**DOI: 10.37943/24QUFK4295**

**Andrii Biloshchytskyi**
Doctor of Technical Sciences, Professor, Vice-Rector for Science and Innovations
a.b@astanait.edu.kz, orcid.org/0000-0001-9548-1959
Astana IT University, Kazakhstan
Professor Department of Information Technologies,
Kyiv National University of Construction and Architecture, Ukraine

**Alexandr Neftissov**
PhD, Associate Professor, Rectorate for Science and Innovation
alexandr.neftissov@astanait.edu.kz, orcid.org/0000-0003-4079-2025
Academy of Physical Education and Mass Sports, Kazakhstan
PhD, Associate Professor, Researcher, Scientific-Innovation Center Industry 4.0
Astana IT University, Kazakhstan

**Ilyas Kazambayev**
Master`s degree, Acting Director of Scientific-Innovation Center Industry 4.0
i.kazambayev@astanait.edu.kz, orcid.org/0000-0003-0850-7490
Astana IT University, Kazakhstan

**Magzhan Amangeldi**
Master's degree, Doctorate student
255242@astanait.edu.kz, orcid.org/0009-0002-7979-0502
Astana IT University, Kazakhstan

**Lalita Kirichenko**
Master`s degree, Junior Researcher of Scientific-Innovation Center Industry 4.0
l.kirichenko@astanait.edu.kz, orcid.org/0000-0001-7069-5395
Astana IT University, Kazakhstan

**Ainur Medetkhan**
Master`s degree, Junior Researcher of Scientific-Innovation Center Industry 4.0
ainur.medetkhan@astanait.edu.kz, orcid.org/0009-0003-3539-6930
Astana IT University, Kazakhstan

# CORRELATION-MATRIX–DRIVEN DIAGNOSTICS OF INDUSTRIAL EMISSIONS: A PEARSON BASELINE WITH SCATTER-PLOT EVIDENCE

**Abstract:** Currently environmental state became very actual in the world, especially in Kazakhstan. Air pollution of industries is a major threat to the environment and health of the people, especially in areas with high reliance on coal-powered power stations in electricity production. Fossil fuels in Kazakhstan are the largest electrical source, and they contribute to the emission of sulfur dioxide ($SO_2$), nitrogen oxides ($NO_x$), carbon monoxide ($CO$), and the particle matter (PM). Although, to formulate diagnostic and monitoring procedures at industry level it is crucial to determine relationships among emissions. The study approaches the Pearson correlation method on data taken from an automated emission monitoring system at the Coal Power Plant in Kazakhstan. The aim of the study is to discover linearity between emission indicators and industrial combustion. The observed correlation heat map and scatter-plots indicate positive trends among the $CO$ and $SO_2$, inverse correlation between $CO$ and $O_2$, and insufficient relation of CO and NO. These results show the key combustion processes, which in-

Copyright © 2025, Authors. This is an open access article under the Creative Commons CC BY-NC-ND license
Received: 21.09.2025 Accepted: 20.10.2025 Published: 30.10.2025

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 24, OCTOBER 2025

volve reduced oxygen supply leading to the incomplete oxidation and simultaneous increased sulfur emissions. The three-dimensional description of $CO$ dependence on $SO_2$ and $O_2$ further explains the coupled emission response and supports the explanation of underlying regularities in the operation. The correlation-based framework has diagnostic capabilities of the early identification of inefficient combustion regimes and enables scalable and data-driven methods of emission control. The research finds that Pearson-based analytics can be used to offer a strong and interpretable predictive modeling and regulatory monitoring foundation of future air-quality management in industries.

**Keywords:** emission diagnostics; Pearson correlation; automated monitoring system; air pollution; combustion efficiency; coal power plant.

## Introduction

*Industrial Air Pollution Context*

Ambient air pollution is considered as one of the global significant health and regulatory problems. Recent WHO Air Quality Guidelines limit the recommended amounts of PM, NO, $SO_2$, $O_3$ and CO, which increases the importance of transparent, reproducible analysis of the emissions at industrial level [1]. Nevertheless, factors such as wintertime stagnation, fuel mix, and topography contribute to seasonal pollution patterns, which make reliable, diagnostics from manufactory stack data highly valuable [2]. Kazakhstan regulatory framework is currently enforced to introduce automated measuring system (ASM) of emissions at large stationary sources, which generates a uniform data stream to support facility level analysis and control [3].

Electricity production in Kazakhstan is crucially dominated by fossil fuels-based power plants, which utilize nonrenewable energy sources such as coal, natural gas, and fuel oil [4]. Coal Power Plants (CPPs) generate most electricity for Industrial and household demands. However, wide use of CPPs often cause negative environmental impact [5]. This includes emissions of greenhouse gases (GHGs) such as $CO_2$ $O_2$, and non GHG pollutants ($SO_2$, $NO_x$, PM) which are considered as by-products of chemical reactions of energy conversion [6]. Particle pollution (PM) also referred to as matter, consists of a mixture of airborne solid and liquid particles [7]. Moreover, while large particles such as dust, dirt and smoke are observable enough, some finer particles can only be observed using specific equipment like electron microscope [8].

