

Article

AI-Driven Random Forest Model and the Six Sigma Approach for Enhancing Offset Printing Process and Product Quality

Diana Bratić ^{1,*} , Petar Miljković ², Denis Jurečić ¹ and Tvrtko Grabarić ¹

¹ Faculty of Graphic Arts, University of Zagreb, Getaldićeva 2, 10 000 Zagreb, Croatia; denis.jurecic@grf.unizg.hr (D.J.); tvrtko.grabarić@grf.unizg.hr (T.G.)

² Department for Multimedia, University North, Jurja Križanića 31b, 42 000 Varaždin, Croatia; pmiljkovic@unin.hr

* Correspondence: diana.bratice@grf.unizg.hr

Abstract: The Six Sigma methodology for quality improvement enabled a high degree of process compliance and enhanced process capability. This research develops a new model for optimizing the offset printing process based on the Six Sigma approach, with the aim of reducing process variability and achieving stable, predictable production outcomes. Special focus was placed on defining Critical Product Characteristics (CPCs) and Critical to Quality (CTQs) points and analysing their impact on process output quality, defined by the sigma level. Based on the research, variability limits of production parameters were defined to ensure consistency and high product quality. The integration of Artificial Intelligence (AI) within the Six Sigma framework allowed for additional automation and model adaptation to changing production conditions. The use of the Random Forest model enabled efficient analysis of critical variability points, prediction of potential deviations, and real-time process adjustment. AI is utilized to improve precision and efficiency in quality management, which further enhances process stability and optimization in line with the dynamic demands of modern production. The proposed model represents an innovative approach that facilitates maintaining stable production results and provides a sustainable foundation for future process optimizations in the printing industry.

Keywords: offset printing; optimization; quality; Six Sigma; process variability; Artificial Intelligence; machine learning; Random Forest



Academic Editor: James M. Griffin

Received: 19 October 2024

Revised: 14 January 2025

Accepted: 15 February 2025

Published: 20 February 2025

Citation: Bratić, D.; Miljković, P.; Jurečić, D.; Grabarić, T. AI-Driven Random Forest Model and the Six Sigma Approach for Enhancing Offset Printing Process and Product Quality. *Appl. Sci.* **2025**, *15*, 2266. <https://doi.org/10.3390/app15052266>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Quality has become a key factor for competitiveness in the industry, and Six Sigma is one of the most effective methodologies for improving quality and reducing variability in production processes. Six Sigma is a methodology based on statistical analysis, aiming to reduce process variability to achieve the highest possible level of quality. The primary goal of Six Sigma is to reach a defect rate of 3.4 defects per million opportunities, which results in an almost perfect process [1–3].

The printing industry, although technologically advanced, faces numerous challenges in maintaining production stability and quality. Introducing the Six Sigma methodology can bring significant benefits. Although Six Sigma was initially developed for production processes in other industries, its application in the printing industry is becoming more prevalent and holds great potential for improving performance and achieving production process stability.

Variability in the printing process can be caused by various factors such as material variations, machine settings, and human factors. By implementing the Six Sigma method-

ology, it is possible to identify and reduce these variations, leading to more consistent production. The optimization process involves defining Critical to Quality (CTQs) points and analyzing process capability. Through the analysis of these points, it is possible to identify the main sources of variability and take steps toward their elimination.

This paper explores the possibilities of applying an AI-driven Six Sigma methodology in graphic production, with a particular focus on the offset printing process. In contrast to traditional applications of Six Sigma, which are typically employed as a framework for optimizing the entire production process, this study introduces the principles of Six Sigma as a numerical indicator for evaluating the quality and consistency of individual printed sheets. The Sigma level serves not only as a quantitative measure to assess the stability and predictability of the offset printing process but also as a basis for the identification and continuous optimization of key process parameters. By integrating statistical methods with the Random Forest model, this study provides a data-driven framework for determining optimal parameter combinations to achieve the highest possible sigma level corresponding to first-class print quality. This approach provides practitioners with an accurate tool to monitor and improve print quality in real time, actionable insights on parameter adjustments, and a methodology to stabilize offset printing processes. It also ensures that quality assessment is both objective and accurate, minimizing reliance on subjective assessment methods. By combining robust statistical analysis and machine learning, this methodology provides a scalable solution that can be adapted to a variety of production environments. It also bridges the gap between theoretical advances and real-world applications in graphic production by enabling consistent, high-quality results and reducing production variability.

The Six Sigma methodology encompasses the definition of critical product or process characteristics, measurement of current performance, data analysis to identify the causes of variability, implementation of improvements, and the establishment of control mechanisms to ensure the maintenance of achieved results. The introduction of the Six Sigma methodology in the graphic industry requires adaptation to the specific characteristics of that industry, but its basic principles can be applied with the goal of achieving more stable and efficient production processes.

Thus, an analysis of existing research and theories was conducted to identify key factors affecting production variability and propose measures for AI-driven optimization and standardization of the offset printing process.

Numerous studies explore the application of the Six Sigma methodology and its modalities across various industries. However, in recent years, the application of Six Sigma has been contextualized with artificial intelligence. The rapid advancement of technology, particularly AI, has significantly impacted organizations, especially within Industry 4.0, where traditional methods like Six Sigma face limitations when dealing with complex, high-volume data [4–8]. Adopting AI-based strategies in manufacturing improves decision-making, productivity, and system performance [9]. Specifically, the integration of Artificial Intelligence (AI) and Lean Six Sigma (LSS) practices has the potential to revolutionize operational excellence through enhanced digital capabilities and innovative applications [10,11].

Given the generally limited number of published works in the field of graphic technology, particularly printing, this paper draws upon the application of AI-driven Six Sigma methodology in other manufacturing sectors.

In the context of industrial production, the primary goal of engineering is continuous improvement in process performance by maximizing manufacturing efficiency and product quality. To achieve these goals, advanced, robust process optimization techniques have been designed, implemented, and applied to the manufacturing process. Starting from the Six Sigma approach, an analysis of the production process of ultrasound (US) probes

for medical imaging was conducted, and the PDCA (Plan-Do-Check-Act) methodology was implemented, with an emphasis on robust optimization [12]. Perera et al. [13] utilized artificial intelligence, specifically Natural Language Processing (NLP), to bridge gaps in understanding the causal mechanisms that contribute to Lean Six Sigma (LSS) success. The study developed a streamlined model, highlighting how LSS elements drive quality performance, customer satisfaction, and overall business efficiency, with validation against the existing Six Sigma frameworks to enhance its explanatory power.

A group of authors employed the Surface Tension Neural Network (STNN) as an innovative tool within the Lean Six Sigma framework to improve efficiency in the food industry. By using neural networks, they achieved a deeper understanding of the relationships between key production variables and enabled real-time process control. This resulted in increased operational efficiency and reduced waste, directly impacting production sustainability [14]. Nader [15], in a case study, illustrates the successful implementation of LSS tools, such as the design of experiments, which resulted in efficiency improvements in a brewery, emphasizing the potential of AI-augmented LSS for various industries.

Sharma and Singh [16] describe a modified LSS 4.0 model for a sustainable textile industry, integrating Lean, Six Sigma, and Industry 4.0 technologies to optimize processes, reduce defects, and enhance sustainability, quality, and competitiveness within the sector.

An advanced algorithm was developed to accurately measure the geometry and printability of shape patterns in nanoparticle-based printed electronic devices, aiming to establish international standards. The algorithm uses image processing techniques to quantify edge waviness and widening across the entire pattern boundary, leading to reduced deviation in pattern dimensions [17].

Artificial intelligence (AI) is transforming various engineering fields, including the corrugated board industry. This study demonstrated how AI, particularly artificial neural networks (ANNs), can predict the crush resistance of corrugated packaging, encompassing typical boxes as well as those with ventilation holes or perforations. By optimizing input parameters such as material properties, box dimensions, and structural features like openings, the study achieved a predictive model with an error below 10%, showing the potential for efficient compressive strength prediction and improved load-bearing calculations in the corrugated packaging industry [18,19].

Although it is about printing, three-dimensional (3D) printing is not necessarily related to graphic production. Integration of Six Sigma (6S) with additive manufacturing (AM) addresses quality management challenges like process repeatability and customization [20]. Three-dimensional (3D) printing, also known as additive manufacturing (AM), has already shown its potential by demonstrating remarkable applications in various manufacturing. A group of authors presented a novel concept of utilizing artificial intelligence (AI) to support quality control in the additive manufacturing (AM) of medical devices made from polymeric materials. The publication aims to demonstrate how AI enhances the efficiency, adaptability, and speed of the inspection process, bringing innovative scientific and technological solutions with significant economic and social impact [21].

An overview of machine learning-driven 3D printing technology, given by Zhang et al. [22], highlights its advancements in process optimization, monitoring, and motion planning. They found that supervised learning is effective for process optimization and surface quality inspection, while reinforcement learning and deep learning excel in path planning for multi-degree of freedom printing platforms, such as those involving a 6-DOF robotic arm. These AI advancements pave the way for innovative applications in additive manufacturing.

Recent advancements in integrating artificial intelligence (AI) with 3D printing have enhanced design precision, material selection, and production efficiency, ultimately fostering an environment-friendly manufacturing process. Authors [23], in their review, highlight

the potential of AI-driven predictive modelling, quality control, and design optimization in advancing 3D printing towards sustainable and efficient production solutions.

Building on the advances in three-dimensional (3D) printing, the work of a group of authors [24] presents an intriguing and forward-looking study that integrates machine learning (ML) to predict raster angles in fused deposition modelling (FDM), optimizing additive manufacturing (AM) processes. Using algorithms like Random Forest Regression (RFR), the research achieves high prediction accuracy, reducing production costs and lead times without compromising quality. This novel application of ML to FDM raster angle estimation represents a promising advancement for improving efficiency in AM.

In the field of 3D printing, interesting is the work that systematically explores the integration of artificial intelligence (AI) in the three-dimensional (3D) printing process for architectural structures, highlighting its transformative potential. AI is utilized to optimize design, printing parameters, and quality control, providing real-time feedback and enhancing efficiency [25].

