubbling on the nitrogen content in a metal sample before SM.

Tables 12 and 13 present the outcomes of a modern regression analysis of the individual factors (variables) listed in Table 10. The tables demonstrate that the correct results were obtained using the ordinary least squares method (OLS).

Table 12. Results of a modern regression analysis (OLS) of the factors influencing the nitrogen content in steel at the beginning of the SM processing.

Coefficient	Value	Standard Error	*t*-Ratio	*p*-Value
c_0	0.00982567	0.00228811	4.2942	0.00006
c_1	-7.89409×10^{-5}	1.89305×10^{-5}	-4.1700	0.00009
c_2	0.00368566	0.00143795	2.5631	0.01262
c_3	-0.00987549	0.00757363	-1.3039	0.19672
c_4	0.00442257	0.00595744	0.7424	0.46046
c_5	-0.00140863	0.000838822	-1.6793	0.09775
c_6	6.15271×10^{-7}	1.35425×10^{-6}	0.4543	0.65106
c_7	6.28702×10^{-6}	3.14335×10^{-6}	2.0001	0.04955

Where: c_0 constant/intercept, c_1–c_7 coefficients of variables (from Table 10).

Table 13. Results of individual tests using modern regression analysis to identify the factors influencing the final nitrogen content in steel at the beginning of the SM processing.

| Parameter	Value
Mean value of the dependent variable	0.003160
Sum of squared residuals	0.000035
R^2—Coefficient of Multiple Determination	0.402680
$F_{(7, 67)}$—F-test	6.452518
Standard deviation of the dependent variable	0.000888
Durbin–Watson test	1.872243

3.4. Parameters Affecting the Amount of Nitrogen in Molten Steel at the End of Secondary Metallurgy

Table 14 shows the most significant factors that influence the nitrogen content of steel at the end of the secondary metallurgy processing.

Table 14. Ranking of factors affecting the nitrogen content in steel at the end of secondary metallurgy (SM).

Rank	Coefficient	Factor	Correlation Coefficient R	Coefficient of Determination R^2
1.	d_1	Steel temperature at the end of SM [°C]	−0.2768	0.0766
2.	d_2	Final carbon in molten steel [%]	0.2268	0.0514
3.	d_3	Addition of FeMn aff. during SM [%]	0.2056	0.0423
4.	d_4	Final manganese in molten steel [%]	0.1930	0.0372

According to Cohen's distribution, the correlation coefficient values from Table 14 are low. Figure 10a–d illustrates the factors from Table 14 that influence the final nitrogen content in steel at the end of the secondary metallurgy processing. It can be seen in Figure 10a that, as the temperature of the steel increases at the end of SM, the amount of dissolved nitrogen in the molten steel decreases. Figure 10b shows that, as the final amount of carbon in the molten steel increases, the amount of nitrogen in the molten steel also increases. Adding FeMn aff. ferroalloy (Figure 10c) increases the amount of nitrogen in the liquid steel, and increasing the amount of manganese in the steel (Figure 10d) increases the amount of nitrogen in the steel at the end of SM. Further analysis of the results is provided in Section 4.4.

(a)

(b)

(c)

(d)
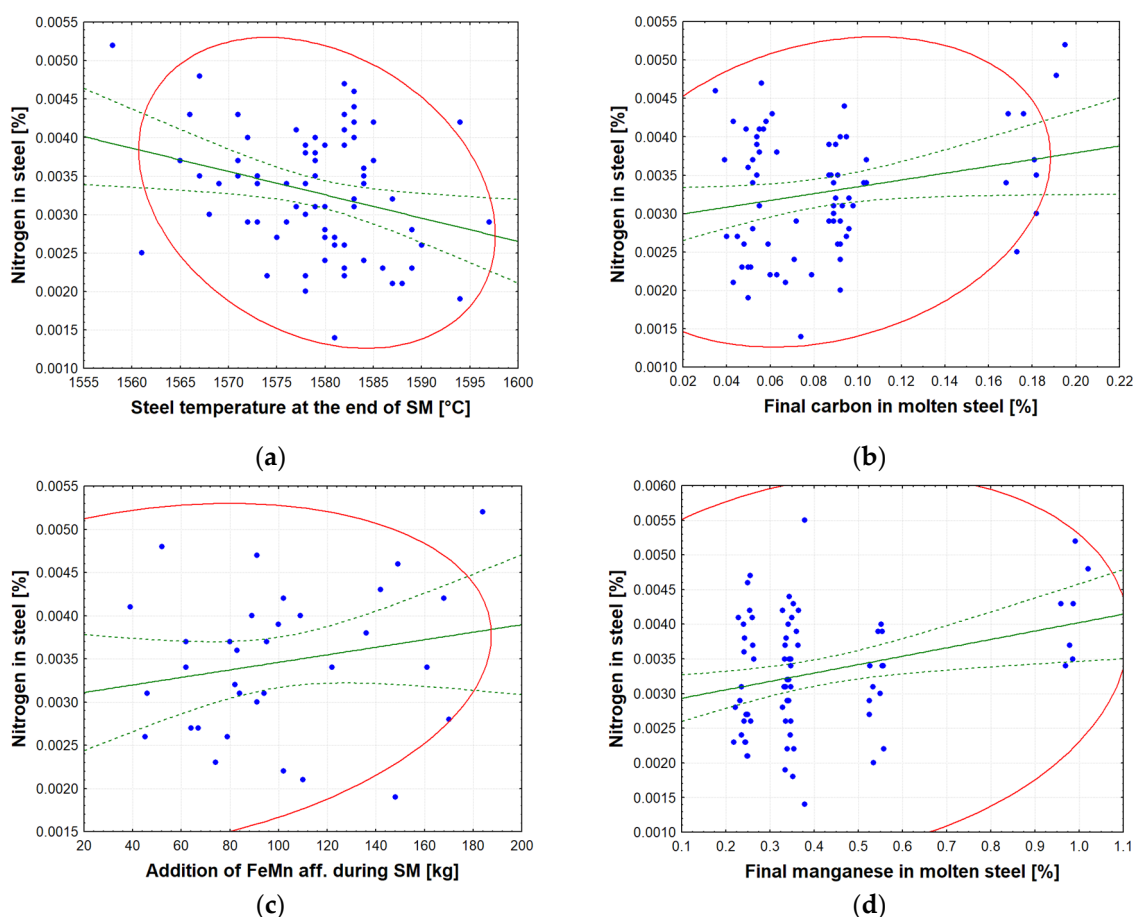
Figure 10. (**a**) The effect of molten steel temperature at the end of secondary metallurgy (SM) on nitrogen content in molten steel; (**b**) the effect of final carbon content in molten steel at the end of SM on nitrogen content in molten steel; (**c**) the effect of the added amount of FeMn aff. (affine) during SM on nitrogen content in molten steel; (**d**) the effect of final manganese content in molten steel at the end of SM on nitrogen content in molten steel.

The conditions for statistical testing of nitrogen presence in steel at the end of secondary metallurgy processing are shown in Table 15.

Table 15. Conditions for statistical testing of nitrogen presence in steel at the end of secondary metallurgy.

Entry Requirement	Value
Significance level, α	0.05 *
Number of observations, n	72
Number of parameters, m	4

* Unless otherwise stated.

Tables 16 and 17 show the results of a regression analysis of the individual factors (variables) listed in Table 14. The tables confirm that the correct results were obtained using the OLS method.

Table 16. Results of a modern regression analysis (OLS) of the factors influencing the nitrogen content in steel at the end of the SM processing.

Coefficient	Value	Standard Error	t-Ratio	p-Value
d_0	0.0257714	0.0241686	1.066	0.2902
d_1	-1.45992×10^{-5}	1.52140×10^{-5}	-0.9596	0.3408
d_2	0.0126950	0.00645446	1.967	0.0534
d_3	3.08438×10^{-6}	1.63553×10^{-6}	1.886	0.0637
d_4	-0.00153525	0.000911391	-1.685	0.0968

Where: d_0 constant/intercept, d_1–d_4 coefficients of variables (from Table 14).

Table 17. Results of individual tests using modern regression analysis to identify the factors influencing the final nitrogen content in steel at the end of the SM processing.

Parameter	Value
Mean value of the dependent variable	0.003270
Sum of squared residuals	0.000033
R^2—Coefficient of Multiple Determination	0.241736
$F_{(4, 67)}$—F-test	2.682089
Standard deviation of the dependent variable	0.000790
Durbin–Watson test	2.029951

4. Discussion

4.1. Model for Predicting Nitrogen Content in Molten Desulphurized Pig Iron

Section 3.1 contains Table 2, which shows the order of influence of the significant factors that cause metal saturation by nitrogen during the desulfurization phase of pig iron. The amount of nitrogen in pig iron after desulfurization is most influenced by the amount of sulfur removed from the pig iron. Sulphur is a strong surface-active element in pig iron, occupying active sites at the metal–gas interface and slowing down the decomposition of molecular nitrogen $\{N_2\}$ from the carrier gas. During desulphurization, the activity of the sulfur in the metal decreases, freeing up reaction sites at the phase interface and accelerating the dissolution of atomic nitrogen [N] into the metal. Therefore, the more sulfur that is removed from pig iron, the more nitrogen will dissolve in it. This is closely related to the amount of carrier gas used in the desulfurization mixture, which is nitrogen. Nitrogen enters the metal via an adsorption apparatus at the phase interface. Here, molecules are dissociated into their atomic form, which then dissolves into the metal volume. The amount of nitrogen supplied as a carrier gas is related to the amount of desulfurization mixture used.

As the amount of desulfurization mixture added to pig iron can vary within a given volume of carrier gas, these two factors both influence the oversaturation of pig iron by nitrogen. The weight of the pig iron is directly proportional to the amount of desulfurization mixture (for pig iron with a consistent sulfur content) and the volume of nitrogen used as the carrier gas. The difference in temperature at the beginning and end of pig iron desulfurization is significant in relation to nitrogen solubility in metal.

As shown in Table 5, the ordinary least squares (OLS) method yielded correct results. The mean value of the dependent variable, which, in this case, is the amount of nitrogen in the metal after desulphurization of pig iron, gives the mean value of the observed dependent variable in the processed dataset.

The sum of squares of residuals is the total of the squares of the differences between the measured and estimated values of the dependent variable. This sum must be minimal, a goal that has been achieved.

The coefficient of multiple determination for residuals is a measure of the degree of suitability of the regression. The values taken are in the range $\langle 0,1 \rangle$. The objective is to attain the maximum possible value of R^2. The interpretation of the obtained value of the coefficient of determination is dependent to a significant extent on the nature of the processed datasets. It is imperative to emphasize that non-stationary variables, which are characteristic of the metallurgical industry, were processed. When such non-stationary variables are obtained from integration systems (metallurgical aggregates, storage tanks, etc.), R^2 values of only around 0.15 are often achieved. However, this does not necessarily indicate a low degree of suitability of the proposed regression [44].

The F-test is a statistical procedure that is employed to ascertain whether the standard deviations of two datasets are equal [68]. The objective of this procedure is to ascertain whether the typical cases in the set of examined numbers differ from each other, i.e., to determine whether $\sigma_2^2 = \sigma_1^2$ primarily applies [69]. The critical value of the F distribution F_{crit} for a significance level of 10% (for a two-tailed test) is $F_{crit}(5, 70) = 1.931$. The null hypothesis has the form H_0: $\sigma_2^2 = \sigma_1^2$, and the alternative hypothesis has the form H_1: $\sigma_2^2 \neq \sigma_1^2$. It can be stated that, if $F_{crit}(5, 70) > F(5, 70)$, i.e., $1.931 > 1.5103$, then the null hypothesis can be accepted, whereby the standard deviations of the datasets at a given significance level are 90% similar to each other.

The Durbin–Watson autocorrelation test was used to determine the presence of autocorrelation, i.e., whether random components influence each other. At the relevant significance level, we test the null hypothesis (H_0: there is no autocorrelation) against the alternative hypothesis (H_1: autocorrelation is present). The test statistic can take values ranging from 0 to 4, with values around 2 indicating the absence of autocorrelation [70]. From the calculated value of the Durbin–Watson autocorrelation coefficient (DW: 1.641496), it can be concluded that the null hypothesis (H_0) of the absence of autocorrelation is accepted. Therefore, the random components are considered to be statistically independent at the relevant significance level (α). A special Granger–Newbold comparison was introduced to indicate spurious regression using Durbin–Watson statistics (DW). To indicate spurious regression, inequality in Equation (28) must be fulfilled; in other words, the multiple determination coefficient (R^2: 0.097373) must exceed the Durbin–Watson statistic value [71,72].

$$R^2 > DW \tag{28}$$

In the regression for which the results are displayed in Table 5, this inequality is not satisfied, which correctly indicates the absence of spurious regression. The time series cointegration analysis was employed to accurately and correctly distinguish spurious regression [72].

The augmented Dickey–Fuller test was employed to test for cointegration. The testing itself is based on the null hypothesis H_0: the variables are not cointegrated, or the alternative hypothesis H_1: the variables are cointegrated. The subsequent data (Figure 11) are the result of an investigation into the cointegration of the dataset in the Gretl 2025a program.

```
Step 8: testing for a unit root in uhat

Augmented Dickey-Fuller test for uhat
including one lag of (1-L)uhat
sample size 74
unit-root null hypothesis: a = 1

    model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
    1st-order autocorrelation coeff. for e: 0.002
    estimated value of (a - 1): -0.769082
    test statistic: tau_c(6) = -4.9592
    asymptotic p-value 0.0246
```

Figure 11. Output for the augmented Dickey–Fuller test of cointegration for model N_{DeS}.

As illustrated in Figure 11, the *p*-value is 0.0246. Given that this value is lower than the significance level $\alpha = 0.05$, it can be concluded that the null hypothesis H_0 must be rejected and the alternative hypothesis H_1 accepted. This indicates that the factors are cointegrated, i.e., they are non-stationary in themselves, but their linear dependencies are stationary. Consequently, spurious regression is not a possibility in mathematical modelling. Similarly, the significantly negative value of tau_c(6) = −4.9592 indicates strong evidence for rejecting the null hypothesis.