*Statistical Approaches on Data Interpretation*

The authors in [9], provided various statistical techniques have been used to examine the relationship between different air pollution measures and Pearson correlation, is often considered a basis of diagnostic diagnostics to show patterns of emission sources in industrial contexts. It offers a quantitative evaluation of linear relationships between variables (PM, $SO_2$, $NO_2$, $O_2$ and CO) that would allow researchers to identify the probable co-emission patterns when focusing on particular industrial actions. Recent studies found utilizing Pearson correlation approach that there are great interdependencies among several airborne pollutants produced by the processes of coal burning and defined metrics of air quality control [10]. Through the transparency and reproducibility of correlation-based analyses, their applicability to emission diagnostics (in combination with automated industrial monitoring systems that provide high-resolution longitudinal data capture) has been firmly established.

Recent developments in environmental informatics emphasize the usefulness of the scatter plot visualization in the interpretation of the correlation matrices of the emission diagnostics [11]. These graphical methods enable one to determine nonlinear correlations and outlier behaviors to supplement Pearson baseline results, thereby increasing the overall granularity of diagnosis. A well-known example of scatter analysis of PM and NOx revealed seasonal and weather-based effects in CPPs in Central Asia [12]. Together with scatter plots, correlation analysis can lead to effective emission control operation in industry fields by isolation of regular pollution events [13].

*Integration of AMS and Correlation-Based Modeling*

Recent research presented the AMS framework with an advanced statistical model and achieved significant improvements in reproducibility and scalability of air quality in the industrial field [14]. These types of systems can provide strong sampling independent monitoring. Empirical findings demonstrate that correlation analysis with AMS can provide early abnormal emission detection to prevent further critical consequences [15]. However, utilization of open-source standardized protocols can enhance air pollution management practices in the world.

Recent research has suggested that correlation-based methods, especially the Pearson method, is still among the most available and the simplest methods to analyze complex industrial data. Correlation matrices can be used to identify regular emission patterns and irregularities when used on automated monitoring outputs, without the need to use intensive parametric modelling [16]. Their visibility and understandability are deemed vital to the power-generation facilities that are subjected to strict environmental standards, particularly where the decision-makers need practical indicators as opposed to the abstract model outputs [17]. Furthermore, more sophisticated correlation models, including either partial or rank based scales, are being added to consider non-linear relationship and temporal changes whilst preserving the intuitive interpretability of the underlying Pearson design [18].

The authors in [19] have incorporated correlation studies with the current monitoring data system. As an example, the dynamic principal component analysis (DPCA) and the moving-window correlation models have been used to trace slow operational changes and predict the abnormalities of constant monitoring streams. It is found that comparative analyses of these methods indicate that, although predictive performance may be superior in machine-learning models (such as random forests or deep regression networks) when compared with simple linear estimators, more traditional methods of correlation have more diagnostic explainability and generally reduced computational costs [20]. Hence, integrating classical correlation analysis with periodic visualization (rolling heatmaps and scatter plots) has emerged as an expedient norm in industrial environmental analytics [21]. Such a combination of methods ensures that emission data of high-frequency monitoring systems are transformed into efficient environmental management knowledge and offered not only in quantitative terms but also in visual forms.

*Aim of the Study*

This research proposes a Pearson correlation baseline represented by a heatmap and supported by specific scatter plots to describe linear co-variation between $\mathrm{CO}$, $\mathrm{SO_2}$, $\mathrm{NO}$, and $\mathrm{O_2}$ in a coal power plant. To determine strength of correlation between two variables and direction the Pearson correlation coefficient test is applied [Some Reference Here]. The raw data is taken from a coal power plant located in Kazakhstan, with more than 15000 records with measured parameters of gases concentration, emissions etc.

***Overview of analytical approach.***

Structured statistical evaluation of relationships between process variables and indicators of pollutants is key component in quantitative interpretation of emissions data. Among the current set of analytical models, correlation-based analysis provides methodologically informative and simple diagnostic. This approach enables distinguishing the prevalent emission patterns, possible control malfunctions, and anomalies of sensors without large modeling assumptions [22]. This study utilized correlation-based workflow to analysis a continuous dataset collected in one of the coal-fired plants in Kazakhstan to identify empirical data con-

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 24, OCTOBER 2025

cerning the relationship among the most significant emission parameters. This was aimed at establishing a baseline model of inter-variable interactions that will be suitable in future predictive and regulatory modeling.

The data set includes over 15,000 high-frequency data that were collected using an automated emission monitoring unit, which was embedded in the control system of the plant. Multiple chemical and operational parameters related to the quality of combustion and exhausts were data channels. To eliminate missing data, filtration is utilized on unnatural values due to maintenance or calibration of sensors and nonphysical outliers before computation. Similarly, numerical homogenization can provide a high level of accuracy and variable formatting. To indicate correlation analysis accuracy, such measures also can provide observed patterns which represent operational behavior trends and the data stability [23].