Furthermore, Lean Six Sigma (LSS) has been enhanced with machine learning and artificial neural networks (ANNs) to tackle high scrap rates in Small Mixed Batch (SMB) production. This study applies the improved LSS methodology in a case involving bakery machine manufacturing, demonstrating process improvements through DMAIC and an ANN-based model to predict critical-to-quality (CTQs) characteristics, ultimately reducing variability in input materials, demonstrating the effectiveness of combining AI with LSS for improved quality and sustainability [26].

Taking into account all the aforementioned studies and analyses, the application of Six Sigma principles as a mathematical framework in the graphics industry offers significant opportunities to increase production stability, reduce variability, and achieve a high level of quality.

Importantly, this study does not employ Lean Six Sigma (LSS) as a structured process improvement methodology, nor does it incorporate its standardized tools such as DMAIC cycles, predefined process control frameworks, or hierarchical belt certification structures. Instead, the Six Sigma level in this research serves solely as a high-resolution quantitative metric for print quality assessment, mathematically derived through a structured correlation between Critical to Quality factors (CTQs) and Critical Process Characteristics (CPCs) to quantify process capability and ensure a statistically grounded evaluation of print stability.

The proposed model introduces a methodologically rigorous and mathematically formalized approach to print quality assessment, establishing a systematic relationship between sigma levels and key process parameters. Unlike conventional Six Sigma implementations that focus on holistic process optimization, this model contextualizes sigma as a probabilistic measure of conformance within a defined quality spectrum, directly linking it to the functional requirements of print output rather than to a broader process control strategy. While Six Sigma methodologies are traditionally applied in high-volume manufacturing to optimize entire production chains, the approach presented in this study is fundamentally different. It does not seek to standardize operational procedures or minimize process inefficiencies across a complex production environment. Instead, it is designed to function as a precise mathematical descriptor of print quality variability, offering a quantifiable means of evaluating deviations in critical process characteristics relative to defined quality thresholds.

The novelty of this model lies in its ability to mathematically formalize print quality evaluation through the integration of sigma levels into a structured analytical framework. This enables a more precise, objective, and scalable methodology for assessing print stability and quality consistency across different production scenarios. While this framework

has been developed within the offset printing domain, its generalizable architecture and analytical scalability enable its application to a broader range of quality-influencing factors within the printing process. Furthermore, the intrinsic adaptability of the model allows for its direct transposition to alternative printing technologies, reinforcing its potential as a universally applicable and theoretically extensible paradigm for the quantitative assessment of print quality across diverse technological environments.

2. Modelling Background

In order to establish models for the optimization and control of the production process, this chapter provides the background of modelling and defining key parameters related to product quality in the context of offset printing. To provide a clearer perspective on the research process, Figure 1 outlines the main stages of the study, the methods employed, and the connections between these stages.

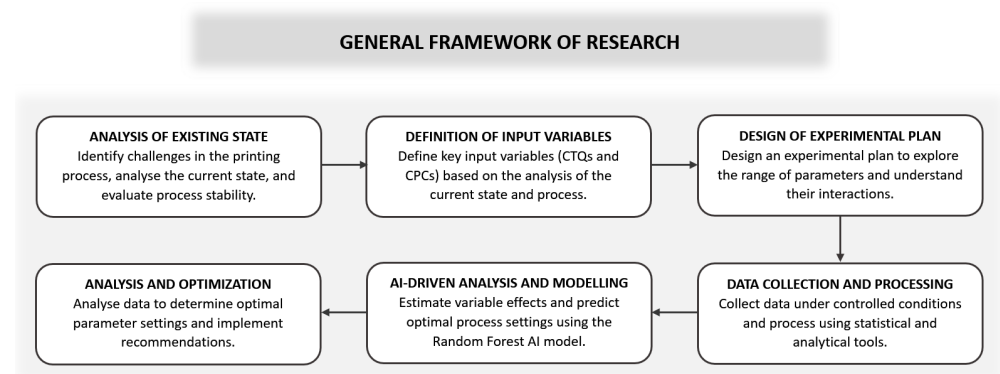


Figure 1. General Framework of Research.

This framework provides an overview of the study and its methodological approach, serving as the basis for the modelling and analysis presented in this chapter.

2.1. Defining Critical to Quality and Critical Product Characteristics

Considering the principles of the Six Sigma methodology, it was necessary to define potential variability limits in the process by determining the functional requirements of the product and the methods for controlling them. Critical Product Characteristics (CPCs) were also defined, analysed, and grouped according to cause categories. The functional requirements of the product are most often defined by its design and can be categorized as [27,28] product usability, ergonomic adaptability, technical reliability, and aesthetic sensitivity.

Based on the functional requirements, the current sigma level ($k\sigma$) of the process was calculated. To identify Critical to Quality (CTQs) points through process analysis, Critical Product Characteristics (CPCs) were defined as dependent variables, which include:

- Dot gain at 40% halftone value for cyan, magenta, yellow and black inks
- Dot gain at 80% halftone value for cyan, magenta, yellow and black inks
- Ink density at 100% halftone value for cyan, magenta, yellow and black inks
- Grey balance
- Print register (alignment)
- Geometric deformations of halftone elements (slurring, doubling, smudging)
- Paper folding register
- Paper cutting register
- Print rub register

Critical Product Characteristics (CPCs) are the points on the product that are closely associated to the Critical to Quality (CTQs) points.

- W_1 : ink temperature ($^{\circ}\text{C}$)—crucial for maintaining ink viscosity stability and uniform ink transfer
- W_2 : dampening solution temperature ($^{\circ}\text{C}$)—ensures stable transfer of the dampening solution to the rollers and process balance
- W_3 : dampening solution acidity (pH)—controls chemical stability during printing
- X_1 : paper temperature ($^{\circ}\text{C}$)—ensures the dimensional stability of the paper during the printing process
- X_2 : paper humidity (%)—is related to the hygroscopic nature of the paper and its stability during the printing process
- Z_1 : ink viscosity (mPa·s)—ensures consistent ink application and print quality
- Z_2 : dampening solution conductivity ($\mu\text{S}/\text{cm}$)—enabled proper interaction between the dampening solution and the paper surface
- Z_3 : dampening solution hardness (dH)—helps to maintain the stability of the solution during the printing process
- Z_4 : alcohol content in the dampening solution (% vol.)—crucial for stable solution transfer and preventing excessive evaporation.

From the analysis of the relevant literature and processes, it was concluded that the most critical and characteristic points for quality in the offset printing process are the ink temperature, the dampening solution temperature, and the pH of the dampening solution [29–55]. These independent variables in the model, identified through preliminary process analysis and research, directly influence the chemical and physical interactions within the printing process. Their control is crucial for maintaining consistency, stability, and high-quality print outcomes. In particular, ink temperature (W_1) affects drying speed, dot gain, and colour balance; dampening solution temperature (W_2) is essential for maintaining process stability and the balance between ink and water; and the dampening solution acidity (W_3) directly affects the chemical stability of the solution and ensures proper ink distribution.

The impact of these variables on the Critical to Quality (CTQs) characteristics was analysed using statistical tools such as correlation analysis and descriptive statistics. The selection of these parameters was based on their strong association with quality metrics such as process stability and print accuracy. The specification limits for these parameters, listed in Table 1, include the lower specification limit (LSL), the upper specification limit (USL), and the target value (TV). These limits were defined to minimize process variability, considering industry standards and experimental research results to minimise process variations to ensure optimal process performance and reduce variability that could negatively impact print quality.

Table 1. Defined Specification Limits of Variables in the Model (LSL, USL, TV).

Var	W_1	W_2	W_3	C40	M40	Y40	K40	C80	M80	Y80	K80	C100	M100	Y100	K100
LSL	21.0	8.0	4.0	9.0	9.0	9.0	12.0	8.2	8.2	8.2	9.0	1.5	1.5	1.35	1.8
USL	24.0	12.0	7.0	17.0	17.0	17.0	20.0	13.2	13.2	13.2	14.0	1.55	1.55	1.40	1.8
TV	22.0	10.2	5.6	13.0	13.0	13.0	16.0	10.7	10.7	10.7	11.5	1.53	1.53	1.38	1.8

This trio of parameters has a direct impact on a wide range of chemical and physical reactions that are critical to the quality of multicolour and multitone reproduction. These reactions can be described and monitored by Critical Product Characteristics (CPCs) such

as dot gain, ink density, grey balance, and print register, ensuring a high level of quality control in the offset printing process.

Additionally, the maximum range for each characteristic was defined, determined by the lower and upper specification limits (LSL or Lower Specification Limit and USL or Upper Specification Limit), as well as the target value (TV) [56].

Table 1 showed the specified limits (LSL—Lower Specification Limit, USL—Upper Specification Limit, TV—Target Value) for the key controlled variables in the experimental model. The target values for W_1 (ink temperature) and W_2 (dampening solution temperature) are not centered but are defined according to the machine specifications, which recommend optimal settings to ensure process stability. For W_1 , the decentered target value ensures the stability of ink viscosity and consistent transfer to the substrate, while for W_2 it helps to maintain the balance between ink and dampening solution as well as the stability of the printing process. In contrast, W_3 (dampening solution acidity) is centered within its limits to ensure chemical stability and consistent print quality. The decentralized target values reflect the technical requirements of the machine and the process-specific characteristics, while the limits for other variables are aligned with industry standards and the analysis of the initial process.

The values listed in Table 1 include the lower (LSL) and upper (USL) specification limits as well as the target values (TVs) for key process variables. The tolerances are defined to ensure that the variables remain within the permissible limits and thus meet the process stability and quality requirements. The target values (TVs) are adjusted according to the machine specifications and process characteristics so that minor variations without affecting the quality of the final product.

The following diagram (Figure 2) illustrates the model of the offset printing process, including key input variables (ink temperature, dampening solution temperature, and dampening solution acidity), output variables (CTQs, CPCs, and sigma level), controlled constants (paper temperature and paper humidity), and absolute constants (ink viscosity, dampening solution conductivity, dampening solution hardness, and alcohol content in the dampening solution). This model illustrates the relationships between parameters, whereby the input variables define the process conditions, the output variables assess quality, and the constants represent controlled conditions.