The parameters outlined above demonstrate the suitability of the proposed configuration of variables (Table 4). As demonstrated in Equation (23), it is possible to formulate a mathematical model to predict the nitrogen content in desulphurized pig iron (N_{DeS}). The resulting form of the model is as follows (29):

$$N_{DeS} = \begin{aligned} &0.000105058 + 0.0220528 \cdot A_1 + 2.5678 \times 10^{-8} \cdot A_2 + 2.50746 \times 10^{-8} \cdot A_3 \\ &- 2.35159 \times 10^{-6} \cdot A_4 + 4.96884 \times 10^{-6} \cdot A_5 \end{aligned} \tag{29}$$

where:

N_{DeS}: predicted nitrogen content in desulphurized pig iron;

A_1: amount of sulfur removed [%];

A_2: amount of blown nitrogen as carrier gas for the desulphurization mixture [l];

A_3: weight of pig iron after desulfurization [kg];

A_4: amount of desulphurization mixture [kg];

A_5: temperature difference of pig iron before and after desulfurization [°C].

Validity ranges of the model (29) are listed in Table 18.

The proposed model was subjected to a rigorous diagnostic process, which involved the utilization of precise testing analyses and graphical tools. This study employed a range of statistical techniques, including tests for normality, heteroscedasticity, multicollinearity, and autocorrelation, to analyze the data. The analysis of the model was facilitated by the utilization of graphical representations, including scatter plots, line graphs, and histograms. The evaluation of the model is based on residuals, which represent the difference between the measured and predicted values of the amount of nitrogen in the desulphurized pig iron. Figure 12a presents a graphical representation of the residual variance, while Figure 12b provides a residual analysis of the timeline.
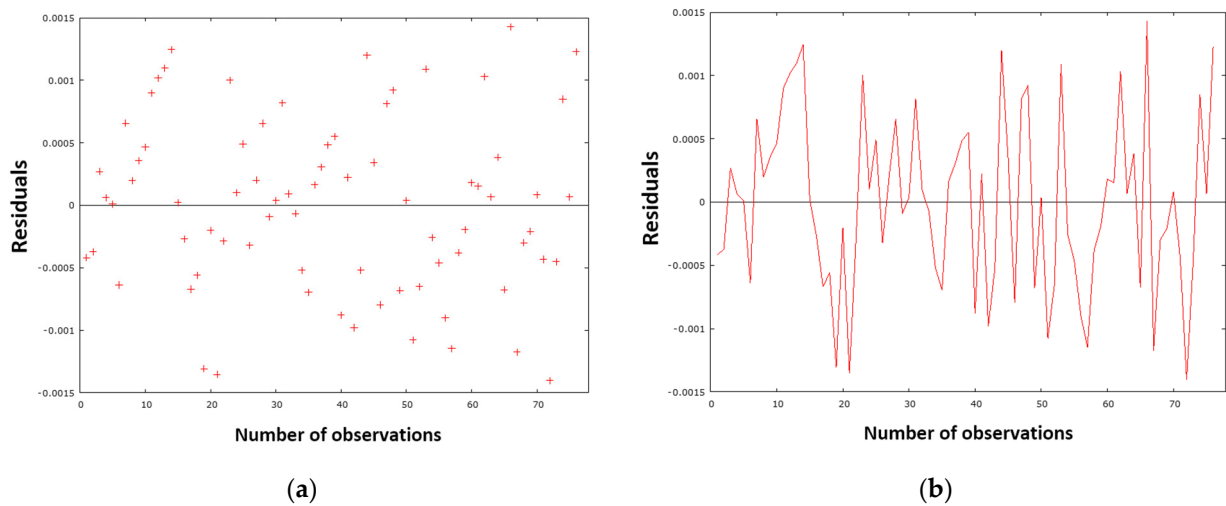
Figure 12. (**a**) Residual dispersion for the processed dataset and model N_{DeS} (29); (**b**) the residual course in the timeline for the processed dataset and model N_{DeS} (29).

As demonstrated in Figure 12a, the residual deviations are randomly dispersed around zero, and no potential trend or pattern can be observed in the graph. Consequently, the model is well designed and meets the assumptions. Figure 12b clearly shows that the sign of the residual values changes sufficiently over time. This finding suggests that the designed model generally does not overestimate or underestimate the calculated nitrogen values in metal when making predictions. The sum of squares of residuals (Table 5) with a value of 0.000036 confirms this opinion, as the value is close to zero. This fact—the normality of the residuals—is also evident from Figure 13, in which the red dots are arranged close to the blue line, demonstrating the normal distribution of the residuals, which is also confirmed by the histogram in Figure 14.
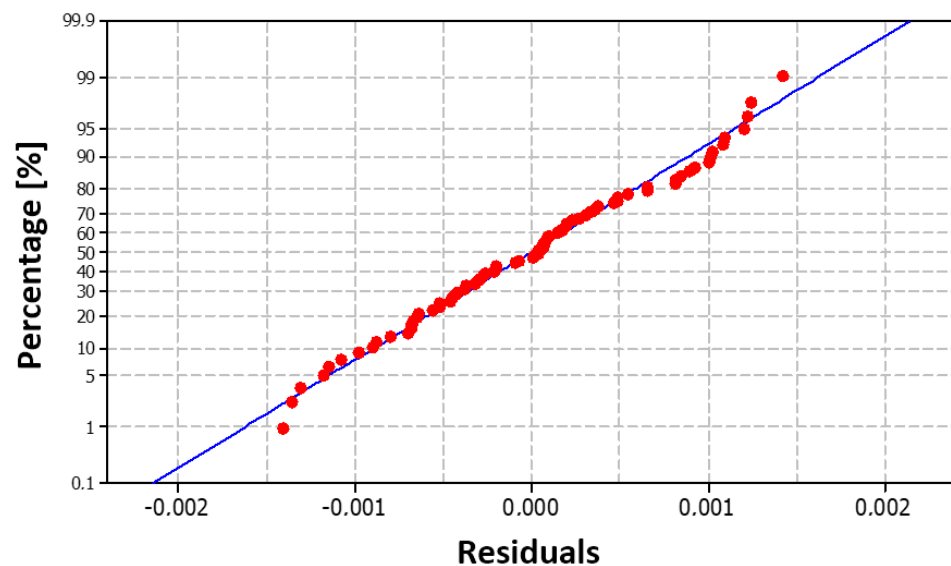
Figure 13. Graph of the normal distribution of residuals for the model N_{DeS} (29).

Table 18. Validity range of the N_{DeS} model (29) for predicting the amount of nitrogen in desulphurized pig iron.

Parameter	Range of Values
Amount of sulfur removed [%]	0.017–0.082
Amount of blown nitrogen as carrier gas for desulphurization mixture [l]	373–14,586
Weight of pig iron after desulfurization [kg]	139,200–145,500
Amount of desulphurization mixture [kg]	140–625
Temperature difference of pig iron before and after desulfurization [°C].	Temp. lowering by 38 °C Temp. increase of 11 °C
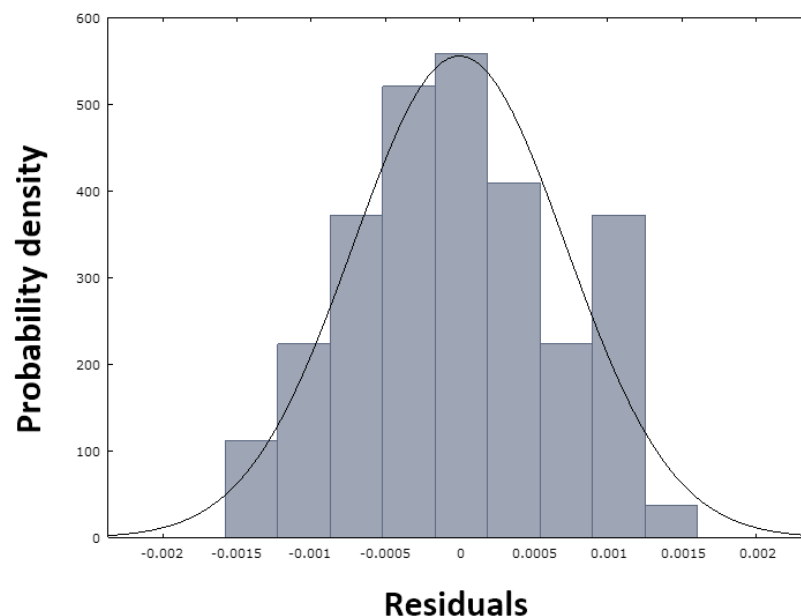
Figure 14. Histogram of the normal distribution of residuals for model N_{DeS} (29).

Figure 15 provides a graphical representation of the comparison between measured and predicted results using the N_{DeS} model (29). The red curve signifies the measured values, whilst the blue curve denotes the predicted values of nitrogen content in pig iron following desulfurization. The green curves represent the 95% confidence interval. The standard deviation of the residuals is 0.000718051. This low value indicates minimal discrepancies between the measured and predicted datasets. Consequently, the N_{DeS} model (29) provides accurate results.

To detect incorrect specifications in the proposed model, heteroscedasticity tests are used to test the non-constant variance of random components [73]. The tests verify whether any significant variables have been omitted from the mathematical model [74]. White's heteroscedasticity test and Breusch–Pagan's heteroscedasticity test were used for testing. For both tests, the null hypothesis H_0 applies: no heteroscedasticity, as opposed to the alternative hypothesis H_1: heteroscedasticity present. The test statistic for White's test of heteroscedasticity is 15.0079. The null hypothesis is rejected if the value of White's test statistic is greater than the critical value $\chi^2(20)$ at the corresponding confidence level α. However, this is not the case, because $\chi^2(20) = 31.41 > 15.0079$. Consequently, the null hypothesis H_0 regarding the absence of heteroscedasticity is accepted. The Breusch–Pagan test statistic for heteroscedasticity is 3.00616. The null hypothesis H_0 is rejected if the value of the Breusch–Pagan statistic is greater than the corresponding critical chi-squared value $\chi^2(5)$ at the chosen confidence level α. However, this is not the case because $\chi^2(5) = 11.07 > 3.00616$. Therefore, we accept the null hypothesis (H_0) of the absence of

heteroscedasticity. Both tests confirmed the absence of heteroscedasticity, and thus the variance of random components exhibits homoscedasticity, meaning that no significant variable is omitted from the proposed model.
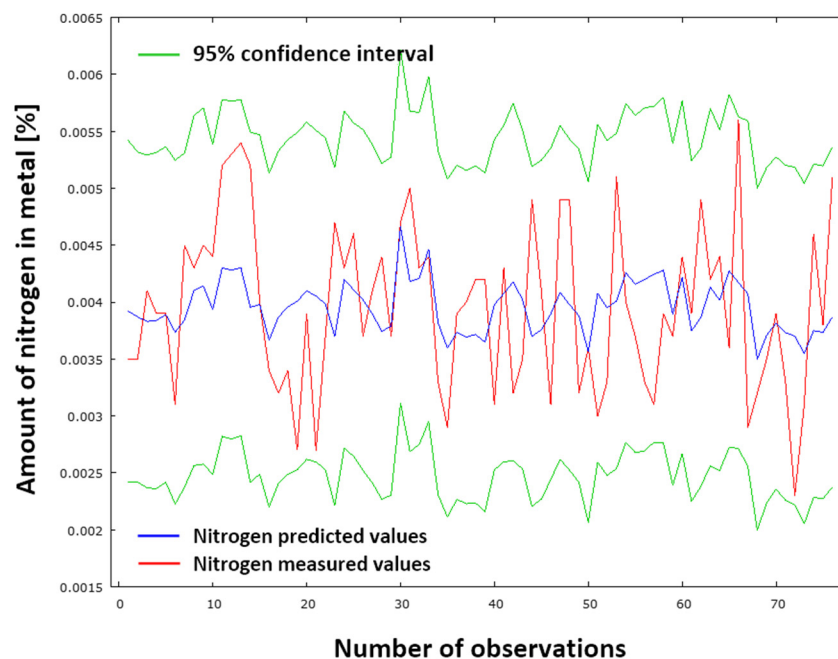
Figure 15. Comparison of measured and predicted nitrogen values according to model N_{DeS} (29).

Multicollinearity testing is used to verify the suitability of the factors affecting the amount of nitrogen in metal. In a multiple regression model, multicollinearity assesses the extent to which two or more prognostic factors are correlated. If there is a high degree of correlation, even a small change in the dataset could result in a significant change in the estimated strength of the coefficient. However, multicollinearity does not reduce the model's overall predictive power and reliability, only affecting the calculations relating to individual predictors [75]. Multicollinearity is assessed by the Variance Inflation Factor (VIF) [76], whose values for the analyzed factors a_1–a_5 (Table 4) are provided in Table 19. The minimum value is 1, with values above 10 indicating high multicollinearity.

Table 19. Results of the VIF test for the factors that influence the nitrogen content of desulphurized pig iron.

Parameter	Value
Amount of sulfur removed	3.184
Amount of blown nitrogen as carrier gas for desulphurization mixture	1.127
Weight of pig iron after desulfurization	1.258
Amount of desulphurization mixture	1.089
Temperature difference of pig iron before and after desulfurization	3.468

Time series extrapolation produces forecasts based on estimates of the parameters of a specific mathematical model whose quality has been confirmed by various statistical tests. Therefore, it can be expected that the resulting forecasts will not differ greatly from reality. The accuracy of the forecasts is assessed using various average characteristics.

The Mean Absolute Error (MAE) is a measure of the average absolute deviation of actual values from estimated (predicted) values. Following the substitution of the given

Equation (24), the resulting statistic is found to be $\text{MAE}_{\text{DeS}} = 1.3374 \times 10^{-10}$. Consequently, it can be deduced that the Mean Absolute Error is negligible.

The Mean Percentage Error (MPE) expresses the degree of distortion. After substituting into Equation (25), the result $\text{MPE}_{\text{DeS}} = -3.5785\%$ was obtained. It can, therefore, be concluded that the proposed model systematically overestimates reality, with the predicted values being, on average, 3.5785% higher than the actual values.

The Mean Absolute Percentage Error (MAPE) is a statistical metric that calculates the average magnitude of forecast errors as a percentage relative to actual values over the entire forecast period. Following the substitution of the observed statistic into Equation (26), the result is $\text{MAPE}_{\text{DeS}} = 16.5411\%$.