***Comparative context.***

The correlation diagnostics include regression and machine learning models more intently on interpretability than the predictive insights. However, multivariate methods such as principal component analysis (PCA) and dynamic PCA can be utilized to detect detailed patterns in collected emissions data. These approaches need significant amount of computation power and have limitations in scalability and integration [24], [25]. Nevertheless, correlation heatmap matrices can be used as initial analysis to estimate performance of the plants without complex experiments. These types of statistical methods are often used as backbone for further prediction approaches such as random fores regressors and deep learning models [26]. Thus, the current study provides initial step of multi-level diagnostic pipeline of industrial monitoring.

The parameters related to emissions showed moderate-strong linear correlation in the analyzed data, which indicates underlying thermochemical reactions and the mechanism of system control. Positive relationships tended to show combustion efficiency and the emissions of pollutants, whereas negative correlation implied a compensatory or regulatory effect in the control logic. These findings agree with the same ones reported in the recent literature on large scale industrial monitoring systems [27]. The resulting correlation matrix and the corresponding scatter plots, therefore, become a diagnostic reference and at the same time a benchmarking tool in the continuous verification of the system.

**Methods and Materials**

*Data source and description*

The data used in the current study was taken in a big power plant that uses coal, which is in the country of Kazakhstan. Continuous Emission Monitoring Systems (CEMS) provided as part of the automated control structure of the plant provided high-frequency data of the operating and environmental values. The total sample size was more than 15,000-time stamped measurements of gas concentrations, temperature, flow and indicators of the particulate. These values were the normal operating range of the plant, and these were first measured in normal load conditions during winter and summer seasons.

*Data preprocessing and quality control*

To ensure completeness and consistency before it was analyzed, the data was first screened. Records relating to maintenance operations, calibration activities, or equipment downtimes were cut out. Interquartile range thresholds were used to eliminate the nonphysical or extreme outliers. Standardization of numerical variables was performed to a consistent level of a non-zero decimal number and in the case of the presence of fields in the category form, the encoding was performed in accordance. It finds that the dataset contains missing values in some recordings, to obtain that these variables were handled using pairwise deletion. The preprocessing steps assisted to enhance statistical robustness and to remove artefacts which could negatively affect on correlation measures computation.

*Analytical procedure*

Pearson correlation coefficent is a standard metric of linear relation to evalueate linear dependence between measured values

For any X and Y with sample size n, the Pearson correlation coefficient $r_{XY}$ is defined as:

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{1}$$

where:
- $\bar{X}$ & $\bar{Y}$: denote sample means,
- $r_{XY} \in [-1,1]$,
- $r_{XY} > 0$ indicates direct proportional relationship, and $r_{XY} > 0$ visa versa.

The t-test is used to access statistical significance:

$$t = \frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}, \qquad t \sim t_{(n-2)} \tag{2}$$

where $t_{(n-2)}$ represents Student's t-distribution with $n-2$ of freedom. This test estimates the null hypothesis $H_0: r_{XY} = 0$ versus alternative $H_1: r_{XY} \neq 0$.

Calculated coefficients were put in a two-sided correlation table and represented in a heatmap with a visual representation of colors using the Seaborn statistical library in Python (v3.10). To confirm the assumption of linearity, Scatter plots of selected pairs of key variables with strong positive or negative relationships were generated, and to visually detect possible outliers.

All the analyses were performed in a separate Conda environment using core libraries (pandas, NumPy, matplotlib, seaborn, scikit-learn). Confidentiality of the data was ensured by anonymizing the identifiers of the plants and eliminating any location-related metadata during publication.

**Results**

The final analytical pipeline in Figure 1 represents workflow from CPP to alert system based on emissions diagnostics. The air pollutants data taken by AMS sensors are collected in large tables. Subsequently, tabular data transfers to the server with several units including preprocessing unit, computing unit, and analytical unit. Firstly, preprocessing unit clears the insufficient and incomplete data. Then, the computing unit calculates the Pearson correlation coefficients for each pair of variables to formulate correlational heatmap matrix. Next, based on this matrix, the analytical unit graphs the scatter plots to indicate trends in the emissions data. Finally, after data analysis and processing server estimates the sensitivity of the data and alerts users in critical cases.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 24, OCTOBER 2025