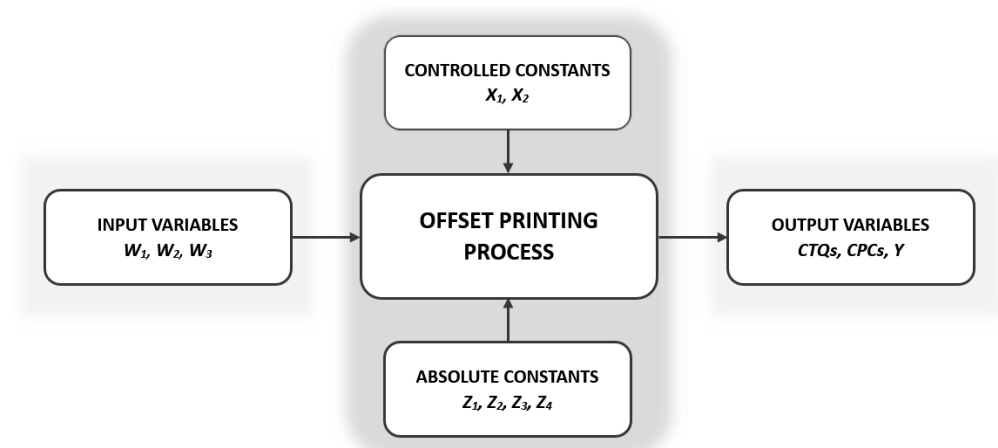


Figure 2. Model of the Offset Printing Process.

The specification limits (Lower Specification Limit, Upper Specification Limit) and target values (Target Value) are defined based on experimental data, and the output variables are used to calculate the sigma level.

For additional clarity, the factors in the model are categorized as follows:

- Controlled variable factors (W_1, W_2, W_3): the main factors or variables in the experimental model with clearly defined values
- Controlled constants (X_1, X_2): Stable input factors or controlled constants that are not varied but are important for the model and are included in the calculation of the number of repetitions
- Non-controlled or absolute constants (Z_1, Z_2, Z_3, Z_4): Factors that are kept as absolute constants during the experiment to eliminate their influence
- Process output Y : CPCs required for calculating the sigma level

2.2. Materials and Equipment

The research was conducted under strictly defined and controlled microenvironment conditions to minimize any potential external impact of temperature and humidity on the printing process. Therefore, the temperature in the production facility was maintained at 22 °C, with a relative air humidity of 45%. Printing was performed on a six-colour KBA Rapida 105 press, B1 format. All prints were made on identical printing substrates, Brigl & Bergmeister BioMatt 130 g/m².

Quality control of the offset printing process was carried out both densitometrically and spectrophotometrically using control strips on the printed sheets. For this purpose, measurement strips for grey balance, relative print contrast, tonal errors, and ink efficiency were used, along with signal strips that indicate geometric deformation of halftone elements (i.e., for detecting slurring, doubling, and smudging), as well as signal strips for detecting print register and alignment errors (Figures 3 and 4).

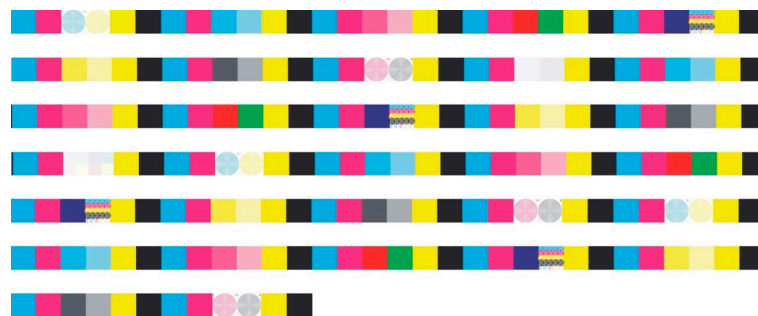


Figure 3. Control Strip KBA K C215.

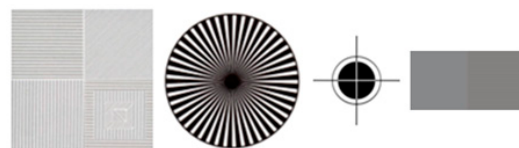


Figure 4. Signal Strip and Grey Balance Strip Ugra.

For measurements, a thermometer and hygrometer for paper (the so-called ‘sword’) Dostmann P300W (Dostmann electronic GmbH, Wertheim am Main, Germany), a densitometer TECHKON SpectroDens D-61462 A603013 Advanced (TECHKON GmbH, Königstein im Taunus, Germany), and a photometer Hanna Instruments HI96735 (Hanna Instruments, Woonsocket, RI, USA) were used, along with the KBA Rapida 105 LogoTronic and ErgoTronic professional software (for regulation and control of all machine conditions) and KBA Rapida 105 DensiTronic (for regulation and control of ink application) (Koenig & Bauer Gruppe GmbH, Würzburg, Germany).

The key process parameters, i.e., main input variables, in the future model are:

- W_1 (ink temperature): the ink temperature is defined as a key process parameter and is automatically set and controlled by the ErgoTronic system integrated into the press.

The ink temperature set point is maintained within a precise range of ± 0.1 °C to ensure stable ink viscosity and consistent application to the offset plate. Continuous monitoring and regulation of the temperature prevents possible variations in ink viscosity, which are essential for consistent printing at all halftone values.

- W_2 (dampening solution temperature): the dampening solution temperature is defined as a key process parameter and is automatically set and controlled by the ErgoTronic system. The temperature is maintained within the specified range with a regulation precision of ± 0.1 °C. Controlling the dampening solution temperature ensures optimal stability in solution transfer to the offset plate, minimizing the transfer of dampening solution to the paper and ensuring the dimensional stability of the paper and the print quality.
- W_3 (acidity of the dampening solution): the acidity of the dampening solution is defined as a key process parameter and is precisely set and controlled by the ErgoTronic system with a regulation accuracy of ± 0.1 . Proper pH control ensures the chemical stability of the dampening solution throughout the printing process and prevents unwanted reactions between the dampening solution and the ink. The pH value is defined according to industry standards.

The output variables (CTQs) associated with Critical Product Characteristics (CPCs) and required for the calculation of the sigma level are controlled and maintained within the defined values through the integration of the DensiTronic and ErgoTronic systems. DensiTronic analyses and measures colour parameters, including dot gain at different halftone values, ink density, and grey balance. The system uses spectral and densitometric measurements to detect deviations from the specified values, with the results displayed on the ErgoTronic console. It also detects geometric deformations of halftone elements (slurring, doubling, and smudging). ErgoTronic uses the data obtained from DensiTronic to precisely adjust the machine settings and controls the registers (print register, paper folding and cutting register, and print rub register). The control strips are additionally verified using a densitometer.

In this study, controlled factors are defined as process parameters with predetermined values that are kept constant throughout the experiment. These parameters play a key role in ensuring the stability of the conditions and the reliability of the research results. They were regulated by precise measuring devices and control systems integrated into the printing machine. The controlled factors are:

- X_1 (paper temperature): defined at 22 °C, measured with a Dostmann P300W thermometer (i.e., 'sword'), to prevent dimensional deformation of the paper and maintain stability during the printing process.
- X_2 (paper humidity): defined at 45%, measured with a Dostmann P300W hygrometer (i.e., 'sword'). The humidity content of the paper is related to its hygroscopic nature and directly affects dimensional stability.
- Z_1 (dampening solution conductivity): defined at 1950 $\mu\text{S}/\text{cm}$, a value chosen based on industry standards and previous research to ensure proper interaction of the solution with the paper surface. Conductivity is measured using a digital conductometer integrated into the ErgoTronic system, enabling continuous monitoring and precise control of conductivity throughout the process. This system automatically adjusts the dampening solution concentration to maintain the defined conductivity level, eliminating the possibility of variations and ensuring process stability.
- Z_2 (alcohol content in the dampening solution): defined at 8.5%, which is the optimal value according to industry standards for stable printing operation and proper distribution of the dampening solution on the roller surface. The alcohol content is measured using an alcohol concentration sensor integrated into the ErgoTronic system,

enabling continuous monitoring and automatic regulation of the alcohol level. This system ensures that the alcohol content remains within the defined limits, preventing excessive evaporation or improper solution mixture, which could affect press stability and print quality.

- Z_3 (ink viscosity): defined at 160 mPa·s for cyan, 180 mPa·s for magenta, 140 mPa·s for yellow and 150 mPa·s for black, according to the specifications of factory-prepared inks used in the experiment. Since the viscosity of the ink is not regulated during the printing process, inks with factory-defined values were used, which comply with industry standards for offset printing. The stability of the ink viscosity was ensured by controlling the ink temperature (W_1) using the ErgoTronic system, minimizing viscosity changes caused by temperature fluctuations during the process. Maintaining proper viscosity was crucial for consistent ink application and uniform print quality across different halftone values.
- Z_4 (dampening solution hardness): defined at 9 dH, in line with industry standards for offset printing. This value ensures optimal stability of the dampening solution and prevents the formation of deposits on rollers and other parts of the press. The hardness of the solution was determined based on the chemical properties of the water used to prepare it and adjusted with appropriate additives before the printing process began. The hardness measurement was performed using the Hanna Instruments HI96735 digital photometer to analyze water hardness. After the initial setting, hardness was not regulated during printing, as this parameter remains stable throughout the process.

The ErgoTronic system ensures precise regulation and maintenance of parameters throughout the printing process, eliminating the need for additional manual measurements.

3. Preliminary Process Analysis

The research was based on determining the process potential and fulfilling the process function through preliminary process capability indices and sigma levels in order to identify all influential factors of process variation. Before determining the key variables in experimental planning, it was necessary to gain insight into the actual state of the printing process. Process analysis was essential to verify whether the process was in a state of statistical control and whether it followed a normal distribution, indicating stability, which allows for the prediction of its performance. When the process is under control, there is a lower likelihood that the observed process parameters will exceed the specified control limits. It was also necessary to verify if the process was properly centered in relation to the already defined and monitored Critical Product Characteristics (CPCs).

In addition to controlling the characteristics related to the functional requirements of the product, it was important to monitor the factors influencing the process. These included the paper temperature (°C) and humidity (%), the ink viscosity (mPa·s) and ink conductivity (μS/cm), the hardness (dH), and the alcohol content in the dampening solution (% vol.). The paper temperature and humidity are defined as controlled constants, while the ink viscosity, dampening solution conductivity, hardness, and the alcohol content in the dampening solution are defined as absolute constants. Therefore, it was not necessary to define LSL (Lower Specification Limit), USL (Upper Specification Limit), or TV (Target Value) for them.