The accuracy of the N_{DeS} model (29) is determined through the substitution of the model's parameters into Equation (27). Subsequently, Equation (30) can be established, thereby determining the accuracy of the N_{DeS} model, as outlined in Equation (31).

$$\text{Model accuracy}: \ N_{\text{DeS}} = 100 - \text{MAPE}_{\text{DeS}} \tag{30}$$

$$\text{Model accuracy}: N_{\text{DeS}} = 83.4589\% \tag{31}$$

4.2. Model for Predicting Nitrogen Content in Molten Crude Steel Before Tapping from BOF

Section 3.2 and Table 6 show the ranking of factors affecting the amount of dissolved nitrogen in crude steel before it is tapped from the BOF. The results can be interpreted as follows: The most significant factor causing increased nitrogen content in crude steel is the total oxygen reblow time, which correlates very well with operational reality. The nitrogen content of high-purity oxygen (oxygen purity level of 95%) can vary significantly, from 70 to 1250 ppm of nitrogen. Reblow is performed either due to an inadequate chemical composition or a low temperature of the crude steel, and this increases the amount of nitrogen dissolved in the metal. The content of manganese, phosphorus, and carbon relates to the amount removed during the refining process. The greater the removal of these elements during heat, the greater the nitrogen content of the crude steel, which is closely related to the blowing time, amount of high-purity oxygen supplied, and intensity of oxygen blowing. The effect of the amount of briquettes added on the increase in dissolved nitrogen in the produced crude steel is related to its binder, molasses. Molasses is used in the production of briquettes at a proportion of up to 10 wt.%. Sugar beet molasses contains nitrogen in its structure and has the chemical formula $C_6H_{12}NNaO_3S$ [77]. Tapping temperature affects the amount of nitrogen dissolved in crude steel. More nitrogen is dissolved at higher tapping temperatures. This is because the solubility of nitrogen in molten steel increases with temperature.

Based on Table 9, it can be observed that correct results were achieved using the OLS method. The standard deviation of the dependent variable (0.000614) when compared to the mean value of the dependent variable (0.002002) yielded a coefficient of variation of approximately 0.31 (SD/Mean). This indicates that the variability of the dependent variable is to a considerable extent governed, and it is estimated to account for approximately 31% of its mean value.

As demonstrated in Table 1, the coefficient of determination, according to Cohen's distribution, manifests only moderate values. However, it is imperative to underscore the non-stationary nature of data from integrated systems and the fact that these are operational data, for which achieving average correlation coefficients is a substantial accomplishment.

The sum of squares of residuals is 0.000017. A low value indicates that the absolute errors of the model are very small, which is a positive sign for the accuracy of predictions.

The F-test tests whether the null hypothesis H_0 primarily applies, i.e., $\sigma_2^2 = \sigma_1^2$, or if the alternative hypothesis applies, i.e., H_1: $\sigma_2^2 \neq \sigma_1^2$. The critical value of the F-

distribution for a significance level of 10% is $F_{crit}(7, 55) = 1.829$. Since $F_{crit}(7, 55) < F(7, 55)$, i.e., $1.829 < 2.326593$, it can be concluded that the null hypothesis can be rejected, meaning that the standard deviations of the datasets are 90% different from each other.

The Durbin–Watson statistic value of 2.000830 is almost ideal. A value close to 2.0 indicates the absence of autocorrelation in the model residuals, thus supporting the null hypothesis H_0 concerning the absence of autocorrelation. This is a highly positive finding, as it fulfils one of the fundamental assumptions of linear regression, namely, the independence of errors. In Granger–Newbold's comparison of spurious regression, it is possible to conclude, on the basis of Equation (28), that, in this case, there is no indication of spurious regression, because the value of the coefficient of determination R^2 is lower than the value of the DW test.

The augmented Dickey–Fuller test was applied to assess cointegration. This test evaluates the null hypothesis, H_0: the variables are not cointegrated, against the alternative hypothesis, H_1: the variables are cointegrated. Figure 16 presents the results of this cointegration analysis, carried out using the statistical tool Gretl 2025a.

```
Step 10: testing for a unit root in uhat

Augmented Dickey-Fuller test for uhat
including one lag of (1-L)uhat
sample size 61
unit-root null hypothesis: a = 1

   model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
   1st-order autocorrelation coeff. for e: -0.005
   estimated value of (a - 1): -1.03311
   test statistic: tau_c(8) = -5.61958
   asymptotic p-value 0.01596
```

Figure 16. Output for the augmented Dickey–Fuller test of cointegration for model N_{BOF}.

As depicted in Figure 16, the *p*-value is 0.01596, which is below the significance level ($\alpha = 0.05$), leading us to reject the null hypothesis (H_0) in favor of the alternative (H_1). Consequently, the series are cointegrated: each is non-stationary on its own, but their linear combination is stationary. Accordingly, spurious regression cannot occur in the mathematical model. Moreover, the markedly negative value of tau_c(8) = -5.61958 offers strong evidence for rejecting the null hypothesis.

The test parameters demonstrate the suitability of the configuration of variables listed in Table 8. Equation (23) can be used to create a mathematical model for predicting the nitrogen content of raw steel before tapping it from the basic oxygen furnace. The resulting model takes the form of Equation (32).

$$
\begin{aligned}
N_{BOF} = \quad & -0.0137962 + 7.06124 \times 10^{-6} \cdot B_1 - 0.000306035 \cdot B_2 - 0.0469186 \cdot B_3 \\
& - 0.00499063 \cdot B_4 - 3.44051 \times 10^{-8} \cdot B_5 + 1.04457 \times 10^{-5} \cdot B_6 \\
& - 3.96352 \times 10^{-7} \cdot B_7
\end{aligned} \tag{32}
$$

where:

N_{BOF}: predicted nitrogen content in crude steel before tapping from BOF;

B_1: oxygen reblow time [s];

B_2: manganese content in crude steel [%];

B_3: phosphorus content in crude steel [%];

B_4: carbon content in crude steel [%];

B_5: briquettes [kg];

B_6: temperature of tapping steel [°C];

B$_7$: oxygen blowing time [s].

The validity ranges of the model (32) are exhibited in Table 20.

Table 20. Validity range of the N$_{BOF}$ model (32) for predicting the amount of nitrogen content in crude steel prior to tapping from BOF.

Parameter	Range of Values
Oxygen reblow time [s]	0–100
Manganese content in crude steel [%]	0.060–0.244
Phosphorus content in crude steel [%]	0.005–0.015
Carbon content in crude steel [%]	0.026–0.105
Briquettes [kg]	0–3260
Temperature of tapping steel [°C]	1620–1685
Oxygen flowing time [s]	1601–2322

The proposed N$_{BOF}$ model (32) followed the same rigorous diagnostic process as the N$_{DeS}$ model (29). This process involved the use of precise testing analyses and graphical representations. The evaluation of the model is grounded in the assessment of residuals. Residuals serve to illustrate the discrepancy between the measured value and the predicted value of the amount of nitrogen in the crude steel prior to tapping. As illustrated in Figure 17a, the residual variance is presented graphically, while Figure 17b offers a residual analysis with respect to the timeline.

(a)

(b)
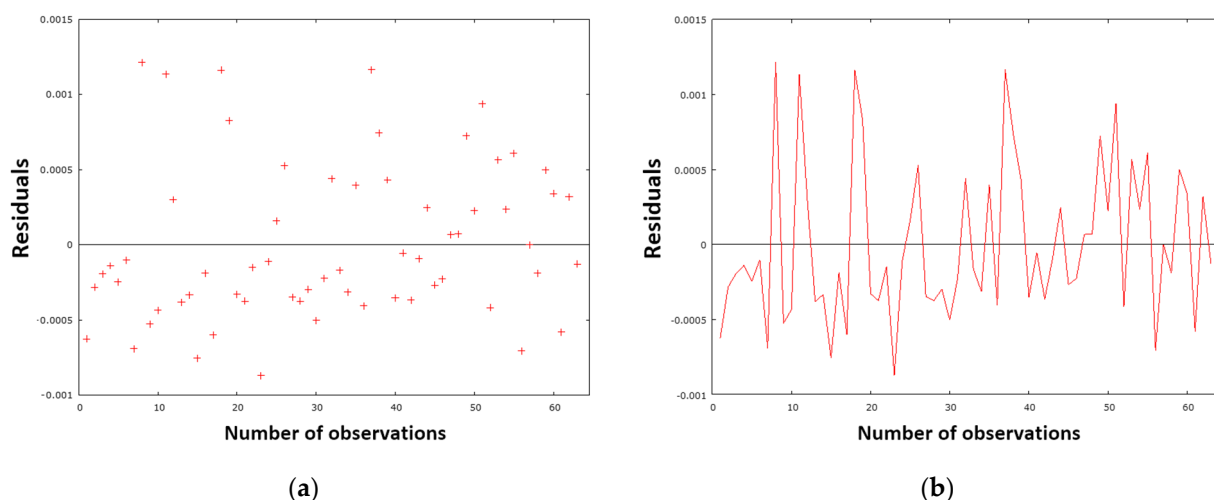
Figure 17. (**a**) Residual dispersion for the processed dataset and model N$_{BOF}$ (32); (**b**) the residual course in the timeline for the processed dataset and model N$_{BOF}$ (32).

As illustrated in Figure 17a, the residual deviations are randomly dispersed around zero. Furthermore, the graph does not exhibit any discernible trend or pattern. Consequently, the model is well designed and meets the assumptions. As demonstrated in Figure 17b, the sign of the residual values undergoes a substantial alteration over the course of the experiment. This finding suggests that the designed model generally does not significantly overestimate or underestimate the calculated nitrogen values in metal when making predictions. The calculation of the sum of squares of residuals (Table 9) yields a value of 0.000017, which corroborates this viewpoint, as it is proximate to zero. This finding—that is, the normality of the residuals—can be seen in Figure 18, where the points (red dots) can be seen to be arranged quite close to the blue line. This arrangement demonstrates a normal distribution of the residuals. This finding is also confirmed by the histogram in Figure 19.
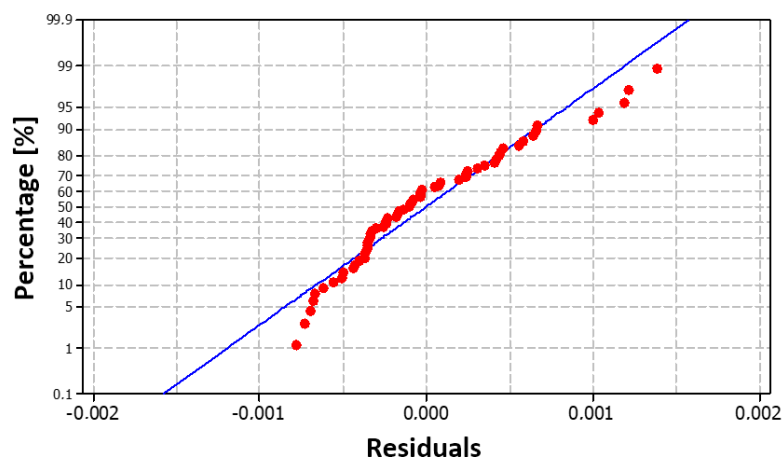
Figure 18. Graph of the normal distribution of residuals for the model N_{BOF} (32).
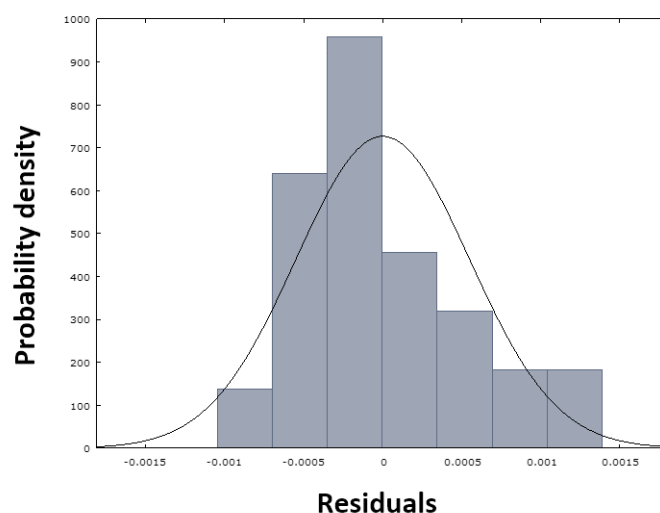

Figure 19. Histogram of the normal distribution of residuals for model N_{BOF} (32).

As illustrated in Figure 20, a graphical representation is provided of the comparison between measured and predicted results using the N_{BOF} model (32). The red curve represents the measured values, whilst the blue curve denotes the predicted values of nitrogen content in crude steel prior to tapping. The green curves represent the 95% confidence interval. The standard deviation of the residuals is calculated to be 0.000548932. The observed value, which is of negligible magnitude, signifies that there is minimal discrepancy between the two datasets, i.e., the measured and predicted ones, respectively. Consequently, it can be posited that the N_{BOF} model (32) provides accurate results.

The White and Breusch–Pagan tests were used to assess the presence of heteroscedasticity, and to verify the null hypothesis (H_0), which assumes the absence of heteroscedasticity, and the alternative hypothesis (H_1), which assumes its presence. The White test yielded a value of 34.3269. The null hypothesis is rejected if the value of the test statistic is greater than the corresponding critical value, $\chi^2(34)$, at the chosen confidence level, α. However, this is not the case because $\chi^2(34) = 48.602 > 34.3269$. Therefore, in White's test, the null hypothesis of no heteroscedasticity is accepted. The Breusch–Pagan test statistic for heteroscedasticity is 8.66205. The null hypothesis (H_0) is rejected if the value of the Breusch–Pagan test statistic is greater than the corresponding critical $\chi^2(7)$ value at the chosen confidence level α. However, this is also not the case here, as $\chi^2(7) = 14.067 > 8.66205$. According to the Breusch–Pagan statistic, the null hypothesis (H_0): no heteroscedasticity is accepted.
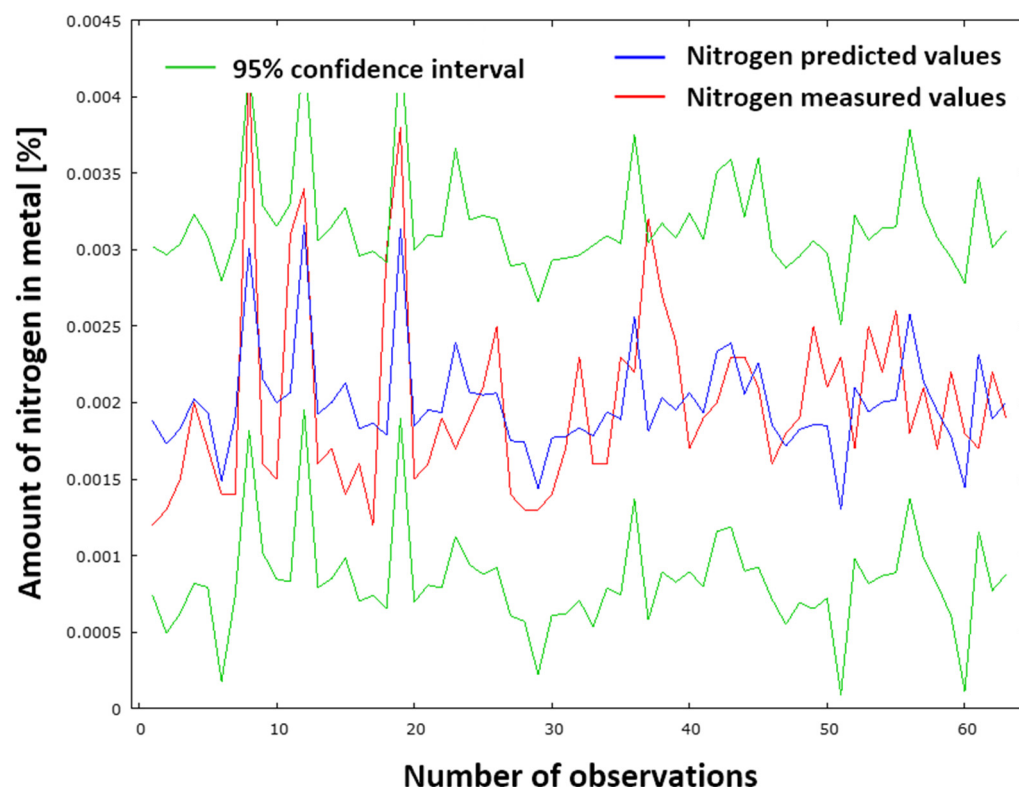
Figure 20. Comparison of measured and predicted nitrogen values according to model N_{BOF} (32).