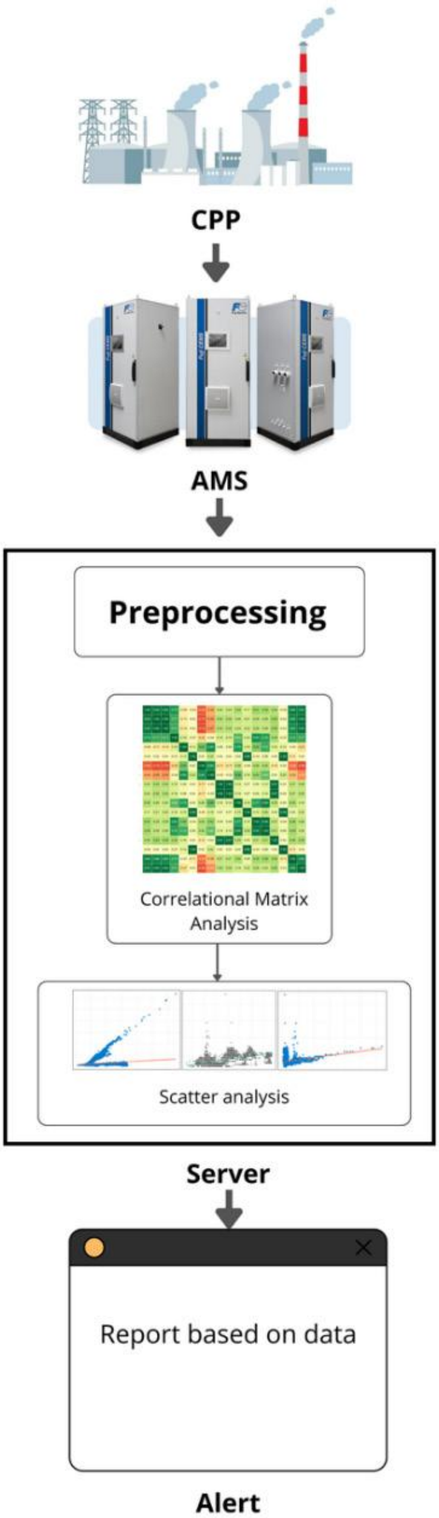


Figure 1. Analytical framework of the study

The correlation-based approach revealed various specific interdependences between the emission parameters. Figure 2 illustrates Pearson correlation table of all the measured process and emission variables. Several operational parameters were found to have strong linear correlations, especially among the steam flow variables and pressures ($r > 0.9$), and the same behaviors were found to be consistent in the control of the plant. It can be observed positive and negative trends of the pollutant constituents.

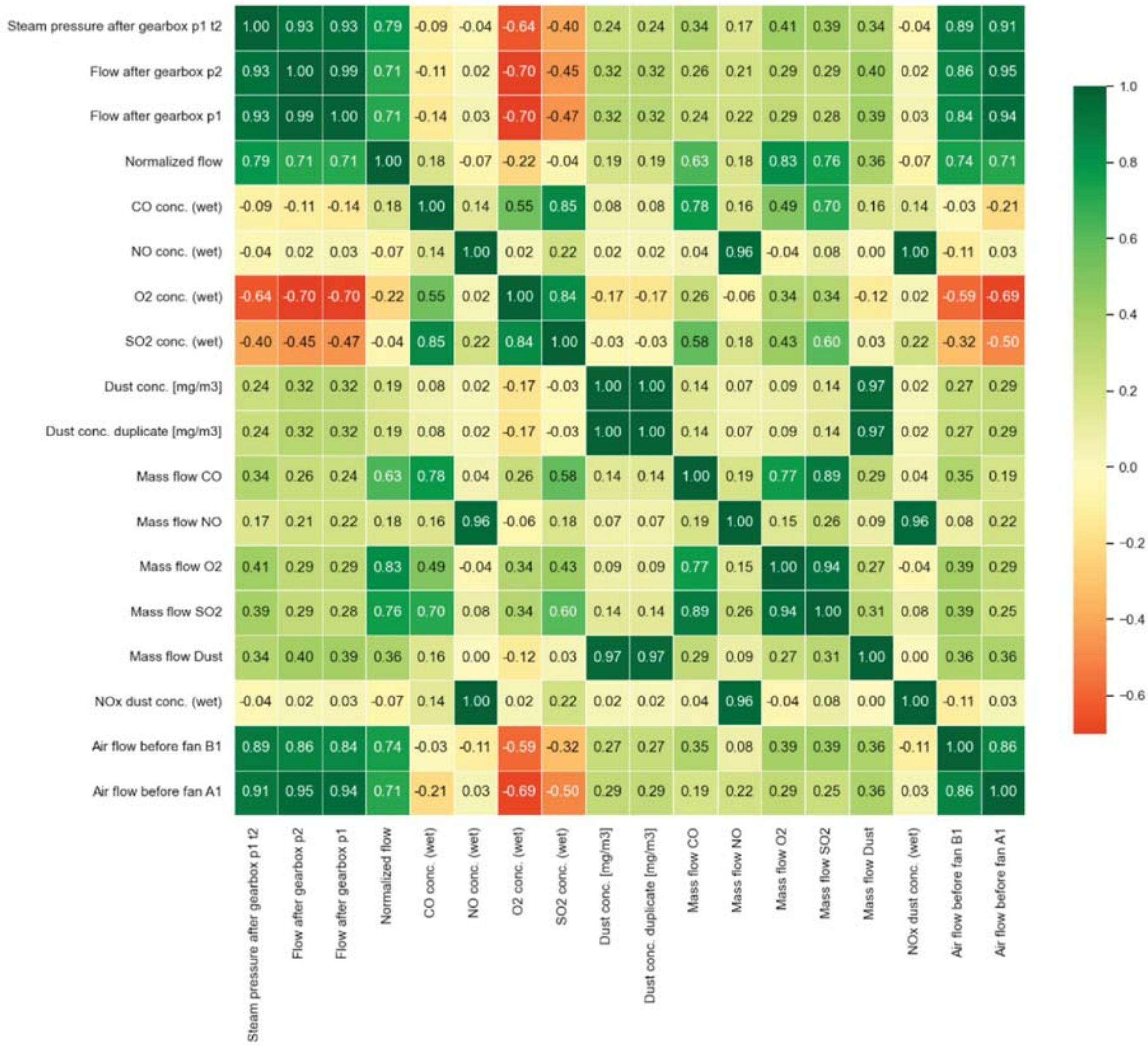


Figure 2. Correlation Matrix of Emission Parameters.