This approach is in line with the principles of the Six Sigma methodology, which emphasizes process stability and control as a prerequisite for reducing variability and improving quality. Constant values for these parameters minimized the influence of uncontrolled factors on the process. This allowed the analysis to focus on optimizing key variables and ensuring process performance within predefined quality standards.

It is important to note that the frequency and sample size of 100 measurements were based on the preliminary process capability assessment, which is carried out at the beginning of the process or after a relatively short monitoring period, where a sample of at least 100 units is taken.

3.1. Descriptive Analysis of Variables in the Model

In the first step, a descriptive analysis of the obtained data was conducted to assess their consistency and stability. This analysis provides insight into the statistical characteristics of the variables, allowing for the identification of deviations from the target values and an evaluation of compliance with the specification limits (LSL and USL). Descriptive analysis also enabled the identification of potential issues in the data, such as outliers, skewness, or high coefficients of variation, which may indicate the presence of variability within the process [57].

Table 2 presents a descriptive analysis of the key variables in this study, which provides a detailed insight into the behaviour of the offset printing process. The following statistical characteristics were quantified:

Table 2. Descriptive Statistical Analysis of Key Variables in the Model.

Var	N	Min	Max	Mean	Median	Mode	Var	Std. Dev.	Co. Var.	Skew.	Kurt.
W ₁	100	26.9	27.2	27.024	27.0	1.0	0.0039	0.0622	0.2230	0.3042	0.3345
W ₂	100	9.2	10.5	10.0	10.1	1.0	0.1147	0.3387	3.3874	−0.6523	−0.4379
W ₃	100	5.5	6.1	5.751	5.7	1.0	0.0187	0.1367	2.3769	0.4870	−0.1629
C40	100	13.5	14.1	13.816	13.8	1.0	0.0161	0.1269	0.9185	0.6402	0.2448
M40	100	13.4	13.8	13.588	13.6	1.0	0.0071	0.0844	0.6213	0.2325	−0.5494
Y40	100	13.6	14.0	13.752	13.7	1.0	0.0106	0.1030	0.7487	0.3418	−0.4664
K40	100	16.6	17.0	16.807	16.8	1.0	0.0108	0.1037	0.6171	−0.1422	−0.3098
C80	100	11.3	11.7	11.486	11.5	1.0	0.0079	0.0888	0.7730	0.1922	−0.2935
M80	100	11.0	11.4	11.253	11.3	1.0	0.0102	0.1010	0.8972	−0.7445	0.2602
Y80	100	11.4	11.9	11.63	11.6	1.0	0.0122	0.1106	0.9506	0.2929	−0.2066
K80	100	12.1	12.6	12.338	12.3	1.0	0.0095	0.0972	0.7879	0.1116	0.4783
C100	100	1.48	1.56	1.5189	1.52	1.0	0.0003	0.0166	1.0909	0.1514	−0.4351
M100	100	1.49	1.56	1.5252	1.52	1.0	0.0002	0.0154	1.0101	−0.0468	−0.5247
Y100	100	1.34	1.39	1.365	1.36	1.0	0.0002	0.0125	0.9167	0.0789	−0.4870
K100	100	1.78	1.82	1.8004	1.8	1.0	0.0000	0.0042	0.2358	1.0581	14.2819

- Minimum and Maximum: define the extreme values of the data for each variable and illustrate the range within which the variables are distributed.
- Mean: Represents the arithmetic mean of the values for each variable and provides information about its central tendency.
- Median: Indicates the middle value of the data and divides the dataset into two equal halves, regardless of the extreme values.
- Mode: indicates the most frequently recorded value, which is useful for identifying dominant patterns in the data.
- Variance: quantifies the overall variability of the data and indicates how far the values deviate from the mean.
- Standard deviation: represents the average deviation of the data from the mean, which is a key indicator of process stability.

- Coefficient of variation: expresses the relative variability as a percentage and enabled the comparison of variables with different measures.
- Skewness: indicates an imbalance in the data distribution. Positive values indicate a higher frequency of low values, while negative values indicate the opposite trend.
- Kurtosis: describes the shape of the data distribution. A high positive value indicates a peaked distribution, while a negative value indicates a flat and broad distribution.

These statistics provide a basic insight into the stability, distribution, and variability of a process and enable model validation and the identification of optimization opportunities.

The analysis showed that all variables are within the specified tolerance limits and the deviations from the target values are negligible. All measured values in Table 2 remained within the specification limits, with the exception of variable W_1 (ink temperature), whose mean value was above the upper limit defined in Table 1 (USL: 24.0 °C). This deviation is the result of the specific operating conditions in the pre-analysis phase, in which the actual operating settings were adapted to industrial practice and not to the experimental conditions. For the purpose of the experiment, the controlled parameters were adjusted within the limits defined by the specifications, ensuring complete process stability and precision of the results.

This indicates process stability and high-quality control in the production process. In addition, the symmetrical distribution of most variables and the low coefficient of variation indicate a high level of measurement consistency, which is critical for maintaining the robustness and predictability of the process in accordance with Six Sigma standards. These results indicate that key production parameters are well defined and controlled, thereby minimizing variability and ensuring process quality and stability. Further analysis of the key variables (W_1 , W_2 , W_3) provides deeper insights into their behaviour and relationships:

- W_1 (ink temperature): showed a low coefficient of variation (Co. Var. = 0.22%) and a low standard deviation (Std. Dev. = 0.0622), which indicates a high stability of this variable. This consistency is crucial for maintaining the viscosity and uniform application of the ink.
- W_2 (dampening solution temperature): showed a relatively higher variability (Co. Var. = 3.39%) and a negative skewness (Skew. = −0.6523), indicating a tendency towards lower temperature values. This indicates the need for precise control to maintain process stability.
- W_3 (dampening solution acidity): showed a standard deviation of 0.1367 and a positive skewness (Skew. = 0.4870), indicating a slight tendency towards higher pH values. Such deviations could have an effect on the chemical stability of the process.

Furthermore, interrelationships among these variables, especially the strong correlation between W_2 and W_3 , confirm their mutual influence on the overall performance of the process. This underlines the importance of collective regulation of these parameters to ensure optimal process conditions.

This descriptive analysis confirms the process stability and provides a robust foundation for the optimization steps in this research. Consistent data from the analysis enabled the development of a reliable, AI-driven Six Sigma model that aims to further reduce variability, improve production performance, and enhance overall process quality.

3.2. Calculation of the Current Sigma Level of the Process

Before implementing the Six Sigma methodology or a model based on it, it was necessary to calculate the current sigma level of the process, which should be at least within the range of average global companies, i.e., $(3 \div 4) \cdot \sigma$ [58]. The calculation of the current sigma level for processes ($k\sigma$), based on Allen [3], is determined by the number of products or outputs from the process, the number of requirements that define the conformity of the

process outputs, and the number of errors in the processes. The mathematical expression is as follows:

$$\delta = \frac{\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n}{G \times z} \quad (1)$$

where: δ is errors per unit, ε is the number of errors for a defined requirement, G is the total number of products (process outputs), and z is the number of requirements.

Since sigma is evaluated per million units, the result (errors per unit) is multiplied by one million to obtain the number of errors per million products or process outputs (N).

$$N = \delta \cdot 10^6 \quad (2)$$

Table 3 showed the current, initial sigma level of the process to verify whether the process has a sigma level above 3, which represents a critical threshold for assessing process stability and quality and for determining the need to introduce the Six Sigma model for further optimization and reduction in variability. These 20 requirements are related to the Critical Product Characteristics (CPCs), key characteristics defined according to quality standards. They cover specific aspects of the product that have a direct impact on its functionality and aesthetic value.

Table 3. Calculation of the Current Sigma Level of the Process.

G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
100	20	0	0	0	0	0	0	0	0	5	4	33	0.01650	16,500	3.625
		ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
		6	9	2	2	1	2	2	0	0	0				

where: ε_1 is the number of errors in dot gain at a 40% halftone value for cyan, ε_2 is the number of errors in dot gain at a 40% halftone value for magenta, ε_3 is the number of errors in dot gain at a 40% halftone value for yellow, ε_4 is the number of errors in dot gain at a 40% halftone value for black, ε_5 is the number of errors in dot gain at an 80% halftone value for cyan, ε_6 is the number of errors in dot gain at an 80% halftone value for magenta, ε_7 is the number of errors in dot gain at an 80% halftone value for yellow, ε_8 is the number of errors in dot gain at an 80% halftone value for black, ε_9 is the number of errors in the 100% coverage value for cyan, ε_{10} is the number of errors in the 100% coverage value for magenta, ε_{11} is the number of errors in the 100% coverage value for yellow, ε_{12} is the number of errors in the 100% coverage value for black, ε_{13} is the number of errors in the print register, ε_{14} is the number of errors in halftone element slurring, ε_{15} is the number of errors in halftone element doubling, ε_{16} is the number of errors in halftone element smudging, ε_{17} is the number of errors in grey balance, ε_{18} is the number of errors in the folding register, ε_{19} is the number of errors in the cutting register, and ε_{20} is the number of errors in print rub resistance.

Table 3 showed the current sigma level of the process ($k\sigma = 3.625$), calculated based on the total number of products ($G = 100$), the number of requirements ($z = 20$), and the total number of defects ($\Sigma\varepsilon = 33$). The Errors per unit ($\varepsilon/u = 0.0165$) are defined as the ratio between the total number of defects and the total number of products and requirements. The parameter ε/u (errors per unit) represents the number of defects per unit and is defined as the ratio of the total number of defects to the total number of products and requirements, as expressed in Equation (1). This metric serves as a key measure of process performance, providing a direct assessment of defect rates within the production process. This value is used to calculate the number of defects per million products ($N = 16,500$) and the sigma level ($k\sigma$). It is important to emphasize that δ is not the standard deviation of the population but a key indicator that enabled the evaluation of process stability and quality within the Six Sigma methodology. Unlike standard deviation, which measures dispersion within a dataset, δ is derived directly from defect rates and reflects process performance rather than statistical variance. The sigma level of $k\sigma = 3.625$ includes a standard sigma shift of 1.5, which is common in the Six Sigma methodology for processes that are not perfectly stable. The sigma shift of 1.5 takes into account potential long-term fluctuations in the process

that may occur due to changes in input parameters, environmental conditions, or other uncontrolled factors.