As part of the solution to multicollinearity, the variance inflation factor (VIF) is evaluated. The values for the analyzed factors b_1–b_7 (Table 8) are shown in Table 21. The test results indicate low to no multicollinearity (1 is the minimal value). This indicates that the independent variables are well separated, meaning that each variable contributes unique information to the model. Based on the VIF test, the proposed N_{BOF} (32) model also shows stability, with reliable regression coefficients that are unaffected by excessive correlation between variables.

Table 21. Results of the VIF test for the factors that influence the nitrogen content of crude steel prior to tapping.

Parameter	Value
Oxygen reblow time	1.166
Manganese content in crude steel	2.609
Phosphorus content in crude steel	2.540
Carbon content in crude steel	1.582
Briquettes	1.184
Temperature of tapping steel	1.550
Oxygen flowing time	1.075

The accuracy of the N_{BOF} model (32) can be evaluated as follows. The Mean Absolute Error (MAE) was calculated using Equation (24), and the result is $MAE_{BOF} = 1.3606 \times 10^{-10}$. It can therefore be concluded that the average absolute error is very small.

The Mean Percentage Error (MPE) was computed based on Equation (25), and, after substituting into the relationship, the result is $MPE_{BOF} = -6.1515\%$. It can therefore be concluded that reality is systematically overestimated by the proposed model, with the predicted values being 6.1515% higher than the actual values on average.

The average size of forecast errors compared to actual values across the entire forecast period is expressed using the Mean Absolute Percentage Error (MAPE). Substituting into Equation (26), we obtain the result $MAPE_{BOF}$ = 22.7696%. The accuracy of the N_{BOF} model (32) can be determined by substituting the model's parameters into Equation (27). This establishes Equation (33), which determines the accuracy of the NBOF model as outlined in Equation (34).

$$\text{Model accuracy}: \ N_{BOF} = 100 - MAPE_{BOF} \tag{33}$$

$$\text{Model accuracy}: N_{BOF} = 77.2304\% \tag{34}$$

4.3. Model for Predicting Nitrogen Content in Molten Steel at the Beginning of Secondary Metallurgy

It has been shown in Section 3.3 (Table 10) that, as the tapping angle is increased, the nitrogen content in molten steel is reduced. This dependency is associated with the length of the tapping steel stream. With a smaller BOF vessel tilt, the length of the tapped steel stream is greater, meaning more tapped steel comes into contact with the atmosphere, creating a larger reaction area. As the converter tilt increases, the steel flowing into the ladle is straighter and shorter, resulting in a smaller reaction surface. The duration of the tapping time is also found to be significantly related to the length of time the steel is in contact with air (79% of air consists of nitrogen). It has been demonstrated that an increase in the duration of the tapping time results in an increase in the nitrogen content dissolved in the steel, due to the prolongation of the steel's exposure to the air. At the beginning of secondary metallurgy, silicon in steel comes from the FeSi ferroalloy. This ferroalloy is added to steel only after a carbonized deoxidizer or carburizing coke is added. This ensures the boiling of the steel and the generation of a large amount of CO bubbles, which subsequently generate CO_2. This reduces the amount of active oxygen at the metal–gas interface, enabling the metal to become supersaturated with atmospheric nitrogen during intense steel boiling. After FeSi is added, the silicon also reacts with the active oxygen in the metal to form SiO_2, which increases the nitrogen transfer coefficient into the metal. Depending on the manufacturer, FeSi contains approximately 80–150 ppm of nitrogen. Moreover, Wagner's interaction coefficient for the Fe–Si–N system is $e_{Si}^N = 0.047$. A positive value indicates that silicon increases the activity coefficient of nitrogen and thus also the equilibrium solubility of nitrogen in molten steel. At the beginning of secondary metallurgy (SM), manganese comes from both the crude steel produced in the BOF and the FeMn aff. alloy added during the SM process. Manganese increases the solubility of nitrogen in steel. Similarly, the FeMn aff. ferroalloy contains 40–80 ppm of nitrogen, depending on the supplier. Wagner's interaction coefficient for the Fe–Mn–N system is $e_{Mn}^N = 0.013$. A positive value indicates that manganese increases the activity coefficient of nitrogen and thus also the equilibrium solubility of nitrogen in molten steel. Oxygen is a highly active element on the surface of metal and occupies active sites at the metal–gas phase interface. Therefore, oxygen in crude steel slows down the dissolution of nitrogen in the metal. Adding a large amount of aluminum in the form of blocks significantly reduces the activity of oxygen in the metal. This reduces the amount of oxygen at the metal–gas interface and increases the nitrogen transfer coefficient, allowing nitrogen to dissolve into the metal and increasing its content. This is the reason why fully-killed steels have a higher nitrogen content than semi-killed steels, as deoxidation removes more oxygen and requires more added aluminum as a deoxidizer. During this process, the metal is mixed intensively and comes into contact with air. This is why the nitrogen that enters the metal during aluminum-based deoxidation comes from the atmosphere. For steel grades that require a very low final nitrogen value, deoxidation using aluminum is performed during

processing at SM with chopped aluminum wire rather than aluminum blocks during tapping from BOF. The efficiency of deoxidation using chopped aluminum wire is 85–92%. Wagner's interaction coefficient for the Fe–Al–N system is $e_{Al}^{N} = -0.017$. A negative value indicates that aluminium decreases the activity coefficient of nitrogen, thereby reducing the equilibrium solubility of nitrogen in molten steel. Therefore, adding 0.03% aluminium to the metal reduces the equilibrium nitrogen content by approximately 3–4 ppm at 1600 °C.

The results presented in Table 13 confirm the validity of the ordinary least squares (OLS) estimation approach. The coefficient of variation, calculated as the ratio of the standard deviation (0.000888) to the mean value (0.003160) of the dependent variable, equals approximately 0.28 (SD/Mean). This indicates that the dispersion of the dependent variable is substantially controlled, with the variability representing approximately 28% of the mean value.

As shown in Table 1, according to Cohen's distribution, the coefficient of determination only exhibits moderate values. However, it is important to emphasize the non-stationary nature of the data from integrated systems, and the fact that these are operational data for which achieving average correlation coefficients is a significant achievement.

The sum of the squares of the residuals is 0.000035. This low value indicates that the model's absolute errors are very small, suggesting that predictions will be accurate.

The F-test determines whether the null hypothesis (H$_0$: $\sigma_2^2 = \sigma_1^2$) or the alternative hypothesis (H$_1$: $\sigma_2^2 \neq \sigma_1^2$) applies. The critical value of the F-distribution for a 10% significance level is $F_{crit}(7, 67) = 1.808$. As $F_{crit}(7, 67) < F(7, 67)$, i.e., $1.808 < 6.453$, the null hypothesis can be rejected. This means that the standard deviations of the datasets are 90% different from each other.

The Durbin–Watson statistic of 1.872243 (Table 13) provides evidence for the absence of autocorrelation among model residuals, thereby supporting the null hypothesis H$_0$ regarding the independence of error terms. This finding is particularly significant as it satisfies a fundamental assumption underlying linear regression analysis, specifically the requirement for error independence. According to the Granger–Newbold criterion in Equation (28), no spurious regression is detected since $R^2 < DW$, confirming model validity.

The augmented Dickey–Fuller test was employed to examine cointegration relationships among the variables. The test framework evaluates the null hypothesis H$_0$ (absence of cointegration) against the alternative hypothesis H$_1$ (presence of cointegration). The cointegration analysis results, conducted using the statistical software Gretl 2025a, are presented in Figure 21.

```
Step 10: testing for a unit root in uhat

Augmented Dickey-Fuller test for uhat
including one lag of (1-L)uhat
sample size 73
unit-root null hypothesis: a = 1

    model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
    1st-order autocorrelation coeff. for e: 0.006
    estimated value of (a - 1): -0.977964
    test statistic: tau_c(8) = -6.0245
    asymptotic p-value 0.003933|
```

Figure 21. Output for the augmented Dickey–Fuller test of cointegration for model N$_{SMB}$.

As illustrated in Figure 21, the *p*-value is 0.003933, which is significantly below the significance level ($\alpha = 0.05$). This indicates that the null hypothesis (H$_0$) is rejected and the alternative (H$_1$) is accepted. Consequently, the series are cointegrated: each is non-

stationary on its own, but their linear combination is stationary and spurious regression is precluded in the mathematical model. Furthermore, the markedly negative value of tau_c(8) = −6.0245 provides substantial evidence to support the rejection of the null hypothesis—series are cointegrated.

The test parameters demonstrate the suitability of the configuration of variables listed in Table 10. Equation (23) can be utilized to formulate a mathematical model for predicting the nitrogen content in steel at the beginning of secondary metallurgy. The resulting model assumes Equation (35).

$$
\begin{aligned}
N_{SMB} = \ & 0.00982567 - 7.89409 \times 10^{-5} \cdot C_1 + 0.00368566 \cdot C_2 - 0.00987549 \cdot C_3 + \\
& 0.00442257 \cdot C_4 - 0.00140863 \cdot C_5 + 6.15271 \times 10^{-7} \cdot C_6 + \\
& 6.28702 \times 10^{-6} \cdot C_7
\end{aligned}
\tag{35}
$$

where:

N_{SMB}: predicted nitrogen content in steel at the beginning of secondary metallurgy,
C_1: tapping angle [°];
C_2: silicon in molten steel prior to argon bubbling [%];
C_3: total aluminum prior to argon bubbling [%];
C_4: carbon in molten steel prior to argon bubbling [%];
C_5: manganese in molten steel prior to argon bubbling [%];
C_6: tapping time [s];
C_7: added aluminum blocks [kg].

The validity ranges of the model (35) are illustrated in Table 22.

Table 22. Validity range of the N_{SMB} model (35) for predicting the amount of nitrogen content in steel at the beginning of secondary metallurgy.

Parameter	Range of Values
Tapping angle [°]	98–117
Silicon in molten steel prior to argon bubbling [%]	0–0.403
Total aluminum prior to argon bubbling [%]	0.006–0.056
Carbon in molten steel prior to argon bubbling [%]	0.026–0.171
Manganese in molten steel prior to argon bubbling [%]	0.174–1.290
Tapping time [s]	249–683
Added aluminum blocks [kg]	200–350

The proposed N_{SMB} model (35) was subjected to the same rigorous diagnostic process as the N_{DeS} model (29) and N_{BOF} model (32). This process involved precise testing, analysis, and graphical interpretation. The model's evaluation is based on the analysis of residuals. Residuals illustrate the discrepancy between the measured and predicted amounts of nitrogen in the crude steel prior to tapping. Figure 22a illustrates the residual variance graphically, while Figure 22b provides a residual analysis over time.

As depicted in Figure 22a, the residual deviations exhibit a random distribution around zero without any observable systematic trend or pattern. This distribution confirms that the model is appropriately specified and satisfies the underlying statistical assumptions. As shown in Figure 22b, the sign of the residual values changes substantially over the course of the experiment. This suggests that the designed model generally does not significantly overestimate or underestimate the calculated nitrogen values in metal when making predictions. Calculating the sum of squares of residuals (Table 13) yields a value of 0.000035, which corroborates this viewpoint as it is close to zero. Figure 23 illustrates this finding, showing that the points (red dots) are arranged quite close to the blue line,

indicating the normality of the residuals. This demonstrates a normal distribution of the residuals. The histogram in Figure 24 also confirms this finding.