Figure 3 shows subset matrix with four major gaseous pollutants ($CO$, $NO$, $SO_2$, and $O_2$). It demonstrates combustion system internal balance. It was found that the carbon monoxide and sulfur dioxide concentrations had a strong positive correlation (r 0.85), implying that the two pollutants are both affected by factors that determine incomplete combustion or fuel composition.

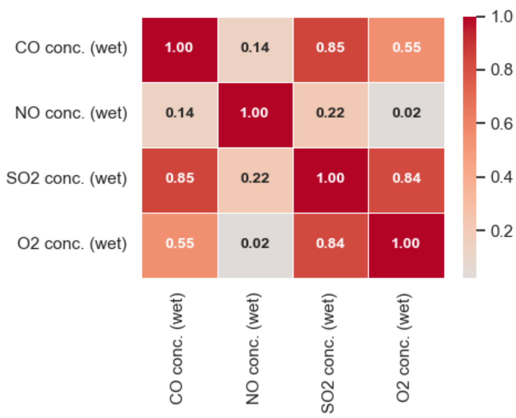


Figure 3. Correlation Between Major Emission Components.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 24, OCTOBER 2025

The correlation is visualized Figure 4, where CO increases proportionally with $SO_2$. It confirms the fact that they have the same source processes.

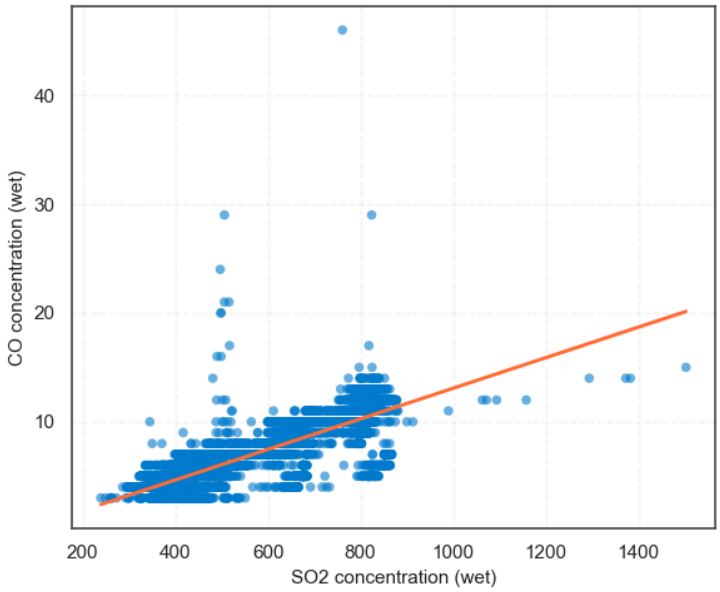


Figure 4. Linear Relationship Between CO and $SO_2$ Concentrations.

Conversely, CO and $O_2$ relationship was negative ($r = -0.55$) as in Figure 5. This opposite tendency is in addition to the principle of combustion-efficiency: increased oxygen concentration generally leads to greater complete oxidation, and, accordingly, to less CO.

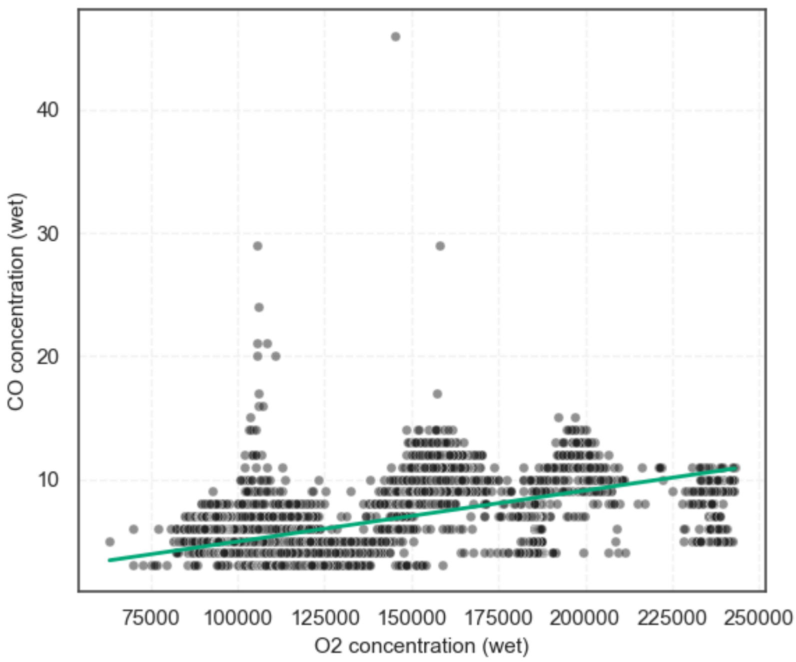


Figure 5. Inverse Relationship Between CO and $O_2$ Concentrations

Figure 6 shows that the CO and NO correlation was low ($r = 0.14$) and this could be because the emitted NO is more vulnerable to temperature and oxidation conditions than the carbon-conversion efficiency.