Here, k represents the process sigma level, which is calculated based on the defect rate and adjusted for long-term process variations using the standard sigma shift. The primary sigma level was calculated before including the sigma shift based on the number of defects per million products ($N = 16,500$) and is 2.13. The initial sigma value was determined using standard Six Sigma conversion tables, which correlate defect rates (DPMO) with corresponding sigma levels. By adding a sigma shift of 1.5, the final sigma level was adjusted to $k\sigma = 3.625$, which better reflects the long-term stability and predictability of the process. This practice enabled more consistent assessments of process quality and ensured that results are comparable to industry standards.

Therefore, it is evident that the current sigma level of the process is 3.625, which meets the initial conditions for the implementation of a model based on the Six Sigma approach.

A calculation of the current sigma level ($k\sigma$) was also performed, considering the number of process outputs, the requirements defining the conformity of process outputs, and the number of errors in the processes, along with the calculation of preliminary process capability. This approach to calculating the current sigma level is directly tied to the product's functional requirements and is based on the number of products coming out of the process, the number of requirements defining the conformity of the process outputs, and the number of errors in the processes. The sigma level value obtained from this calculation, based on the aforementioned set of dependent variables, was used as a dependent output variable in the optimization model.

3.3. Calculation of the Preliminary Process Capability

Understanding the structure of the process and quantifying its performance are essential for improving it and successfully implementing and executing the Six Sigma model. Therefore, process capability analysis is an extremely important and well-defined tool of statistical process control, serving the function of continuous quality improvement [59]. Process capability is its ability to produce a product that meets specifications. A process will be capable when all products are produced within the given tolerances based on objective evidence regarding process performance [1]. In other words, process capability refers to a process that can produce a product within specific tolerance intervals for certain quality characteristics [59]. Process capability is defined as the range that encompasses all potential values of specified quality characteristics produced by the process under defined conditions [60,61]. The Six Sigma methodology evaluates process performance by considering the shift (σ) within the process; however, assessing process capability through Process Capability Indices (PCIs) offers a more comprehensive understanding of the process. PCIs focus on key aspects such as process consistency, uniformity, and losses within the process [62,63].

Preliminary process capability assessment is performed at the beginning of the process or after a relatively short period of process monitoring (with a sample size of at least 100 units). In this context, the term 'performance' is used instead of 'capability' in the index nomenclature, but the indices are calculated in the same way as process capability indices. Process capability, measured using the C_p index, relates to the variation in the process around the mean value. Therefore, the index measures potential capability, assuming that the process mean is equal to the midpoint of the specification limits and that the process operates under statistical control [1,60]. If the process is centered within the tolerance limits, i.e., when the $\bar{x} = (LSL + USL)/2$, the index is equal to:

$$C_p = \frac{USL - LSL}{6\sigma} \quad (3)$$

As the minimum acceptable value for this index, a value of 1.33 is taken, meaning that the width of the range of acceptable values or process tolerance should not exceed 75% of the USL–LSL span. Since the process is not centered within the tolerance limits but is closer to one of the limits, i.e., the mean is often not at the midpoint, the resulting measure is the C_{pk} index.

$$C_{pk} = \text{Min} \left[\frac{USL - \bar{\bar{x}}}{3\sigma}, \frac{\bar{\bar{x}} - LSL}{3\sigma} \right] \quad (4)$$

It is also considered acceptable if its value is at least 1.33.

For the purposes of this study, the following preliminary process capability indices were calculated (Table 4):

- Pp —potential process capability
- Pr —capability ratio
- Ppk —demonstrated performance
- Ppl —lower potential capability
- Ppu —upper potential capability
- K —non-centring correction factor

Table 4. Preliminary Process Capability Indices.

Var	Lower Const C (±)	Upper Const C (±)	LSL	Nom	USL	Pp	Pr	Ppk	Ppl	Ppu	K
W_1	−3.0000	3.0000	26.00	27.00	28.00	5.3634	0.1865	5.2346	5.4921	5.2346	0.0240
W_2	−3.0000	3.0000	8.00	10.20	12.00	1.9681	0.5081	1.9681	1.9681	1.9681	0.0000
W_3	−3.0000	3.0000	4.00	5.60	7.00	3.6577	0.2734	3.0457	4.2698	3.0457	0.1673
C40	−3.0000	3.0000	9.00	13.00	17.00	10.5065	0.0952	8.3632	12.6498	8.3632	0.2040
M40	−3.0000	3.0000	9.00	13.00	17.00	15.7935	0.0631	13.4718	18.1151	13.4718	0.1470
Y40	−3.0000	3.0000	9.00	13.00	17.00	12.9492	0.0772	10.5148	15.3837	10.5148	0.1880
K40	−3.0000	3.0000	12.00	16.00	20.00	12.854	0.0777	10.2613	15.4481	10.2613	0.2018
C80	−3.0000	3.0000	8.20	10.70	13.20	9.3859	0.1065	6.4350	12.3369	6.4350	0.3144
M80	−3.0000	3.0000	8.20	10.70	13.20	8.2541	0.1212	6.4283	10.0799	6.4283	0.2212
Y80	−3.0000	3.0000	8.20	10.70	13.20	7.5378	0.1327	4.7337	10.3418	4.7337	0.3720
K80	−3.0000	3.0000	9.00	11.50	14.00	0.9350	1.0695	0.5879	1.2821	0.5879	0.3712
C100	−3.0000	3.0000	1.50	1.53	1.55	0.5029	1.9883	0.3802	0.3802	0.6257	0.2440
M100	−3.0000	3.0000	1.50	1.53	1.55	0.5409	1.8487	0.5366	0.5453	0.5366	0.0080
Y100	−3.0000	3.0000	1.35	1.38	1.40	0.6660	1.5015	0.3996	0.3996	0.9324	0.4000

The data required to calculate the Pp values, including the sample means, are listed in Table 2 and were used for the calculations shown in Table 4.

The value of the Pp index should be greater than 1.33, but it is evident that the variables $K80$, $C100$, $M100$, and $Y100$ have a lower value than this threshold. Since the Pr index is the reciprocal of the Pp index, deviations in the same variables are expected, meaning the index value is greater than 1. The Ppl and Ppk indices, representing the lower and upper process capabilities, are also below the defined values for these variables. The Pp and Pr indices alone do not indicate how the process is positioned relative to the specification limits, which can be determined by comparing these two indices. Identical values suggest complete process centering (the index values are equal to the Cp index), while a value less

than 1 indicates non-conformance. Moreover, the process shifts towards the specification limit with the smaller index value. The Pp index can be corrected for non-centring by calculating the non-centring correction factor K . This gives the demonstrated performance index Ppk . For perfectly centered processes, K equals zero, and the Ppk index is equal to the Pp index. As the process moves away from the target value (midpoint of the tolerance range), K increases, and the value of the Ppk index becomes smaller than the Pp index.

As shown in Table 4, all Ppk index values are smaller than the Pp index values, and the correction factor K equals zero only for the process variable W_2 , where the values of the Pp , Ppl , and Ppk indices are equal, indicating that the process is fully centered (green markings of index values). Orange markings in the table indicate the borderline index values, while red markings indicate deviations. Deviations are evident in the 80% halftone value of black and in the 100% ink coverage. These deviations are due to the fact that for this substrate type, a slightly thinner ink layer can achieve the same visual effect and satisfactory aesthetic quality from the consumer's perspective. At the same time, it simplifies and reduces the cost of the production process by shortening drying time and reducing the amount of silicate needed in the infrared drying process. From the process analysis, it can be concluded that the process meets the initial conditions for implementing a model based on the Six Sigma approach.

4. Multi-Criteria Process Analysis

4.1. Defining Factors and Factor Combinations in Multi-Criteria Analysis

Given that the experimental concept on which the model for optimizing the offset printing process was based relies on the ability to control influential factors in the studied process, the design of the research model itself was based on a stochastic approach, more specifically with technologically acceptable parameters. Therefore, all limiting factors in the problem were considered, as well as the optimality criteria defined by the objective function. This objective function is essentially a formalized (analytical) form of assessing the desired outcome in the process.

The next step, in accordance with the defined critical product characteristics, Critical to Quality (CTQs) points, and process analysis, was to determine the maximum range for each characteristic, i.e., the factors of multi-criteria analysis and the variation in the process for each of the mentioned characteristics. Given that the process analysis revealed the most common deviations of ± 0.2 from the target value, the ranges in which the three controlled factors (independent variables) will extend, as well as their combinations, are shown in Table 5. The deviation of ± 0.2 from the target values refers to the most frequently observed differences between the actual and target values of the controlled factors under stable conditions of the initial process. This deviation is not defined as a tolerance but as the empirically observed range of variation within which the factors W_1 (ink temperature), W_2 (dampening solution temperature), and W_3 (dampening solution acidity) have varied under normal conditions.

Table 5. Combinations of Optimization Factor Triads.

Variable	W_1 Ink Temperature Range ($^{\circ}\text{C}$)	W_2 Dampening Solution Temperature Range ($^{\circ}\text{C}$)	W_3 Dampening Solution Acidity Range (pH)
Combination	26.8 = $W_{1,1}$	10.0 = $W_{2,1}$	5.4 = $W_{3,1}$
	27.0 = $W_{1,2}$	10.2 = $W_{2,2}$	5.6 = $W_{3,2}$
	27.2 = $W_{1,3}$	10.4 = $W_{2,3}$	5.8 = $W_{3,3}$

This range of ± 0.2 was defined as the basis for determining the modalities of the controlled factors in the experimental model in order to study their interactions on the output variable and determine the optimal settings for each one. The modalities of W_1 (ink temperature) were defined as $W_{1,1}$, $W_{1,2}$, and $W_{1,3}$, while the modalities of W_2 (dampening solution temperature) and W_3 (dampening solution acidity) were defined in the same way ($W_{2,1}$, $W_{2,2}$, $W_{2,3}$, and $W_{3,1}$, $W_{3,2}$, $W_{3,3}$).