(a)   **(b)**
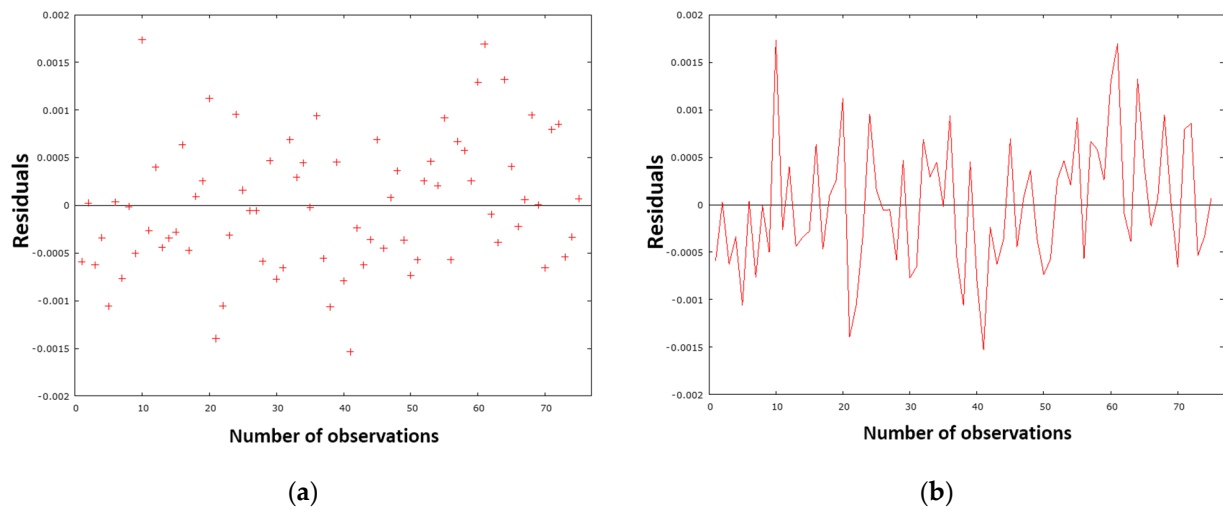
Figure 22. (**a**) Residual dispersion for the processed dataset and model N_{SMB} (35); (**b**) the residual course in the timeline for the processed dataset and model N_{SMB} (35).
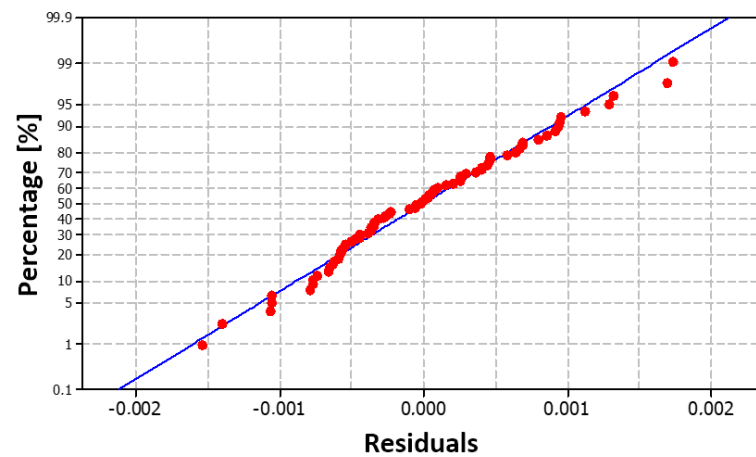

Figure 23. Graph of the normal distribution of residuals for the model N_{SMB} (35).
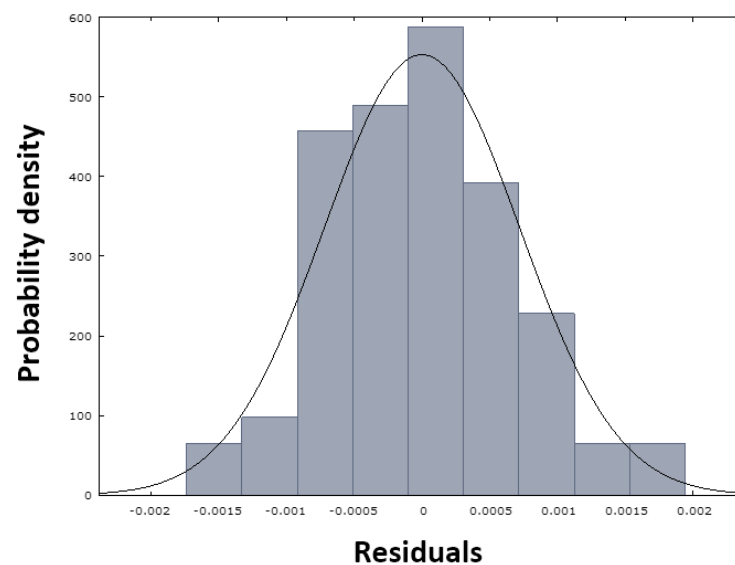
Figure 24. Histogram of the normal distribution of residuals for model N_{SMB} (35).

Figure 25 presents a comparative analysis of measured versus predicted values generated by the N_{SMB} model (35). The graphical representation displays measured values (red curve), predicted nitrogen concentrations in crude steel prior to tapping (blue curve), and the 95% confidence interval (green curves). The calculated standard deviation of residuals is 0.000721066, indicating minimal discrepancy between observed and predicted datasets. This negligible deviation demonstrates that the N_{SMB} model (35) exhibits satisfactory predictive accuracy.
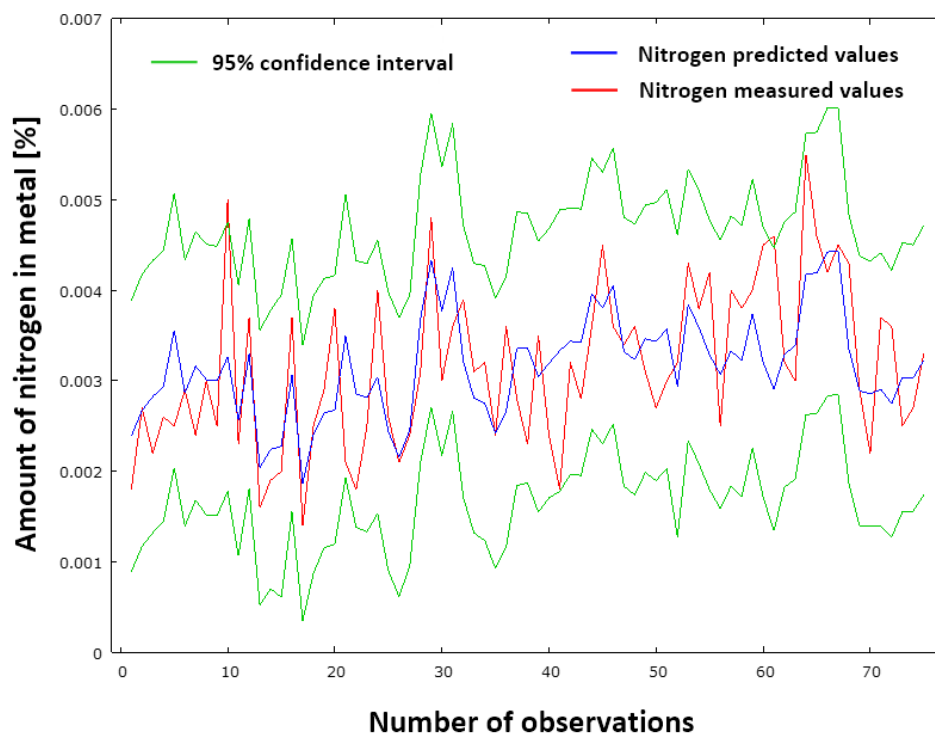
Figure 25. Comparison of measured and predicted nitrogen values according to model N_{SMB} (35).

The White and Breusch–Pagan tests were employed to evaluate the presence of heteroscedasticity and to verify the null hypothesis (H_0), which assumes the absence of heteroscedasticity, and the alternative hypothesis (H_1), which assumes its presence. The White test yielded a value of 42.766. The null hypothesis is to be rejected if the value of the test statistic is greater than the corresponding critical value, $\chi^2(35)$, at the chosen confidence level, $\alpha = 0.05$. However, this is not the case, because $\chi^2(35) = 49{,}802 > 42{,}766$. Consequently, in White's test, the null hypothesis of no heteroscedasticity is accepted. The Breusch–Pagan test statistic for heteroscedasticity is 4.94841. The null hypothesis (H_0) is rejected if the value of the Breusch–Pagan test statistic is greater than the corresponding critical $\chi^2(7)$ value at the chosen confidence level. However, this is not applicable in this instance, as $\chi^2(7) = 14.067 > 4.94841$. It is evident that, in accordance with the Breusch–Pagan statistic, the null hypothesis (H_0) is accepted, namely that there is no heteroscedasticity.

As part of the solution to multicollinearity, the variance inflation factor (VIF) is evaluated. The values for the analyzed factors c_1–c_7 (Table 12) are shown in Table 23. The VIF values indicate that the regression model is relatively favorable. Those that do not exhibit multicollinearity have values close to the ideal of 1, while variables such as carbon, manganese, and silicon in steel exhibit slight multicollinearity but do not exceed the critical VIF value of 10. The VIF values for carbon, manganese, and silicon indicate their correlated behavior. However, their higher VIF test values are not a shortcoming of the model but rather reflect actual metallurgical relationships. This correlation stems from their shared roles in steelmaking processes as they naturally form part of the chemical composition of

both raw iron and steel. Due to their similar affinity for oxygen at high temperatures, they react similarly with oxygen, are subject to similar thermodynamic laws in steel production and processing, and influence each other's final properties. In the context of steel finishing in secondary metallurgy, this correlation is both expected and technologically justified, confirming the accuracy of the statistical analysis observations. Intensive mixing of the steel during tapping from the Basic Oxygen Furnace (BOF), the boiling of the steel at the bottom of the ladle, and the addition of aluminum to deoxidize the steel significantly reduce the amount of nitrogen dissolved in the metal by transporting nitrogen to the metal–slag interface. In secondary steel metallurgy, the correlation between carbon, manganese, and silicon dissolved in steel is a natural phenomenon with fundamental practical significance. Understanding and utilizing this correlation enables more efficient process control, improves product quality in terms of nitrogen content, and generates economic savings. For modern steel producers, this correlation is an invaluable tool for optimizing production processes and ensuring consistent steel quality. Due to this, each variable provides unique information to the N_{SMB} model (35).

Table 23. Results of the VIF test for the factors that influence the nitrogen content in steel at the beginning of the secondary steelmaking.

Parameter	Value
Tapping angle	1.102
Silicon in molten steel prior to argon bubbling	4.056
Total aluminum prior to argon bubbling	1.133
Carbon in molten steel prior to argon bubbling	6.466
Manganese in molten steel prior to argon bubbling	6.760
Tapping time	1.289
Added aluminum blocks	1.211

The accuracy of the N_{SMB} model can be evaluated as follows: The Mean Absolute Error (MAE) was calculated using Equation (24), giving a result of $MAE_{SMB} = 2.407 \times 10^{-11}$. Therefore, it can be concluded that the average absolute error is very negligible.

The mean percentage error (MPE) was computed based on Equation (25). After substituting this into the relationship, the result is $MPE_{SMB} = -5.3582\%$. Therefore, it can be concluded that the proposed model systematically overestimates reality, with the predicted values being, on average, 5.3582% higher than the actual values.

The Mean Absolute Percentage Error (MAPE) is used to express the average size of forecast errors compared to actual values across the entire forecast period. Substituting this into Equation (26) gives $MAPE_{SMB} = 20.0341\%$. The accuracy of the N_{SMB} model can be determined by substituting its parameters into Equation (27). This establishes Equation (36), which determines the accuracy of the NBOF model, as outlined in Equation (37).

$$\text{Model accuracy}: \ N_{SMB} = 100 - MAPE_{SMB} \tag{36}$$

$$\text{Model accuracy}: \ N_{SMB} = 79.9659\% \tag{37}$$

4.4. Model for Predicting Nitrogen Content in Molten Steel at the End of Secondary Metallurgy

The most significant factors affecting the amount of nitrogen dissolved in molten steel (see Table 14) can be described as follows: The solubility of nitrogen in liquid steel is governed by Sievert's law, whereby the equilibrium solubility of nitrogen in steel increases with temperature. Despite thermodynamics predicting higher nitrogen solubility at higher temperatures, industrial observations demonstrate a contrary trend, whereby the ultimate nitrogen content in steel decreases with increasing temperature during secondary

metallurgy. This phenomenon can be attributed to the predominance of kinetic factors over thermodynamic equilibrium. At the beginning of the SM process, a significant number of CO bubbles are generated in the ladle, thereby assisting in the mixing of the melt. It has been demonstrated that at elevated temperatures, there is an increase in the volume of CO bubbles, and that the reaction [C] + [O] = {CO} proceeds at a faster rate [78]. However, at this stage, the elevated presence of surface–active elements, such as oxygen and sulfur, inhibits the process of rapid desulfurization. In the later stages of SM, when the generation of CO is reduced due to the depletion of reagents in the metal, argon assumes the role of the mixing agent. However, it has been demonstrated that, at elevated temperatures, the inhibitory effect of surface–active elements is reduced [79]. It has been demonstrated that, by reducing the amount of oxygen in the metal, it is possible to effectively remove nitrogen from the metal using the residual amount of CO bubbles in combination with argon [80]. The argon is fed into the metal through a porous plug located at the bottom of the casting ladle. As the temperature of the metal is increased, the viscosity of the steel is also reduced, thus facilitating the movement of CO and Ar bubbles. Experimental evidence has demonstrated that elevating the temperature from 1550 °C to 1620 °C enhances the saturation solubility of nitrogen, whereas concurrently increasing the rate constant for nitrogen removal and the mass transfer coefficient has also been observed [81]. During the process of deoxidation of steel tapped from BOF, a carbonized deoxidizer or carburizing coke is added. This process ensures that the steel boils and a large number of CO bubbles are generated. Subsequently, CO_2 is created. This reduces the amount of active oxygen at the metal–gas interface, enabling the metal to become supersaturated with atmospheric nitrogen during intense steel boiling. Adding FeMn aff. (not nitrogenous FeMnN) increases the nitrogen content of steel. The nitrogen in FeMn comes from atmospheric nitrogen that comes into contact with molten FeMn during the production process. Carousel tapping of FeMn creates a large reaction surface between the ferroalloy and the atmosphere, causing the absorption of large amounts of atmospheric nitrogen into the FeMn. The ferroalloy FeMn aff. has been found to contain nitrogen at concentrations ranging from 40 to 80 ppm, with variations depending on the supplier. The interaction coefficient for the Fe–Mn–N system, as determined by Wagner, is $e^N_{Mn} = 0.013$. A positive value indicates that manganese increases the activity coefficient of nitrogen, thus increasing the equilibrium solubility of nitrogen in molten steel. The final manganese content at the end of secondary metallurgy is closely related to the amount of FeMn added during the secondary metallurgy stage of steel processing.

The results presented in Table 17 confirm the validity of the ordinary least squares (OLS) estimation approach. The coefficient of variation, calculated as the ratio of the standard deviation (0.000790) to the mean value (0.003270) of the dependent variable, equals approximately 0.24 (SD/Mean). Consequently, the dataset displays low relative dispersion, indicating that variability constitutes only 24% of the central tendency. This supports the ordinary least squares assumptions, thereby confirming the methodological soundness of the OLS estimator and validating the reliability of the inferences derived from Table 17.