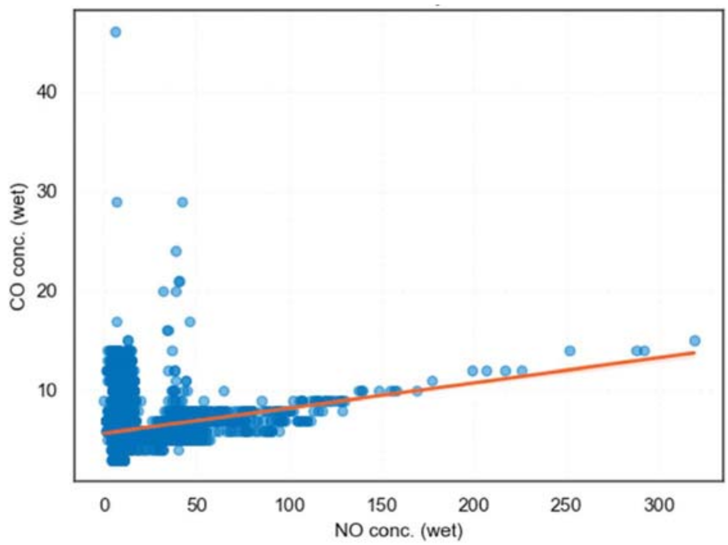


Figure 6. Weak Correlation Between CO and NO Concentrations

Figure 7 demonstrates the three-dimensional interaction between the concentration of CO, $SO_2$, and $O_2$. The color distribution, as well as surface gradient proves the fact that the regions of increased CO have mainly high values of $SO_2$ and low values of $O_2$, which are the signs of drop combustion completeness and less excess-air working. These trends show that the system has a coupled emission behaviour where it is the oxidation of sulfur and carbon that depends on each other and the oxidation of nitrogen oxides under varying thermal conditions.

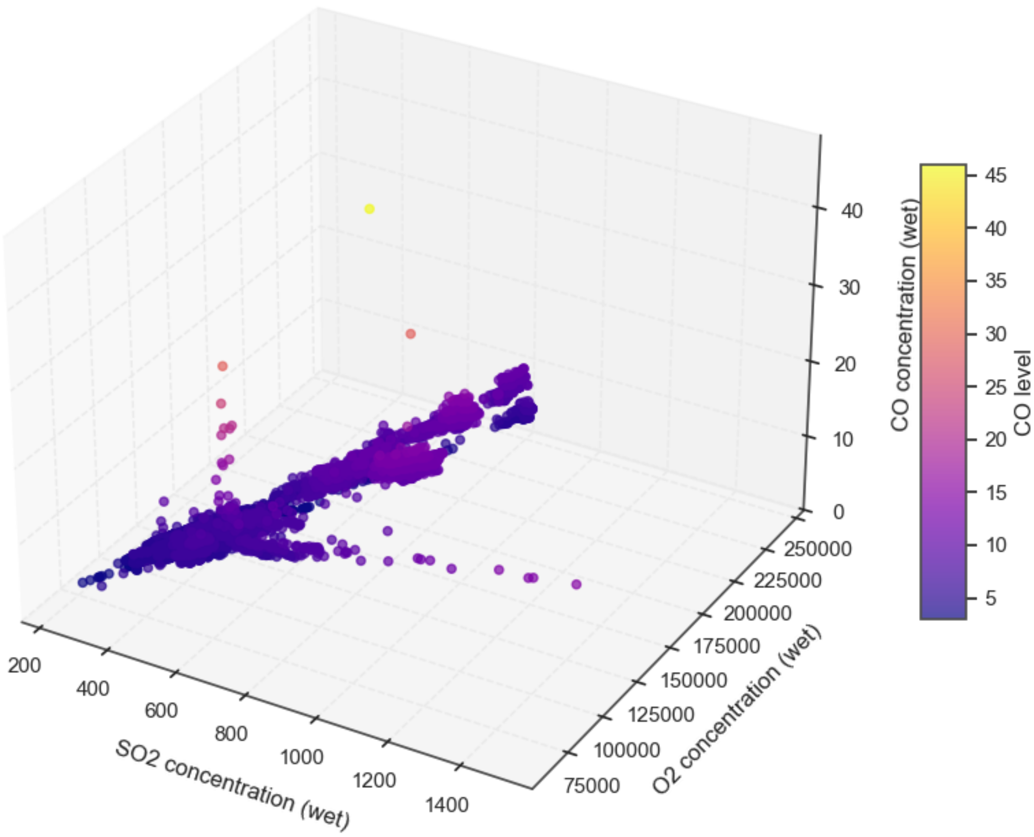


Figure 7. 3D Relationship of CO with $SO_2$ and $O_2$ Concentrations.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 24, OCTOBER 2025

Overall, the findings support the relevance of the Pearson correlation methodology as a diagnostic measure of establishing operational relationships in sets of emissions. The overall linearity of multiple pairs of pollutants, and the possibility to plot combined relationships in 3D, indicates that correlation-based analytics can offer a stable base on which other predictive or anomaly-detection models can be built to rely on.