An experimental model is based on these modalities, including all possible combinations ($3^3 = 27$ combinations), to evaluate their impact on the sigma level of the process. The main objective of this approach is to identify the optimal combination of modalities that leads to the highest sigma level, thus minimizing process variability and ensuring consistent production quality. This approach enabled a systematic study of how the variability of controlled factors affects process stability, providing a basis for improving and optimizing production.

Furthermore, the selection of these three controlled factors is based on a detailed analysis of recent literature in the field of the offset printing process. The studies show that W_1 , W_2 , and W_3 are key parameters that have a direct impact on print quality and process stability, making them the most important input variables for the analysis. For example, maintaining a stable ink temperature (W_1) regulates viscosity, while W_2 and W_3 ensure the balance between ink and dampening solution, thus minimizing errors in the transfer of information from the offset cylinder to the paper.

The values listed in Table 5 are based on target values (TV) defined according to the machine specifications and industry standards, while the mean and median values in Table 2 reflect the actual process behaviour in the initial state.

According to the established combinations of optimization triads, 27 series (3^3) of partial measurements, i.e., partial processes, were carried out with a defined number of experimental repetitions under strictly defined plant conditions and printing press settings.

4.2. Defining the Number of Experimental Repetitions

Since the degree of the multi-factorial plan is directly related to the degree of the mathematical model and given that this is a second-degree mathematical model, a second-degree multi-factorial experimental design will be applied. After defining the process input values, i.e., independent influencing factors (independence assumed in the first approximation), it was necessary to determine the levels of factor variation. Orthogonal plans with two-level factor variation are most commonly used, i.e., $x_{i,min}$ and $x_{i,max}$. Plans in which factors are varied at two levels are called 2^k , which represents the number of factors, or independent variables [56].

The total number of experimental repetitions for each combination of measurement conditions was calculated using the following formula:

$$N = 2^k + 2k + n \quad (5)$$

where: N is the total number of experimental repetitions, k is the number of factors, and n is the number of repetitions at the center of the experiment to ensure statistical stability by minimizing the influence of variability in the experimental conditions. Central point repetitions play a crucial role in estimating inherent process variation and improving the reliability of the derived mathematical model.

In this study, the value of $n = 10$ was determined based on the factorial experimental design and the structural requirements of the second-degree mathematical model. The selection of ten central repetitions was made to ensure a sufficient number of data points for evaluating process stability while maintaining experimental feasibility. The applied design incorporated 3^3 combinations of three controlled factors, while two additional fixed

factors were introduced to maintain stable experimental conditions. The number of central repetitions was optimized to improve model accuracy and mitigate variability due to uncontrolled influences, in accordance with established experimental design principles.

The formula for calculating the number of repetitions is based on five factors, where X_1 (paper temperature) and X_2 (paper humidity) are treated as controlled constant factors. Although they were not varied in the experiment, their inclusion in the formula ensures consistency of experimental conditions and reduces potential variability.

The three controlled factors (W_1, W_2, W_3) were defined with three levels, resulting in a total of 27 combinations (3^3) in the experimental model. In addition, the constant factors (X_1, X_2) were set to stable initial values to ensure the consistency of the input conditions during all measurements.

This approach enabled the analysis of interactions between variable factors at the sigma level, while the constant factors contributed to the stability of the experimental process. In addition, the uncontrolled factors (Z_1, Z_2, Z_3, Z_4) were kept constant during the experiments to eliminate their influence on the results.

Using $k = 5$ and $n = 10$, the total number of repetitions was calculated as:

$$N = 2^5 + 2 \cdot 5 + 10 = 52 \quad (6)$$

Therefore, with 3 factors, 3^3 or 27 combinations were obtained for the controlled measurement conditions, and 52 repeated measurements were carried out for each condition.

4.3. Defining the Sampling Plan

A simple random sample of size n elements is obtained from a population that has N elements if the selection is made such that each sample of size n that can be chosen from that population has the same probability of being selected. However, for the purposes of this study, the systematic sampling method was used. In this method, the sample fraction f was calculated as the ratio of the number of units in the sample to the total population size [64]:

$$f = \frac{n}{N} \quad (7)$$

where: f is the sample fraction, n is the number of units in the sample (52), and N is the total population size, corresponding to the total production volume of 15,000 printed sheets. This distinction clarifies the statistical approach used and ensures consistency in the interpretation of the parameters.

From which it follows:

$$f = \frac{15,000}{52} = 0.00347 \quad (8)$$

The reciprocal value of the sample fraction, known as the sampling step, is calculated as:

$$\frac{N}{n} = \frac{15,000}{52} = 288 \quad (9)$$

This means that one of 288 units of the population was selected. To extract the sample, the starting point was determined by choosing a number within the range [1, 288], in this case the number 100. This number was chosen to facilitate tracking on the machine counter. From this point, every 288th unit was selected until all 52 units were included in the sample.

5. Modelling

The values of the output factor Y (current sigma level of the process) were calculated in the same manner as the initial sigma level when assessing the process's suitability for further analysis and the introduction of the Six Sigma-based model (Table A1). As expected, all current sigma levels fall within the range of $(3.6 \div 3.8) \cdot \sigma$, which is within the $(3 \div 4) \cdot$

σ range. This satisfies the initial condition for designing the optimization model based on the Six Sigma approach.

The first step in modelling was selecting the mathematical apparatus that would provide the model, in the first approximation, with a sufficiently accurate representation of the actual, unknown, analytical form of the response function. Defining the mathematical model not only involves determining its degree but also making a predictive selection of the independent influencing parameters that enter the model. All other potential parameters must remain constant [65,66].

5.1. Approximation of the Model Function

A mathematical model, if adequate, represents a suitable approximation of the actual, unknown, analytical form of the response function.

For the purposes of this study, a multivariable quadratic function is assumed, of the form:

$$F(x_1, x_2) = a \cdot x_1^2 + b \cdot x_2^2 + c \cdot x_1 x_2 + d \cdot x_1 + e \cdot x_2 + f \quad (10)$$

If, in a process where the actual state function is unknown, inputs $(x_i, i = 1, 2, 3, \dots, k)$, are defined, then the mathematical process model can be written in the form:

$$\eta = \varnothing(x_1, x_2, x_3, \dots, x_k) \quad (11)$$

The function η is a hypothetical quantity, and in the mathematical model obtained after experimental testing, the experimental error ε is also present, so the following expression holds:

$$\eta = \varnothing(x_1, x_2, x_3, \dots, x_k) + \varepsilon \quad (12)$$

Since there is non-linearity in the model and a linear function is not an adequate approximation, a higher-order polynomial, typically second-order, is required:

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ij} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \varepsilon \quad (13)$$

Given that data on the value of the function f in points $x_1, x_2, \dots, x_m$, i.e., $f(x_1), f(x_2), \dots, f(x_m)$, has been obtained, the goal is to approximate the function f with a polynomial:

$$h(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 \quad (14)$$

where n is the degree of the polynomial, and $n < m$ ensures that the polynomial is properly fitted to the available data points without overfitting, allowing for a balance between model flexibility and accuracy in function representation.

For each polynomial, corresponding errors or deviations of the approximation are defined:

$$\varepsilon_i = f(x_i) - h(x_i), i = 1, 2, \dots, m \quad (15)$$

The best approximation is the one for which the sum of squared deviations is minimal, i.e., for which:

$$\min f(a_0, a_1, \dots, a_n) = \sum_{i=1}^m [f(x_i) - h(x)]^2 \quad (16)$$

or:

$$\min f(a_0, a_1, \dots, a_n) = \sum_{i=1}^m [f(x_i) - (a_n x_i^n + a_{n-1} x_i^{n-1} + \dots + a_1 x_i + a_0)]^2 \quad (17)$$

Given the relationship between three parameters (W_1, W_2, W_3) and the output value Y , the goal was to obtain an equation that best describes this relationship. Once the equation is obtained, it can easily be optimized to find the specific input values that produce the best output, i.e., 6σ . To simplify the process of finding an appropriate equation that best

describes the system's response to input parameters, the least squares method was used. This method was the most appropriate and straightforward in this case, as it attempts to minimize the deviation from the model and the known output data y , ensuring that the deviation is no greater than 1% from the measured results, as was later proven.

Thus, the goal is to minimize the function:

$$R^2 \equiv \sum [y_i - f(x_i, a_1, a_2, \dots, a_n)]^2 \quad (18)$$

By finding the minimum, the process description that is closest to the measured outputs y_i based on the inputs x_i is obtained.

It was determined that parameters a i b exist such that, for each value of the independent variable X , the dependent variable Y can be written as:

$$Y = x^T A x + B x + c + \varepsilon \quad (19)$$

where: A is an upper triangular matrix, B is a vector, c is a scalar, x is the input vector, and ε is a normal random variable with an expectation $E(\varepsilon) = 0$ and standard deviation $V(\varepsilon) = \sigma^2$, and σ^2 is constant for all x .

Of all possible lines $y = ax + b$, the most probable regression line is the one for which the sum of squared deviations is minimal.

$$f(a_1, a_2, \dots, b_1, b_2, \dots, c) = \sum_{i=1}^N \left[\vec{y}_i - \vec{x}_i^T A \vec{x}_i + B^T \vec{x}_i + c \right]^2 \quad (20)$$

The sum of squared deviations is minimized when the following hold simultaneously:

$$\frac{\partial f}{\partial a_1}, \frac{\partial f}{\partial a_2}, \dots, \frac{\partial f}{\partial b_1}, \dots, \frac{\partial f}{\partial c} = 0 \quad (21)$$

The polynomial that approximates the function is:

$$y = c_1 W_1^2 + c_2 W_2^2 + c_3 W_3^2 + c_4 W_1 W_2 + c_5 W_1 W_3 + c_6 W_2 W_3 + c_7 W_1 + c_8 W_2 + c_9 W_3 + c_{10} \quad (22)$$

For the future model, the equation is a function of three variables:

$$Y = f(w_1, w_2, w_3) \quad (23)$$

For simplicity of notation and simulation, the given function is written in vector form, with input parameters in vector form. This allows the result vector to be obtained for any input matrix for each individual case from the table.