In the context of analyzing industrial data from steel production, the value of $R^2 = 0.241736$ can be considered acceptable. Industrial processes are characterized by high variability and complex interactions between process variables. Consequently, even a low value of R^2 can be informative and useful, especially if the regression coefficients are statistically significant and the result of the coefficient of determination can, therefore, be considered significant in terms of the nature of the data being processed.

The sum of the squares of the residuals is 0.000033. It is evident that the low value indicates that the model's absolute errors are minimal, thereby suggesting that predictions will be accurate.

The F-test is a statistical procedure used to determine whether the null hypothesis (H_0: $\sigma_2^2 = \sigma_1^2$) or the alternative hypothesis (H_1: $\sigma_2^2 \neq \sigma_1^2$) applies. The critical value of the F-distribution for a 10% significance level is $F_{crit}(4, 67) = 2.031$. As $F_{crit}(4, 67) < F(4, 67)$, i.e., $2.031 < 2.682089$, the null hypothesis can be rejected. This indicates that the standard deviations of the datasets differ by 90%.

With a value of 2.029951, the Durbin–Watson statistic is almost ideal. A value close to 2.0 indicates an absence of autocorrelation in the model residuals, thus supporting the null hypothesis (H_0) of an absence of autocorrelation. This is a highly positive finding as it fulfils one of the fundamental assumptions of linear regression: the independence of errors. According to Granger and Newbold's comparison of spurious regression, the relation in Equation (28) suggests that there is no indication of spurious regression in this case, as the value of the coefficient of determination R^2 is lower than the DW test value.

The augmented Dickey–Fuller test was used to analyze the cointegration relationships between the variables. This test evaluates the null hypothesis (H_0: absence of cointegration) against the alternative hypothesis (H_1: presence of cointegration). The results of the cointegration analysis, which was conducted using the Gretl 2025a statistical software, are presented in Figure 26.


```
Step 7: testing for a unit root in uhat

Augmented Dickey-Fuller test for uhat
including one lag of (1-L)uhat
sample size 70
unit-root null hypothesis: a = 1

    model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
    1st-order autocorrelation coeff. for e: -0.011
    estimated value of (a - 1): -0.650824
    test statistic: tau_c(5) = -4.559
    asymptotic p-value 0.03379
```

Figure 26. Output for the augmented Dickey–Fuller test of cointegration for model N_{SME}.

As shown in Figure 26, the *p*-value is 0.03379, which is below the significance level $\alpha = 0.05$. This indicates that the null hypothesis (H_0) is rejected, and the alternative hypothesis (H_1) is accepted. Consequently, the series are cointegrated: while each series is non-stationary on its own, their linear combination is stationary. Therefore, spurious regression can be ruled out in the mathematical model. Furthermore, the markedly negative value of tau_c(5) = -4.559 provides substantial evidence in support of rejecting the null hypothesis. For this reason, the series are cointegrated.

The test parameters demonstrate the suitability of the configuration of variables listed in Table 14. Equation (23) can be employed to formulate a mathematical model for predicting the nitrogen content in steel at the conclusion of secondary metallurgy. The resulting model assumes the form of Equation (38).

$$N_{SME} = 0.0257714 - 1.45992 \times 10^{-5} \cdot D_1 + 0.0126950 \cdot D_2 + 3.08438 \times 10^{-6} \cdot D_3 - 0.00153525 \cdot D_4 \tag{38}$$

where:

N_{SME}: predicted nitrogen content in steel at the end of secondary metallurgy;

D_1: steel temperature at the end of SM [°C];

D_2: final carbon in molten steel [%];

D_3: addition of FeMn aff. during SM [%];

D_4: final manganese in molten steel [%].

The validity ranges of the model (38) are illustrated in Table 24.

Table 24. Validity range of the N_{SME} model (38) for predicting the amount of nitrogen content in steel at the end of secondary metallurgy.

Parameter	Range of Values
Steel temperature at the end of SM [°C]	1558–1597
Final carbon in molten steel [%]	0.039–0.195
Addition of FeMn aff. during SM [%]	0–184
Final manganese in molten steel [%]	0.218–1.380

The proposed N_{SME} model (38) was subjected to the same rigorous diagnostic process as previous models. The process entailed methodical testing, thorough analysis, and graphical interpretation. The evaluation of the model is based on an analysis of residuals. As illustrated in Figure 27a, the residual variance is presented graphically, while Figure 27b provides a residual analysis over time.

(**a**)

(**b**)
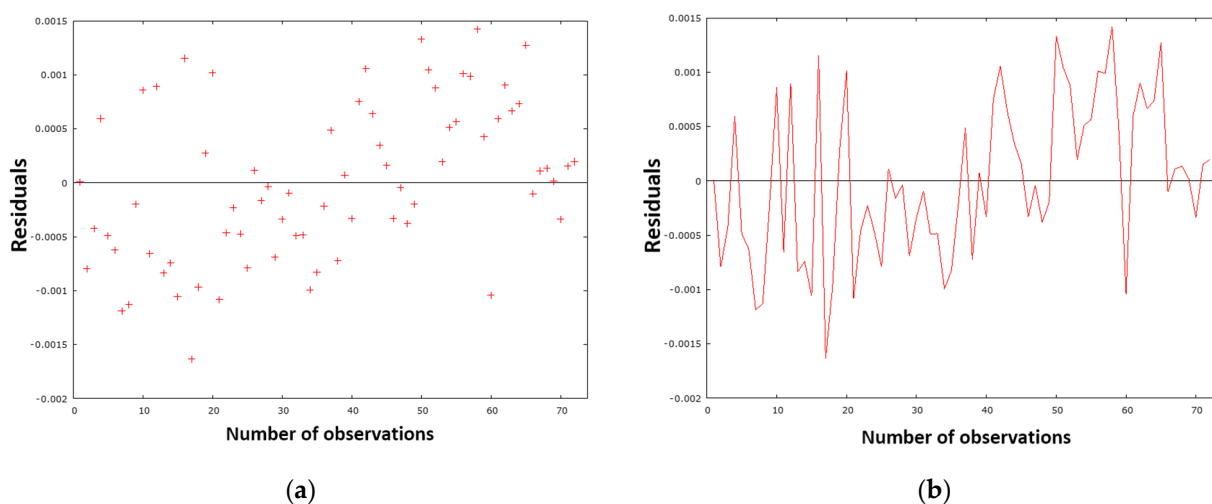
Figure 27. (**a**) Residual dispersion for the processed dataset and model N_{SME} (38); (**b**) the residual course in the timeline for the processed dataset and model N_{SME} (38).
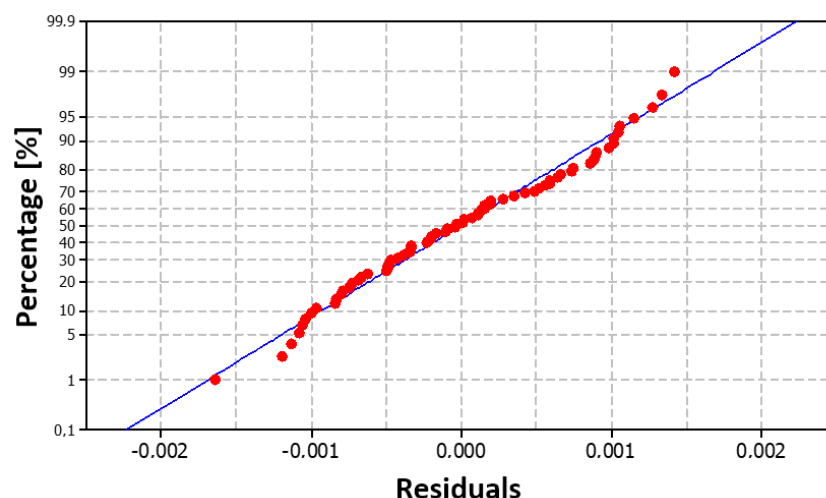
The distribution of residuals in Figure 27a suggests that the random homoscedasticity assumption is valid, which is favorable for the model's reliability. However, the time series in Figure 27b reveals potential issues of serial correlation and systematic sampling (around observations 15–20), though testing for autocorrelation of residuals does not confirm these. It can therefore be concluded that the model is adequately specified.

Both diagnostic plots (Figures 28 and 29) suggest that the residuals exhibit an approximately normal distribution, with slight deviations from perfect normality, particularly at extreme values and at the peak of the distribution. For practical purposes, however, we can consider the assumption of residual normality to be sufficiently satisfied.

Figure 30 shows a comparison of the measured and predicted values generated by the N_{SME} model (38). The graph shows the measured values (red curve), the predicted nitrogen concentrations in crude steel prior to tapping (blue curve), and the 95% confidence interval (green curves). The calculated standard deviation of residuals is 0.00070872, indicating

a minimal discrepancy between the observed and predicted datasets. This negligible deviation shows that the N_{SME} model (38) has satisfactory predictive accuracy.

**Figure 28.** Graph of the normal distribution of residuals for the model N_{SME} (38).
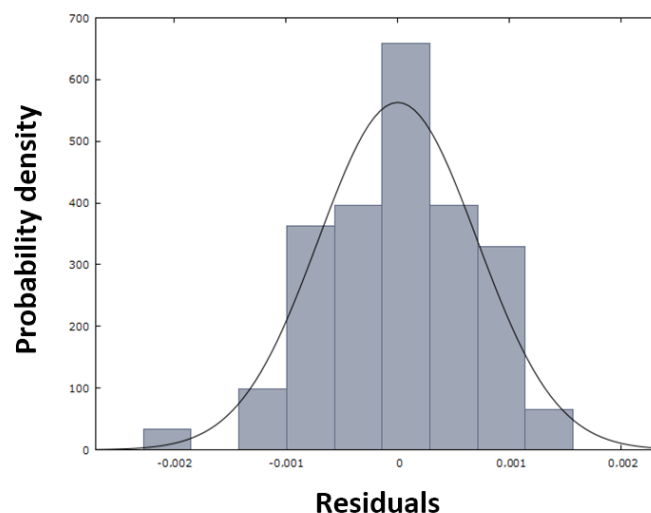
Figure 29. Histogram of the normal distribution of residuals for model N_{SME} (38).

The White and Breusch–Pagan tests were employed to evaluate the presence of heteroscedasticity and to verify the null hypothesis (H_0), which assumes the absence of heteroscedasticity, and the alternative hypothesis (H_1), which assumes its presence. The White test yielded a value of 15.7133. The null hypothesis is to be rejected if the value of the test statistic is greater than the corresponding critical value, $\chi^2(14)$, at the chosen confidence level, $\alpha = 0.05$. However, this would not be the case because $\chi^2(14) = 23.685 > 15.7133$. Due to this, in White's test, the null hypothesis of no heteroscedasticity is accepted. The Breusch–Pagan test statistic for heteroscedasticity is 3.90746. The null hypothesis (H_0) is rejected if the value of the Breusch–Pagan test statistic is greater than the corresponding critical $\chi^2(4)$ value at the chosen confidence level. However, this is not applicable in this instance, as $\chi^2(4) = 9.488 > 3.90746$. It is evident that, in accordance with the Breusch–Pagan statistic, the null hypothesis (H_0) is accepted, namely, that there is no heteroscedasticity.

The variance inflation factor (VIF) is evaluated as part of the solution to multicollinearity. The VIF values for the analyzed factors d_1–d_4 (Table 16) are shown in Table 25. These values indicate that the regression model is relatively favorable. The resulting VIF statistics indicate either no multicollinearity (values around 1) or slightly increased multicollinearity

(values around 6). However, all values are below the critical threshold of 10, indicating that serious multicollinearity problems do not threaten the model. The analysis of VIF values confirms the statistical robustness of the regression model. Higher values for carbon and manganese are technologically justified by the chemical dependence of these elements in steelmaking processes. Therefore, the model can be considered suitable for further analysis without the need to eliminate variables or make further structural adjustments. Because of this, each variable provides unique information to the N_{SME} model (38).
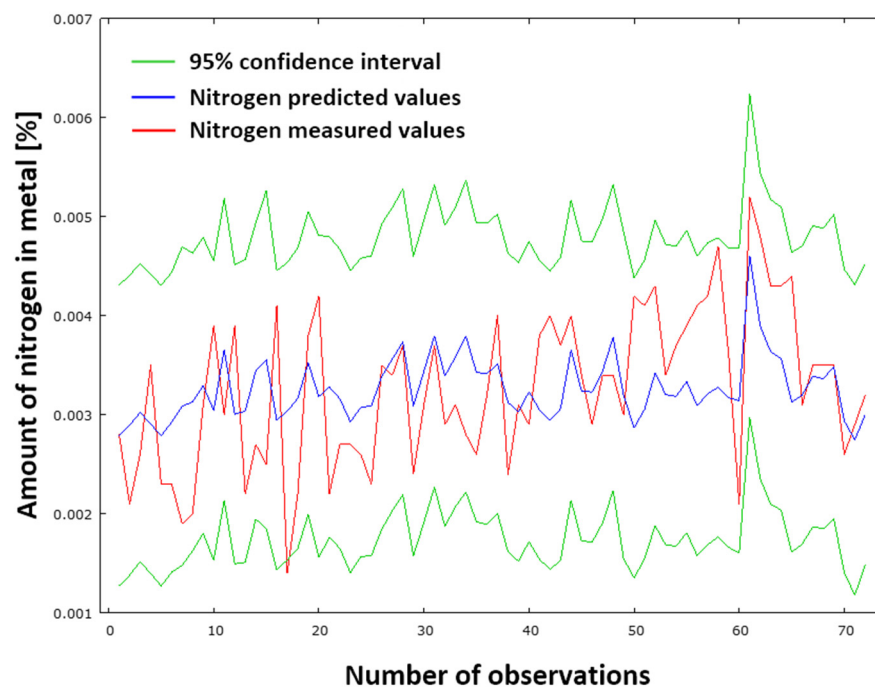
Figure 30. Comparison of measured and predicted nitrogen values according to model N_{SME} (38).

Table 25. Results of the VIF test for the factors that influence the nitrogen content in steel at the end of the secondary steelmaking.

Parameter	Value
Steel temperature at the end of SM [°C]	1.665
Final carbon in molten steel [%]	6.103
Addition of FeMn aff. during SM [%]	1.009
Final manganese in molten steel [%]	5.867

The accuracy of the N_{SMB} model can be evaluated as follows: The Mean Absolute Error (MAE) was calculated using Equation (24), giving a result of $MAE_{SME} = 0.00056649$. Therefore, it can be concluded that the average absolute error is very small.