**Discussion**

The correlations of gaseous pollutants observed provide information on the dynamics of combustion and the formation of emissions that are active in the coal-fired power unit under study. The strongly positive correlation between $CO$ and $SO_2$ concentrations is a positive response, which results in incomplete combustion and fuel sulfur content, which is in line with the results of large-scale monitoring campaigns in East Asia and Central Europe. The ineffectiveness of the air-fuel mixing and local oxygen deficiency have been suggested in those studies to cause the co-variation of these pollutants. The implication of such relationship is that both sulfur oxidation and carbon monoxide production might increase when the excess-air levels are low or when the boiler is in unstable conditions.

The negative correlations between the $CO$ and $O_2$ that were measured in this experiment concur with the overall principles of combustion stoichiometry: the higher the oxygen concentration, the more $CO$ is oxidized to $CO_2$, further lowering the $CO$ concentration of the flue gases. Such a tendency may be used as a diagnosis tool since the absence of the negative correlation between the air temperature and the expected one can indicate the disruption of the airflow system or the failures of the sensors. On the contrary, the low affinity between CO and NO indicates that the formation of nitrogen-oxides is mostly influenced by temperature and residence time and not by the oxidation of carbon. A weak or even non- Stefanic CO -NO relationship in fluidized-bed and staged-combustion systems has been reported in previous investigations utilizing CEMS.

The 3-D representation of the CO as a function of $SO_2$ and $O_2$ concentration gives a better comprehensive view of the emission behavior at operating dynamics. Areas with concomitant elevated $CO$ and $SO_2$ and reduced $O_2$ are a clear indication of inefficient combustion zones, validating the idea of using multivariate correlation surfaces in the process of online diagnostics. The observation validates the argument that correlation matrices, despite being simple, can be used to successfully isolate coupled or competing emission mechanisms with high-frequency monitoring data. Therefore, correlation-based analytics may form a comprehensible initial step of hybrid diagnostic models that are subsequently supplemented by regression or machine-learning models of anomaly detection and predictive maintenance.

**Conclusion**

The current research provided a diagnostic framework based on correlation to examine emission parameters in a coal-burning power station in Kazakhstan. A Pearson correlation coefficient was utilized with over 15,000 observations of high frequency data on an automated monitoring system. Findings indicated that there were good linear associations between the air pollutants namely, CO and $SO_2$ and that there were reverse associations between $CO$ and $O_2$, which indicate underlying combustion and air fuel interactions. A limitation of this study is the lack of provided additional factors to prove hypothesis founded in correlational analysis. This study has identified weak dependencies between CO and NO which highlights that thermal factors affect nitrogen oxides of different formation. The research has also shown that correlation analysis has strong potential to determine the covariance between pollutants and operational factors. Overall, this study strengthens idea that statistical analysis can be used to implement predictive models for emission management systems. Further research could use-fully extended with time dependent correlations and non-linear feature extraction to provide adaptive and robust emissions monitoring system.

**Acknowledgment**

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan within the project BR21882258 "Development of Intelligent Information and Communication Systems Complex for Environmental Emission Monitoring to Make Decisions on Carbon Neutrality".

## References

[1] Krzyzanowski, M., & Cohen, A. (2008). Update of WHO air quality guidelines. *Air Quality, Atmosphere & Health*, *1*(1), 7-13.

[2] Huang, Z., Yu, Q., Ma, W., & Chen, L. (2019). Surveillance efficiency evaluation of air quality monitoring networks for air pollution episodes in industrial parks: Pollution detection and source identification. *Atmospheric Environment*, *215*, 116874.

[3] Biloshchytskyi, A., Kazambayev, I., Toxanov, S., Kirichenko, L., Biloshchytska, S., & Abzhanova, D. (2025, April). Development of Emission Determination Method Concept. In 2025 IEEE Conference on Technologies for Sustainability (SusTech) (pp. 1-7). IEEE.

[4] Energy, G. (2023). Monitor. *Global Gas/Oil infrastructure Tracker. Global Energy Monitor*.

[5] Wang, G., Deng, J., Zhang, Y., Zhang, Q., Duan, L., Hao, J., & Jiang, J. (2020). Air pollutant emissions from coal-fired power plants in China over the past two decades. *Science of the Total Environment*, *741*, 140326.

[6] Xu, Y., Wu, H., Dong, Z., Wang, Q., & Chen, X. (2024). Life cycle energy use efficiency and greenhouse gas emissions of circulating fluidized bed coal-fired plant with coal gangue and coal co-combustion. *Environment, Development and Sustainability*, *26*(8), 20049-20071.

[7] Zhang, L., Zhou, G., Ma, Y., Jing, B., Sun, B., Han, F., ... & Chen, X. (2021). Numerical analysis on spatial distribution for concentration and particle size of particulate pollutants in dust environment at fully mechanized coal mining face. *Powder Technology*, *383*, 143-158.