Furthermore, in matrix form, the given non-linear function is:

$$Y = W^T \cdot A \cdot W + B \cdot W + C \quad (24)$$

where: Y is the vector of output values, W is the vector of input values, A is the matrix of constants from c_1 to c_6 , B is the vector of constants from c_7 to c_9 , and C is the constant c_{10} .

The full form of the non-linear function is:

$$Y = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}^T \cdot \begin{bmatrix} c_1 & c_4 & c_5 \\ 0 & c_2 & c_6 \\ 0 & 0 & c_3 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} + \begin{bmatrix} c_7 & c_8 & c_9 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} + c_{10} \quad (25)$$

For each combination of W_1 , W_2 , and W_3 from Table 5, a different vector Y . This equation is subtracted from the measured values, then squared, and the result is assigned to a variable, continuing the process until all 27 cases are covered.

5.2. AI Model Selection

As traditional tools like Six Sigma and statistical process control reach their limitations, AI-based methods are becoming essential for enabling sustainable and efficient manufacturing practices [67,68]. Artificial Intelligence methods, such as machine learning algorithms (e.g., Random Forest, support vector machines), artificial neural networks (ANNs), deep learning models, and natural language processing (NLP), can be effectively utilized in Six Sigma (6σ) projects for optimization, predictive analysis, and process automation [69]. The Random Forest model is suitable for predictive control in non-linear production processes due to its ability to accurately incorporate control process knowledge, providing more reliable future state predictions and enhancing overall control performance [70].

The Random Forest formed by decision trees is a mainstream bagging algorithm that performs better than single algorithms. The Random Forest model's function depends on using classification and regression trees (CARTs) for establishing base estimators [71]. Random Forest is one of the most popular machine learning algorithms used for solving regression and classification problems. It is based on the idea of aggregating multiple decisions through an ensemble of decision trees, where each tree contributes to the final prediction. This approach not only improves the model's accuracy but also reduces variability and overfitting, which often occur with simpler models like individual decision trees [72,73].

The Random Forest model was chosen for its interpretability, ability to handle large datasets, robustness, feature importance assessment, and reduction in overfitting due to its use of the 'bagging' (bootstrap aggregating) technique. It is well-suited for this non-linear model because it can efficiently identify and model complex non-linear relationships between the input variables (W_1 , W_2 , W_3) and the output variable Y , while also mitigating overfitting, which further enhances the model's accuracy in predicting outcomes. Random Forest relies on the structure of decision trees, which use various mathematical techniques to make decisions. Each tree is trained on different subsets of data, and the final prediction is obtained by averaging the predictions of all trees (in the case of regression) or by majority voting (in the case of classification). This approach improves prediction accuracy and reduces the risk of overfitting.

Random Forest is an ensemble method composed of multiple decision trees. Each tree makes its own estimate (prediction), and the final prediction of the Random Forest model is the average of all individual tree estimates (for regression) or the majority of votes (for classification). For each model trained, a random sample of data are taken from the original dataset through a process called 'bootstrap sampling'. At each node of the tree, instead of considering all features for the split, Random Forest randomly selects only a subset of features. This technique helps reduce correlation among the trees and makes the model more robust. Once the Random Forest model is trained, predictions for new data are made by aggregating the averages of the individual trees. Given that Random Forest uses a set of multiple trees (N) to calculate the final prediction as the average of all individual predictions ($f_i(x)$) obtained from each tree in the forest, it can be mathematically expressed as:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f_i(x) \quad (26)$$

where: $\hat{y}$ is the final prediction of the model, N is the number of trees in the forest, and $f_i(x)$ is the prediction of the i -th tree for input x .

The flowchart (Figure 5) illustrates the process of the Random Forest algorithm for regression, including bootstrap sampling to create multiple training subsets, building decision trees for each subset, and aggregating their predictions to derive a final output. This method aims to enhance prediction accuracy and reduce the variance of the model.

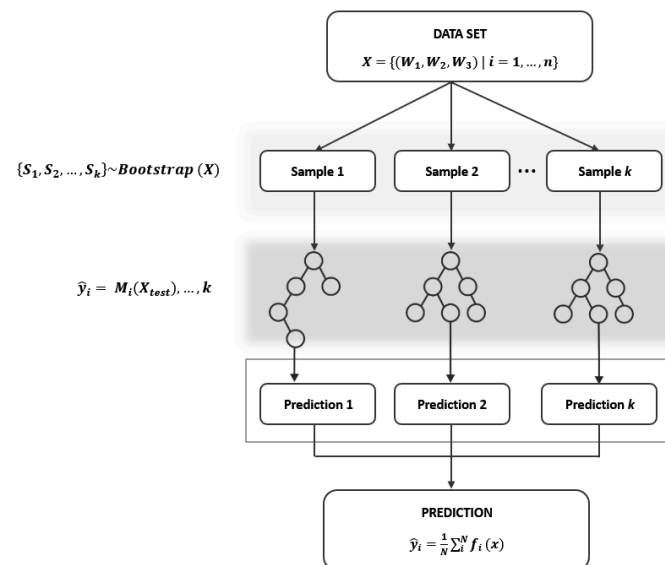


Figure 5. Random Forest Regression Flowchart with Bootstrap Sampling.

5.3. Model and Interpretation of Results

For the creation of the optimization model, the Python programming language was used due to its flexibility and extensive support for machine learning and data analysis, along with the application of appropriate libraries and algorithms to enable precise modelling and process optimization (Figure A1).

The following libraries were used:

- Pandas (import pandas as pd) for loading data from a CSV file (pd.read_csv) and managing data using the DataFrame structure. A DataFrame represents a tabular data format that allows for easy access and analysis (Figure 6).

```
import pandas as pd
data = pd.read_csv('C:/Users/User/Desktop/Applied Sciences/ApplSci_Data.csv', sep=';')
```

Figure 6. Data loading using Pandas library for easy data manipulation and analysis.

- Scikit-Learn (sklearn) for building, training and evaluating model, specifically train_test_split was used to split the dataset into a training set and a testing set (80% of the data is used for training the model and the remaining 20% is used for testing), which allows for an objective assessment of the model's performance on previously unseen data (Figure 7).

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure 7. Splitting Dataset into Training and Test Sets for Model Evaluation.

Additionally, RandomForestRegressor was used to create and train a Random Forest model with 100 estimators using the training dataset (Figure 8). This approach helps to capture complex patterns in the data and provides reliable predictions for the outcome variable.

```
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

Figure 8. Creating and training Random Forest Regressor to predict output variables.

Finally, Mean_squared_error (MSE) was also used to evaluate the model's quality. MSE measures the average squared difference between the actual values and the predicted values. A lower MSE indicates more accurate predictions by the model (Figure 9).

```
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(y_test, y_pred)
print(f"\nMean Squared Error: {mse:.4f}")
```

Figure 9. Calculating Mean Squared Error (MSE) to evaluate the Accuracy of the Model.

After running the Python code for training the Random Forest regression model, the obtained Mean Squared Error (MSE) was 0.000063. The Mean Squared Error (MSE) is a metric that estimates the average squared error between the actual values and the values predicted by the model. Specifically, the obtained value of 0.000063 indicates that the predicted values are very close to the actual values, with a minimal difference, which suggests a high accuracy of the model. This is an extremely low MSE value, implying that the model has learned the pattern in the data well and can accurately predict the output values, meaning the errors between actual and predicted values are insignificant. The predicted sigma level of 3.6947 indicates the expected process quality for the new combination of input parameters. Specifically, a sigma level of 3.6947 suggests that the process is of very good quality but still has room for improvement to reach a higher sigma level (for example, a 6σ level, which represents near-flawless production with a very low defect rate). Based on this predicted value, it can be concluded that, with the new combination of input parameters, the production process is expected to maintain stable quality, but it does not yet reach the highest Six Sigma standard. The Random Forest model uses the available data and, based on learned patterns, predicts the output sigma level (Y) for a new combination of input values (W_1, W_2, W_3). The results indicate relatively good stability, but with its limitations compared to higher sigma levels.

Additionally, a data visualization was created using available Python libraries to further analyse the model's accuracy and to display the correlation between the actual and predicted values (Figure A2). The following libraries were used for this purpose:

The following libraries were used:

- Pandas (import pandas as pd) for loading the CSV file and manipulating tabular data (Figure 10).

```
import pandas as pd
data = pd.read_csv('C:/Users/User/Desktop/Applied Sciences/ApplSci_Data.csv', sep=';')
```

Figure 10. Feature selection and preprocessing using Pandas library.

- NumPy (import numpy as np) for working with numerical data and performing mathematical operations (Figure 11).

```
import numpy as np

# Creating an array and calculating its mean in one step
array_mean = np.mean(np.array([10, 20, 30, 40, 50]))
```

Figure 11. NumPy for numerical data operations and supporting mathematical calculations.

- matplotlib (import matplotlib.pyplot as plt) for visualizing the data and displaying the prediction results relative to the actual values. matplotlib allows for plotting graphs

and visually representing the results, which helps in understanding the model's performance (Figure 12).

```
import matplotlib.pyplot as plt
plt.scatter(Y, predictions, color='blue', label='Predictions')
plt.plot([Y.min(), Y.max()], [Y.min(), Y.max()], color='red', linestyle='--', label='Ideal Prediction')
plt.title('Prediction vs. Actual Values')
plt.xlabel('Actual Values (Y)')
plt.ylabel('Predicted Values (Y)')
plt.legend()
plt.grid()
plt.show()
```

Figure 12. Using matplotlib to visualize predictions vs. actual values for model accuracy.

- scikit-learn–RandomForestRegressor (from sklearn.ensemble import RandomForestRegressor) for training the model based on the given input data (Figure 13).

```
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor()
model.fit(X, Y)
```

Figure 13. Random Forest Regressor to predict output values.

- scikit-learn–mean_squared_error (from sklearn.metrics import mean_squared_error) for evaluating the model's performance by measuring the Mean Squared Error (MSE) (Figure 14).