The mean percentage error (MPE) was computed based on Equation (25). After substituting this into the relationship, the result is $MPE_{SME} = -5.6189\%$. Consequently, it can be deduced that the proposed model systematically overestimates reality, with the predicted values being, on average, 5.6189% higher than the actual values.

The Mean Absolute Percentage Error (MAPE) is used to express the average size of forecast errors compared to actual values across the entire forecast period. Substituting into Equation (26) yields $MAPE_{SME} = 19.6271\%$. The accuracy of the N_{SME} model can be determined by substituting its parameters into Equation (27). This establishes Equation (39), which determines the accuracy of the N_{SME} model, as outlined in Equation (40).

$$\text{Model accuracy}: \ N_{SME} = 100 - MAPE_{SME} \tag{39}$$

$$\text{Model accuracy: } N_{SME} = 80.3729\% \tag{40}$$

5. Conclusions

The presented research introduces a comprehensive system of predictive models for nitrogen content control in steel across individual production stages in the basic oxygen furnace process. Four specialized mathematical models based on the Ordinary Least Squares (OLS) method were successfully developed, covering key technological stages: pig iron desulfurization (N_{DeS}), crude steel before tapping from BOF (N_{BOF}), beginning of secondary metallurgy (N_{SMB}), and end of secondary metallurgy (N_{SME}).

The results demonstrate that the proposed models achieve satisfactory predictive accuracy with the following values: $N_{DeS} = 83.46\%$, $N_{BOF} = 77.23\%$, $N_{SMB} = 79.97\%$, and $N_{SME} = 80.37\%$. A hierarchy of factors influencing nitrogen content was identified for each stage, revealing that the most significant factors vary depending on the technological phase of production.

The acquired values were then subjected to a series of verification methods, including a heteroscedasticity test, a normality test, an autocorrelation test, a collinearity test, and a graphical examination of the distribution of residuals. In order to verify the accuracy of the proposed model using modern regression analysis, econometric cointegration analysis was also applied. All algorithms and methodologies that were subjected to rigorous testing during the steel production and processing phases that were monitored were found to provide conclusions that serve to confirm the proposed models for predicting nitrogen in metal. Consequently, it can be concluded that the results obtained from the proposed prediction formulas are reliable within the calculated accuracy of the model and boundary conditions.

The data being exclusively sourced from U. S. Steel Košice, Slovakia, means that the developed models reflect the unique technological conditions, raw material characteristics, and operational procedures specific to this facility. Through effective knowledge transfer and subsequent integration into process control systems, these models could facilitate initiative-taking parameter optimization while reducing dependence on laboratory testing times, thereby preventing excessive nitrogen content in the final steel products.

The proposed models are designed to serve as a practical tool for predicting nitrogen values in metal at specific production stages. This has the potential to enhance the efficiency of the production process and reduce the costs associated with chemical analysis of metal for the presence of nitrogen. The models created are valid within the specified ranges of values for individual parameters (Tables 18, 20, 22, and 24), which also correspond to the prescribed values of the chemical composition of the analyses steel grades, whose chemical composition is given in Section 2. In order to enhance the accuracy and robustness of predictive models, the following recommendations are proposed: the implementation of advanced machine learning algorithms, such as neural networks, random forest, or support vector machines, which have the capacity to better capture nonlinear relationships between variables; and the expansion of the dataset with additional operational parameters and longer time series, in order to increase statistical significance.

Author Contributions: Conceptualization, J.D.; methodology, J.D.; validation, J.D., B.B. and P.D.; formal analysis, J.D.; investigation, J.D.; resources, J.D. and B.B.; data curation, J.D.; writing—original draft preparation, J.D.; writing—review and editing, B.B., P.D. and M.H.; visualization, J.D.; supervision, B.B. and M.H.; project administration, J.D. and B.B.; funding acquisition, J.D. All authors have read and agreed to the published version of the manuscript.

Nomenclature

The following abbreviations are used in this manuscript:

K_N	Equilibrium constant of reaction [Pa$^{-1/2}$]
a_N	Activity of elemental nitrogen dissolved in metal [$-$]
f_N	Activity coefficient of elemental nitrogen dissolved in metal [$-$]
p_{N_2}	Partial pressure of gas in molecular form in a gaseous atmosphere above molten metal [Pa]
[%N]	Equilibrium concentration of elemental nitrogen dissolved in metal [wt.%]
C	Constant that depends on the given gas $C_{N^{3-}} = 6.01 \cdot 10^{-13}$[J.K^{-1}]
ΔH	Heat effect of dissolving 1 mole of gas [J.mol^{-1}]
k	Boltzmann constant k = $(1.380658 \pm 0.000012) \cdot 10^{-23}$ [J.K^{-1}]
T	Absolute temperature [K]
D	Diffusion coefficient [m^2.s^{-1}]
δ	Thickness of the metal diffusion layer (Nernst diffusion layer) [m]
S	Surface of the gas-melt phase interface [m^2]
$[N]_{srfc}$	Nitrogen concentration in the surface layer of the melt [mol.m^{-3}]
$[N]_{vol}$	Nitrogen concentration in the melt volume [mol.m^{-3}]
t	Time [s]
A	Size of the gas bubble surface [m^2]
V	Gas bubble volume [m^3]
$C_{[N]_{eq.}}$	Equilibrium nitrogen concentration [wt.%]
$C_{[N]}$	Nitrogen concentration [wt.%].
k	Mass transfer coefficient [m^2.s^{-1}]
β	Nitrogen dissolution rate constant
n	Number of observations
X, Y	Quantities that can be written as x_i and y_i, where i = 1, 2, 3, ..., n,
$\bar{x}, \bar{y}$	Average values of variables x and y
s_x, s_y	Standard deviation of variables x and y
$s^2_{x|y}, s^2_{y|x}$	Number of parameters to be measured
s^2_x, s^2_y	Parameters that can be written as xi and yi and in this case i = 1, 2, 3, ..., n
N_n	Theoretical, predicted, nth balanced value of the explained variable N (nitrogen)
z_0	Constant/Intercept
z_1–z_n	Coefficient of variable
Z_1–Z_n	Value of an independent variable
e_t	Residuals (difference between measured and calculated values)
y_t	Measured monitored (dependent) variable
N_{stage}	Identification of a model for predicting the amount of nitrogen in molten metal during individual stages of steel production

Appendix A

The following charts (Figure A1a–d) are intended to provide a clear visual representation of the models' deployment in the specific phases of the BOF production route.

(a)

(b)

(c)

(d)
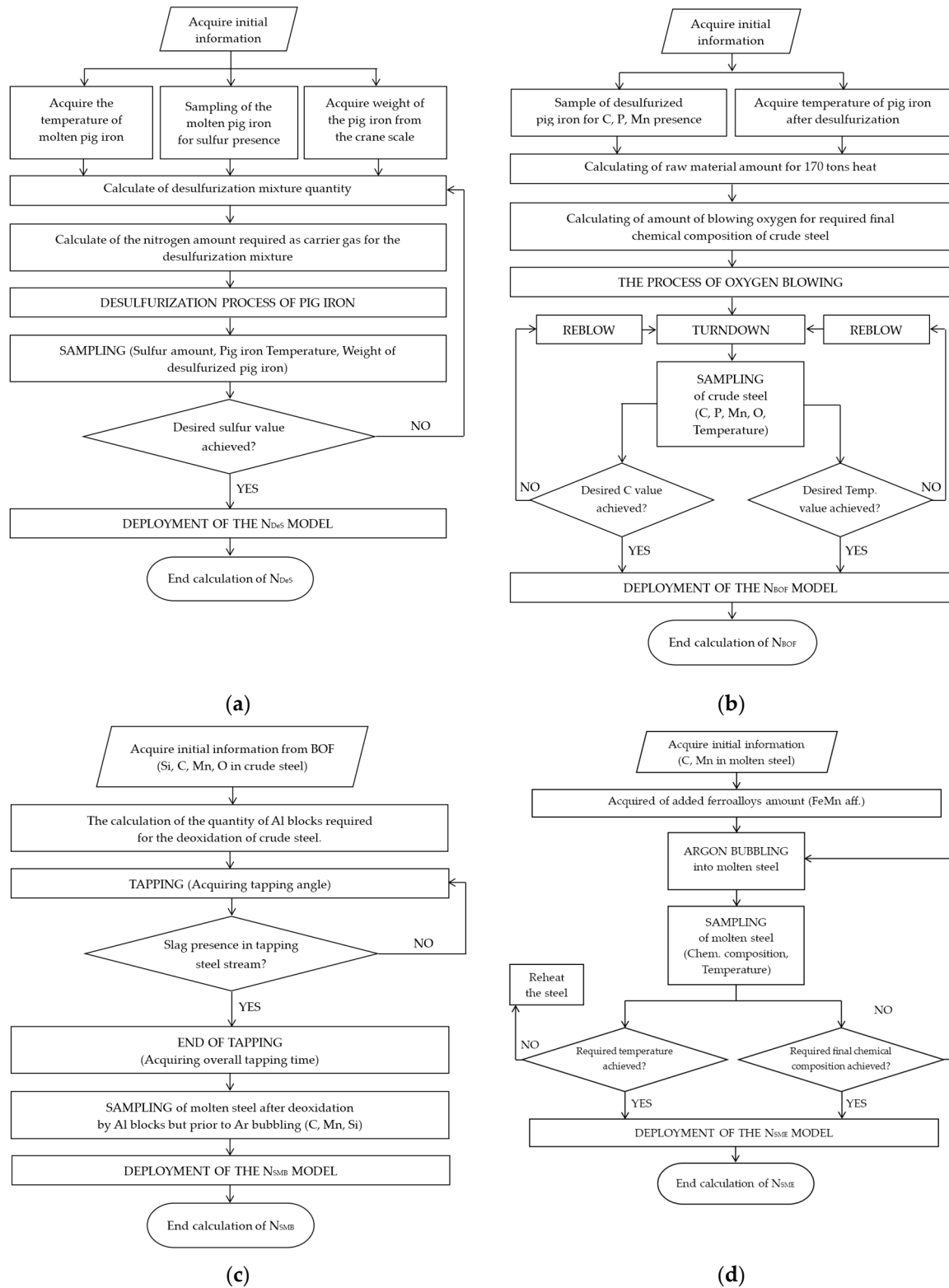
Figure A1. Flow charts of model deployment in specific phases of BOF production cycle. (**a**) Deployment of N_{DeS} model after desulfurization of molten pig iron; (**b**) deployment of N_{BOF} model prior to tapping from BOF; (**c**) deployment of N_{SMB} model at the beginning of secondary metallurgy; (**d**) deployment of N_{SME} model at the end of secondary metallurgy.

References

1. Gavriljuk, V.G. Nitrogen in Iron and Steel. *ISIJ Int.* **1996**, *36*, 738–745. [CrossRef]
2. Levey, P.R.; van Bennekom, A. A Mechanistic Study of the Effects of Nitrogen on the Corrosion Properties of Stainless Steels. *Corrosion* **1995**, *51*, 911–921. [CrossRef]
3. Tsuchiyama, T.; Fukumaru, T.; Egashira, M.; Takaki, S. Calculation of Nitrogen Absorption into Austenitic Stainless Steel Plate and Wire. *ISIJ Int.* **2004**, *44*, 1121–1123. [CrossRef]
4. Baba, H.; Kodama, T.; Katada, Y. Role of Nitrogen on the Corrosion Behavior of Austenitic Stainless Steels. *Corros. Sci.* **2002**, *44*, 2393–2407. [CrossRef]
5. Hertzman, S.; Naraghi, R.; Wessman, S.; Pettersson, R.; Borggren, U.; Jonsson, J.Y.; Pettersson, N.H.; Karami, M.K.; Kohan-Zade, A. Nitrogen Solubility in Alloy Systems Relevant to Stainless Steels. *Met. Mater. Trans. A* **2021**, *52*, 3811–3820. [CrossRef]
6. Chai, G.; Siriki, R.; Nordström, J.; Dong, Z.; Vitos, L. Roles of Nitrogen on TWIP in Advanced Austenitic Stainless Steels. *Steel Res. Int.* **2023**, *94*, 2200359. [CrossRef]
7. Dong, H.; Chai, G.; Guo, X. High Nitrogen Steels. *Steel Res. Int.* **2023**, *94*, 2300505. [CrossRef]
8. Park, W.-I.; Jung, S.-M.; Sasaki, Y. Fabrication of Ultra High Nitrogen Austenitic Stainless Steel by NH3 Solution Nitriding. *ISIJ Int.* **2010**, *50*, 1546–1551. [CrossRef]
9. Zhao, F.; Liu, X.; Zhang, Z.; Xie, J. Effect of Nitrogen Content on the Mechanical Properties and Deformation Behaviors of Ferritic-Pearlitic Steels. *Mater. Sci. Eng. A* **2022**, *855*, 143918. [CrossRef]
10. Hänninen, H.; Romu, J.; Ilola, R.; Tervo, J.; Laitinen, A. Effects of Processing and Manufacturing of High Nitrogen-Containing Stainless Steels on Their Mechanical, Corrosion and Wear Properties. *J. Mater. Process. Technol.* **2001**, *117*, 424–430. [CrossRef]
11. Bazaleeva, K.O. Mechanisms of the Influence of Nitrogen on the Structure and Properties of Steels (A Review). *Met. Sci. Heat. Treat.* **2005**, *47*, 455–461. [CrossRef]
12. Liu, Z.; Fan, C.; Yang, C.; Ming, Z.; Lin, S.; Wang, L. Dissimilar Welding of High Nitrogen Stainless Steel and Low Alloy High Strength Steel under Different Shielding Gas Composition: Process, Microstructure and Mechanical Properties. *Def. Technol.* **2023**, *27*, 138–153. [CrossRef]
13. Woo, I.; Kikuchi, Y. Weldability of High Nitrogen Stainless Steel. *ISIJ Int.* **2002**, *42*, 1334–1343. [CrossRef]
14. Saxena, A.; Sengupta, A.; Chaudhuri, S.K. Effect of Absorbed Nitrogen on the Microstructure and Core Loss Property of Non-Oriented Electrical Steel. *ISIJ Int.* **2005**, *45*, 299–301. [CrossRef]
15. Misra, S.; Fruehan, R.J. *Hydrogen and Nitrogen Control in Ladle and Casting Operations*; Carnegie Mellon University Pittsburgh: Pittsburgh, PA, USA, 2005; p. 62.
16. Turkdogan, E.T. *Fundamentals of Steelmaking*; Institute of Materials: London, UK, 2010; ISBN 978-1-907625-73-2.
17. Pitkälä, J.; Xia, J.; Jokilaakso, A. *CFD Modeling of Nitrogen Dissolution into a Steel Bath During Gas Purging*; CSIRO: Melbourne, Australia, 1999; pp. 35–40.
18. Inomoto, T.; Kitamura, S.; Yano, M. Kinetic Study of the Nitrogen Removal Rate from Molten Steel (Normal Steel and 17 mass%Cr Steel) under CO Boiling or Argon Gas Injection. *ISIJ Int.* **2015**, *55*, 1822–1827. [CrossRef]
19. Slater, C.; Spooner, S.; Davis, C.; Sridhar, S. Observation of the Reversible Stabilisation of Liquid Phase Iron during Nitriding. *Mater. Lett.* **2016**, *173*, 98–101. [CrossRef]
20. Trotter, D.; Varcoe, D.; Reeves, R.; Hornby, S. Use of HBI and DRI for Nitrogen Control in Steel Products. In Proceedings of the Electric Furnace Conference, San Antonio, TX, USA, 10–13 November 2002; Volume 31, pp. 39–50.
21. Seetharaman, S. *Fundamentals of Metallurgy*; CRC Press: Boca Raton, FL, USA, 2005; ISBN 978-1-85573-927-7.
22. Ghosh, A.; Chatterjee, A. *Ironmaking and Steelmaking: Theory and Practice*; Eastern economy edition 3. print; PHI Learning: New Delhi, India, 2010; ISBN 978-81-203-3289-8.
23. Gupta, C.K. *Chemical Metallurgy: Principles and Practice*; John Wiley & Sons: Hoboken, NJ, USA, 2006; ISBN 978-3-527-60525-5.
24. Rosenqvist, T. *Principles of Extractive Metallurgy*; Tapir Academic Press: Trondheim, Norway, 2004; ISBN 978-82-519-1922-7.
25. Gavriljuk, V.G.; Berns, H. *High Nitrogen Steels: Structure, Properties, Manufacture, Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1999; ISBN 978-3-540-66411-6.
26. Derin, B.; Alan, E.; Suzuki, M.; Tanaka, T. Phosphate, Phosphide, Nitride and Carbide Capacity Predictions of Molten Melts by Using an Artificial Neural Network Approach. *ISIJ Int.* **2016**, *56*, 183–188. [CrossRef]
27. Liapina, T. Phase Transformations in Interstitial Fe-N Alloys. 2005. Available online: https://www.osti.gov/etdeweb/biblio/20841020 (accessed on 27 August 2025).
28. Dutta, S.K.; Chokshi, Y.B. Secondary Steelmaking. In *Basic Concepts of Iron and Steel Making*; Dutta, S.K., Chokshi, Y.B., Eds.; Springer: Singapore, 2020; pp. 497–536, ISBN 978-981-15-2437-0.
29. Shamsuddin, M. Secondary Steelmaking. In *Physical Chemistry of Metallurgical Processes, Second Edition*; Shamsuddin, M., Ed.; Springer International Publishing: Cham, Switzerland, 2021; pp. 293–351, ISBN 978-3-030-58069-8.
30. Fruehan, R.J. *The Making, Shaping, and Treating of Steel*, 11th ed.; AISE Steel Foundation, Ed.; AISE Steel Foundation: Pittsburgh, PA, USA, 1998; ISBN 978-0-930767-03-7.