[8] Luo, H., Zhou, W., Jiskani, I. M., & Wang, Z. (2021). Analyzing characteristics of particulate matter pollution in open-pit coal mines: implications for green mining. *Energies*, *14*(9), 2680.

[9] Wang, K., Xu, H., Peng, Y., & Zhou, S. (2024, August). A novel harmonic emission source locating algorithm based on Pearson correlation coefficient. In *2024 International Applied Computational Electromagnetics Society Symposium (ACES-China)* (pp. 1-3). IEEE.

[10] Thota, S., & Syed, M.B. (2024). Analysis of feature selection techniques for prediction of boiler efficiency in case of coal based power plant using real time data. *International Journal of System Assurance Engineering and Management*, *15*(1), 300-313.

[11] Sittig, D.F., Sherman, J.D., Eckelman, M.J., Draper, A., & Singh, H. (2022). i-CLIMATE: a "clinical climate informatics" action framework to reduce environmental pollution from healthcare. *Journal of the American Medical Informatics Association*, *29*(12), 2153-2160.

[12] Ness, J.E., Heath, G., & Ravi, V. (2022). An Overview of Policies Influencing Air Pollution from the Electricity Sector in Central Asia.

[13] Gu, X., Li, B., Sun, C., Liao, H., Zhao, Y., & Yang, Y. (2022). An improved hourly-resolved NOx emission inventory for power plants based on continuous emission monitoring system (CEMS) database: A case in Jiangsu, China. *Journal of Cleaner Production*, *369*, 133176.

[14] Lukyanets, A., Gura, D., Savinova, O., Kondratenko, L., & Lushkov, R. (2023). Industrial emissions effect into atmospheric air quality: mathematical modeling. *Reviews on Environmental Health*, *38*(2), 385-393.

[15] Hansen, K.N.J., Steinshamn, H., Hansen, S., Koesling, M., Dalgaard, T., & Hansen, B.G. (2025). Comparing different statistical models for predicting greenhouse gas emissions, energy-, and nitrogen intensity. *Computers and Electronics in Agriculture*, *234*, 110209.

[16] Susanto, A.P.Y., Syafei, A.D., Zuki, F.M., Hermana, J., Assomadi, A.F., Erfian, A., & Lazuardi, D.J. (2024). Development of particulate matter emission factors based on CEMS data from PT PLN (Persero)-Owned coal-fired power plants. *Case Studies in Chemical and Environmental Engineering*, *10*, 100971.

[17] Gu, J., Sward, J. A., & Zhang, K. M. (2023). Predicting power plant emissions using public data and machine learning. Environmental Science: Advances, 2(12), 1696-1707.

[18] Yu, H., & Hutson, A. D. (2024). A robust Spearman correlation coefficient permutation test. *Communications in Statistics-Theory and Methods*, *53*(6), 2141-2153.

[19] Xie, X., Qian, N., & Qiu, P. (2024). Online monitoring of air quality using PCA-based sequential learning. *The Annals of Applied Statistics*, *18*(1), 600-625.

[20] Ji, C., & Sun, W. (2022). A review on data-driven process monitoring methods: Characterization and mining of industrial data. *Processes*, *10*(2), 335.

[21] Hoffmann, L., Gilardi, L., Schmitz, M. T., Erbertseder, T., Bittner, M., Wüst, S.,... & Rittweger, J. (2024). Investigating the spatiotemporal associations between meteorological conditions and air pollution in the federal state Baden-Württemberg (Germany). *Scientific Reports*, *14*(1), 5997.

[22] Yan, W., Wang, J., Lu, S., Zhou, M., & Peng, X. (2023). A review of real-time fault diagnosis methods for industrial smart manufacturing. *Processes*, *11*(2), 369.

[23] Erfian, A., Syafei, A. D., Zuki, F. M., Hermana, J., Assomadi, A. F., Susanto, A. P. Y., & Lazuardi, D. J. (2025). Determining Carbon Dioxide Emission Factors of Indonesia Coal-Fired Power Plants with CEMS Measurement Data. *Aerosol and Air Quality Research*, *25*(1), 1-12.

[24] Smadi, A. A., Ajao, B. T., Johnson, B. K., Lei, H., Chakhchoukh, Y., & Abu Al-Haija, Q. (2021). A Comprehensive Survey on Cyber-Physical Smart Grid Testbed Architectures: Requirements and Challenges. *Electronics*, *10*(9), 1043. https://doi.org/10.3390/electronics10091043

[25] Zhou, Y., Wang, K., Zhang, Y., Liang, D., & Jia, L. (2025). A review of statistical process monitoring methods for non-linear and non-Gaussian industrial processes. *The Canadian Journal of Chemical Engineering*, *103*(7), 3092-3119.

[26] Alnaim, A., Sun, Z., & Tong, D. (2022). Evaluating machine learning and remote sensing in monitoring NO2 emission of power plants. *Remote Sensing*, *14*(3), 729.

[27] He, H., Schäfer, B., & Beck, C. (2022). Spatial heterogeneity of air pollution statistics in Europe. *Scientific Reports*, *12*(1), 12215.