31. Bodsworth, C. *The Extraction and Refining of Metals*; CRC Press: Boca Raton, FL, USA, 1994; ISBN 978-0-8493-4433-6.

32. Engh, T.A.; Sigworth, G.K.; Kvithyld, A. *Principles of Metal Refining and Recycling*; Oxford University Press: London, UK, 2021; ISBN 978-0-19-881192-3.

33. Baptizmansky, V.I.; Medzhibozhsky, M.Y.; Okhotsky, V.B. *Converter Processes in Steel Production. Theory, Technology, Equipment Design*; Vyshcha Shkola-Main Publishing House: Kyiv, Ukraine, 1984.

34. Pitkälä, J.; Holappa, L.; Jokilaakso, A. A Study of the Effect of Alloying Elements and Temperature on Nitrogen Solubility in Industrial Stainless Steelmaking. *Met. Mater. Trans. B* **2022**, *53*, 2364–2376. [CrossRef]

35. Zhang, Y.; Liu, J.; Yang, H.; Li, W.; Zhang, S.; Zhou, W. Kinetics of Nitrogen Absorption/Vacuum Denitrigenization and Precipitation Behavior of Nitrogen Bubbles in 42CrMoA Molten Steel. *Steel Res. Int.* **2025**, *96*, 2400071. [CrossRef]

36. Luo, J.; He, Z.; Hua, Z.; Fan, C. Research on Microstructure and Mechanical Properties of Ultrasonic-Assisted Gas Metal Arc Welding Additive Manufacturing with High-Nitrogen Steel Welding Wire. *Metals* **2025**, *15*, 491. [CrossRef]

37. Kim, M.; Geller, C.B.; Freeman, A.J. The Effect of Interstitial N on Grain Boundary Cohesive Strength in Fe. *Scr. Mater.* **2004**, *50*, 1341–1343. [CrossRef]

38. Duraipandi, R.; Nani Babu, M.; Moitra, A. Fatigue Crack Growth Behavior of Nitrogen-Alloyed Low-Carbon Austenitic Stainless Steel at Room Temperature. *JOM* **2023**, *75*, 478–487. [CrossRef]

39. Gu, J.; Li, J.; Chen, Y. Microstructure and Strengthening-Toughening Mechanism of Nitrogen-Alloyed 4Cr5Mo2V Hot-Working Die Steel. *Metals* **2017**, *7*, 310. [CrossRef]

40. Oloro, J.O. Formulation of Linear Regression Model For Steel Production Prediction For Oil and Gas Operations in Nigeria. *J. Mater. Environ. Sci.* **2020**, *11*, 1019–1032.

41. Goldstein, D.A.; Fruehan, R.J. Mathematical Model for Nitrogen Control in Oxygen Steelmaking. *Met. Mater. Trans. B* **1999**, *30*, 945–956. [CrossRef]

42. Bae, J.; Li, Y.; Ståhl, N.; Mathiason, G.; Kojola, N. Using Machine Learning for Robust Target Prediction in a Basic Oxygen Furnace System. *Met. Mater. Trans. B* **2020**, *51*, 1632–1645. [CrossRef]

43. van Dam, M. Predicting Nitrogen Concentrations Using Machine Learning Techniques. Master's Thesis, Tilburg University, Tilburg, The Netherlands, 2021.

44. Xu, L.; Li, W.; Zhang, M.; Xu, S.; Li, J. A Model of Basic Oxygen Furnace (BOF) End-Point Prediction Based on Spectrum Information of the Furnace Flame with Support Vector Machine (SVM). *Optik* **2011**, *122*, 594–598. [CrossRef]

45. Usman, M. Multivariate Time Series Prediction for Endpoint Prediction of Temperature, Phosphorus, and Carbon in the Basic Oxygen Furnace. Master's Thesis, University of Skövde, Skövde, Sweden, 2024.

46. Sheik, S.; Mohammed, R.; Teeparthi, K.; Raghuvamsi, Y. Machine Learning-Based Prediction of Intergranular Corrosion Resistance in Austenitic Stainless Steels Exposed to Various Heat Treatments. *J. Inst. Eng. (India) Ser. D* **2024**, *106*, 491–504. [CrossRef]

47. *ASTM E1019-18*; Standard Test Methods for Determination of Carbon, Sulfur, Nitrogen, and Oxygen in Steel, Iron, Nickel, and Cobalt Alloys by Various Combustion and Inert Gas Fusion Techniques. ASTM International: West Conshohocken, PA, USA, 2018. [CrossRef]

48. Gozhyj, A.P.; Kalinina, I.A.; Bidyuk, P.I. Systematic Use of Nonlinear Data Filtering Methods in Forecasting Tasks. *AAIT* **2023**, *6*, 345–361. [CrossRef]

49. Bhandari, P. Correlation Coefficient | Types, Formulas & Examples. Available online: https://www.scribbr.com/statistics/correlation-coefficient/ (accessed on 24 March 2025).

50. Turney, S. Pearson Correlation Coefficient. Available online: https://www.scribbr.com/statistics/pearson-correlation-coefficient/ (accessed on 24 March 2025).

51. Onyango, J.P.; Plews, A.M. *A Textbook of Basic Statistics*; East African Publishers: Nairobi, Kenya, 1987; ISBN 9966-46-251-8.

52. Shevlyakov, G.L.; Oja, H. *Robust Correlation: Theory and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2016; ISBN 978-1-119-26453-8.

53. Roux, B.L.; Rouanet, H. *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006; ISBN 978-1-4020-2236-4.

54. Zighed, D.A.; Tsumoto, S.; Ras, Z.W.; Hacid, H. *Mining Complex Data*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008; ISBN 978-3-540-88066-0.

55. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Routledge: New York, NY, USA, 2013; ISBN 978-0-203-77158-7.

56. Yachikov, I.; Naizabekov, A.; Lezhnev, S.; Myasnikova, A.; Trofimov, E.; Panin, E.; Samodurova, M. Mathematical Modeling of the Non-Stationary Thermal State of a Composite Coating of Close to Equimolar Composition During Laser Remelting. *J. Chem. Technol. Metall.* **2024**, *59*, 1215–1226. [CrossRef]

57. Ozili, P.K. The Acceptable R-Square in Empirical Modelling for Social Science Research. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4128165 (accessed on 10 July 2025).

58. Smant, D.J.C. Reading Computer Output of a Simple Linear Regression. Available online: https://web.archive.org/web/20060522175346/http://people.few.eur.nl/smant/econometrics/intro_pr_ectr_1.pdf (accessed on 10 July 2025).

59. Petkov, V.; Hadjiski, M.; Boshnakov, K. Diagnosis of Metallurgical Ladle Refractory Lining Based on Non-Stationary On-Line Data Processing. *Cybern. Inf. Technol.* **2013**, *13*, 122–130. [CrossRef]

60. Good, P. *Ordinary Least Squares*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2012; pp. 141–162.

61. Mosteller, F.; Tukey, J.W. Data Analysis and Regression. *J. R. Stat. Soc. Ser. A (General)* **1978**, *141*, 549. [CrossRef]

62. Magnus, J.R. On Using the T-Ratio as a Diagnostic. *Econometrics* **2019**, *7*, 24. [CrossRef]

63. Trochim, W.M.K. The T-Test. Available online: https://conjointly.com/kb/statistical-student-t-test/ (accessed on 14 July 2025).

64. Trafimow, D. A Frequentist Alternative to Significance Testing, p-Values, and Confidence Intervals. *Econometrics* **2019**, *7*, 26. [CrossRef]

65. Nahm, F. What the P Values Really Tell Us. *Korean J. Pain.* **2017**, *30*, 241–242. [CrossRef]

66. Corotto, F.S. *Wise Use of Null Hypothesis Tests: A Practitioner's Handbook*; Elsevier: Amsterdam, The Netherlands, 2022; ISBN 978-0-323-95285-9.

67. Myttenaere, A.D.; Golden, B.; Grand, B.L.; Rossi, F. Mean Absolute Percentage Error for Regression Models. *Neurocomputing* **2016**, *192*, 38–48. [CrossRef]

68. Snedecor, G.W.; Cochran, W.G. *Statistical Methods*; Iowa State University Press: Ames, IA, USA, 1989; ISBN 978-0-8138-1561-9.

69. NIST/SEMATECH. e-Handbook of Statistical Methods. F-Test for Equality of Two Variances. Available online: https://www.itl.nist.gov/div898/handbook/eda/section3/eda359.htm (accessed on 17 July 2025).

70. Osborne, J.W. Regression & Linear Modeling: Best Practices and Modern Methods. Available online: https://sk.annas-archive.org/md5/1feb3a379f59b4305b9ded659c399953 (accessed on 14 July 2025).

71. Wang, C.S.-H.; Hafner, C.M. A Simple Solution of the Spurious Regression Problem. *Stud. Nonlinear Dyn. Econom.* **2018**, *22*, 1–14. [CrossRef]

72. Swamy, P.a.V.B.; von zur Muehlen, P.; Mehta, J.S.; Chang, I.-L. Spurious Regressions in Econometrics: Reconsideration. 2019. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3320044 (accessed on 17 July 2025). [CrossRef]

73. Aslam, M.; Paasha, G.R. and G.R. Adaptive Estimation of Heteroscedastic Linear Regression Models Using Heteroscedasticity Consistent Covariance Matrix. *J. Stat.* **2024**, *16*, 28–44.

74. Berenguer-Rico, V.; Wilms, I. Heteroscedasticity Testing after Outlier Removal. *Econom. Rev.* **2021**, *40*, 51–85. [CrossRef]

75. Shrestha, N. Detecting Multicollinearity in Regression Analysis. *AJAMS* **2020**, *8*, 39–42. [CrossRef]

76. Kim, J.H. Multicollinearity and Misleading Statistical Results. *Korean J. Anesth.* **2019**, *72*, 558–569. [CrossRef] [PubMed]

77. ChemBlink Molasses, Beet Molasses. Available online: https://www.chemblink.com/products/68476-78-8.htm (accessed on 28 July 2025).

78. Chatterjee, S.; K Rout, B. Nitrogen Control in the Basic Oxygen Steelmaking Process. In Proceedings of the AISTech 2023 Proceedings, Detroit, MI, USA, 8–11 May 2023; pp. 752–765.

79. Lule, R.; Lopez, F.; Espinoza, J.; Torres, R.; Morales, R.D. The Production of Steels Applying 100% DRI for Nitrogen Removal. Available online: https://www.midrex.com/wp-content/uploads/Production_of_Steels_Applying.pdf (accessed on 15 August 2025).

80. Ardelean, E.; Ardelean, M.; Heput, T.; Drăgoi, F. Research on Increasing the Nitrogen Removal Efficiency by Changing the Secondary Treatment Parameters. *Solid. State Phenom.* **2014**, *216*, 267–272. [CrossRef]

81. Zhang, F.; Li, J.; Liu, W.; Jiao, A. The Thermodynamics and Kinetics of a Nitrogen Reaction in an Electric Arc Furnace Smelting Process. *Materials* **2022**, *16*, 33. [CrossRef] [PubMed]