```
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(Y, predictions)
```

Figure 14. Calculating MSE to evaluate model performance.

The effectiveness of the developed model is illustrated by the correlation between the predicted and actual values of the output variable.

Figure 15 illustrates the comparison between actual values and predicted values using the Random Forest model, trained to predict the output variable Y based on the input variables W_1 , W_2 , and W_3 . On the x -axis, the actual values of the output variable are represented, while the y -axis shows the predicted values. The blue data points correspond to individual predictions made by the model, and the red dashed line represents the ideal scenario where the predicted values match the actual values exactly. Each data point reflects one observation, with its horizontal position representing the actual measured value of Y and its vertical position representing the predicted value of Y from the model. If the model were perfectly accurate, all data points would lie on the red dashed line, indicating a perfect correlation between the actual and predicted values. The proximity of most blue points to the red dashed line indicates that the model achieves high predictive accuracy with minimal deviation. This suggests that the Random Forest model performs well in predicting the output variable Y . Although there are some points that do not fall exactly on the ideal line, these deviations are relatively small, indicating low prediction errors.

In summary, the graph demonstrated that the Random Forest model provides reliable predictions of the output variable Y , with a relatively low Mean Squared Error (MSE). The model successfully approximates the relationship between the input variables and the output variable, making it suitable for practical applications where accurate predictions of sigma levels are required. This is critical for optimizing and controlling processes. Additionally, the model can be further refined by incorporating additional input parameters, potentially enhancing its ability to predict sigma levels under specific production process conditions.

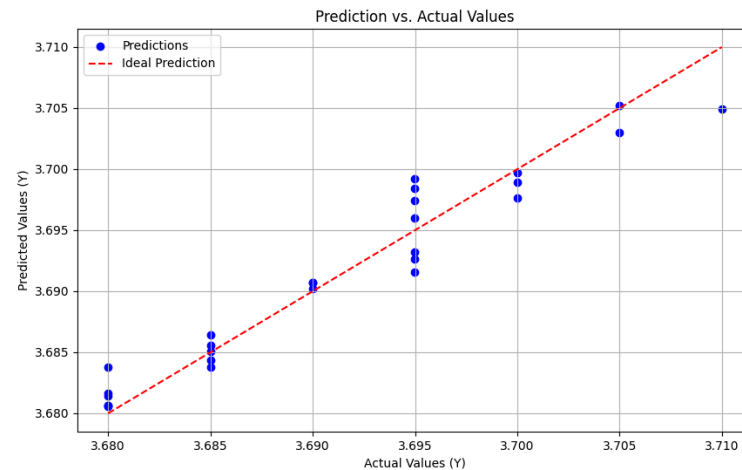


Figure 15. Correlation Between Predicted and Actual Values in the Optimization Model.

6. Conclusions

The conclusion of this study synthesizes the use of the Six Sigma methodology, artificial intelligence (AI), and the analysis of Critical Product Characteristics (CPCs) and Critical to Quality (CTQs) points in optimizing the offset printing process. Process analysis enabled the identification of key points of variability, the definition of the product's functional requirements, and the precise establishment of lower and upper tolerance limits for each characteristic. The application of the Random Forest model within the AI framework allowed for effective process optimization, accounting for multiple input variables such as ink temperature, viscosity, and alcohol content in the dampening solution, to achieve the highest possible stability and quality of the finished products.

The results demonstrate that the synergy of AI techniques and the Six Sigma approach can significantly reduce variability in the printing process while decreasing the number of non-conforming products. The developed optimization model showed a high ability to predict the sigma level and maintain the process within specified tolerance limits, as confirmed by the low Mean Squared Error (MSE) values. Furthermore, the model can serve as a basis for further research that could extend the application of this methodology to other sectors of the graphic industry or adapt parameters for additional production lines.

This research emphasizes the importance of advanced analytical methods for process improvement in the printing industry, with a particular focus on maintaining product quality and consistency under conditions with varying input parameters. Future research could focus on integrating deeper layers of machine intelligence into the control of production processes to achieve even greater levels of optimization and efficiency.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank the company Radin Print d.o.o. for their technical support.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Calculation of the Sigma Level of the 27 Partial Process.

1	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	2	2	17	0.016346	16,346	3.680
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			3	5	2	1	1	1	0	0	0	0				
2	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	1	2	17	0.016346	16,346	3.680
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			3	5	2	1	1	1	1	0	0	0				
3	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	4	2	17	0.016346	16,346	3.680
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			2	4	2	1	1	0	0	1	0	0				
4	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	3	3	16	0.015384	15,384	3.685
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			1	5	1	1	1	0	0	0	1	0				
5	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	2	1	16	0.015384	15384	3.685
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			2	4	2	1	1	1	1	0	0	0				
6	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	1	2	14	0.013461	13,461	3.695
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			3	4	1	1	0	0	1	0	1	0				
7	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	1	3	15	0.014423	14,423	3.690
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			2	3	2	1	1	1	0	0	1	0				

Table A1. *Cont.*[illegible]

Table A1. Cont.

24	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	2	3	17	0.016346	16,346	3.680
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			2	3	2	1	1	1	1	1	0	0				
25	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	1	2	17	0.016346	16,346	3.680
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			2	5	1	1	1	1	1	1	1	0				
26	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	2	1	16	0.015384	15,384	3.685
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			1	4	2	1	1	1	1	1	1	0				
27	G	z	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6	ε_7	ε_8	ε_9	ε_{10}	$\Sigma\varepsilon$	ε/u	N (PPM)	$k\sigma$
52	20		0	0	0	0	0	0	0	0	1	1	17	0.016346	16,346	3.680
			ε_{11}	ε_{12}	ε_{13}	ε_{14}	ε_{15}	ε_{16}	ε_{17}	ε_{18}	ε_{19}	ε_{20}				
			3	5	1	2	1	1	1	0	1	0				

Appendix B

```

import pandas as pd
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error

# Loading data from CSV file
data_path = 'C:/Users/User/Desktop/Applied Sciences/ApplSci_Data.csv'
data = pd.read_csv(data_path, sep=';', decimal='.')

# Printing loaded data
print("Loaded data:")
print(data)

# Defining input (X) and output (y) variables
X = data[['W1', 'W2', 'W3']]
y = data['Y']

# Splitting data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Creating and training the Random Forest model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Making predictions on the test set
y_pred = model.predict(X_test)

# Calculating the mean squared error (MSE) rounded to 6 decimals
mse = mean_squared_error(y_test, y_pred)
print(f"\nMean Squared Error: {mse:.6f}")

# Prediction for a new data combination, rounded to 4 decimals
new_combination = pd.DataFrame({'W1': [27.00], 'W2': [10.20], 'W3': [5.60]})
prediction = model.predict(new_combination)

print(f"\nPredicted sigma level for the new combination: {prediction[0]:.4f}")

# Saving predictions to a CSV file for further analysis
results = pd.DataFrame({'Actual Values': y_test, 'Predicted Values': y_pred})
results.to_csv('C:/Users/User/Desktop/Applied Sciences/predictions_test.csv', index=False)

print("Predictions saved to CSV file.")

```

Figure A1. Python Code for Modelling and Prediction Using Random Forest Regression.

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error

# Loading data from CSV file
data_path = 'C:/Users/User/Desktop/Applied Sciences/ApplSci_Data.csv'
data = pd.read_csv(data_path, sep=';', decimal='.')

# Printing loaded data
print("Loaded data:")
print(data)

# Defining input (X) and output (Y) variables
X = data[['W1', 'W2', 'W3']]
Y = data['Y']

# Creating and training the Random Forest model
model = RandomForestRegressor()
model.fit(X, Y)

# Making predictions based on input data
predictions = model.predict(X)

# Calculating the mean squared error (MSE) rounded to 6 decimals
mse = mean_squared_error(Y, predictions)
print(f"\nMean Squared Error: {mse:.6f}")

# Visualization of predictions vs. actual values
plt.figure(figsize=(10, 6))
plt.scatter(Y, predictions, color='blue', label='Predictions')
plt.plot([Y.min(), Y.max()], [Y.min(), Y.max()], color='red', linestyle='--', label='Ideal Prediction')
plt.title('Prediction vs. Actual Values')
plt.xlabel('Actual Values (Y)')
plt.ylabel('Predicted Values (Y)')
plt.legend()
plt.grid()
plt.show()

```

Figure A2. Python Code for Prediction and Data Analysis Using Random Forest Regression.

References

- Joglekar, A.M. *Statistical Methods for Six Sigma in R&D and Manufacturing*, 1st ed.; John Wiley and Sons: Hoboken, NJ, USA, 2003; pp. 13–140, 147–148.
- Goetsch, D.L.; Davis, S.B. *Quality Management: Introduction to Total Quality Management for Production, Processing and Services*, 5th ed.; Pearson Prentice Hall: New York, NY, USA, 2006; pp. 589–596.
- Allen, T.T. *Introduction to Engineering Statistic and Lean Sigma*, 3rd ed.; Springer: London, UK, 2019; pp. 8–9, 37–54.
- Kahraman, C.; Yanik, S. Process Improvement Using Intelligent Six Sigma. In *Intelligent Decision Making in Quality Management: Theory and Applications*, 1st ed.; Kahraman, C., Yanik, S., Eds.; Springer: Cham, Switzerland, 2016; Volume 1, pp. 363–387.
- Abdolrasol, M.G.M.; Hussain, S.M.S.; Ustun, T.S.; Sarker, M.R.; Hannan, M.A.; Mohamed, R.; Ali, J.A.; Mekhilef, S.; Milad, A. Artificial Neural Networks Based Optimization Techniques: A Review. *Electronics* **2021**, *10*, 2689. [\[CrossRef\]](#)
- Skalli, D.; Charkaoui, A.; Cherrafi, A. Progress and Trends in Industry 4.0 and Lean Six Sigma Integration. *Lect. Notes Mech. Eng.* **2023**, *1*, 87–95. [\[CrossRef\]](#)
- Najafi, B.; Najafi, A.; Farahmandian, A. The Impact of Artificial Intelligence and Blockchain on Six Sigma: A Systematic Literature Review of the Evidence and Implications. *IEEE Trans. Eng. Manag.* **2024**, *1*, 10261–10294. [\[CrossRef\]](#)
- Tissir, S.; Cherrafi, A.; Chiarini, A.; Elfezazi, S.; Shokri, A.; Antony, J. Lean Six Sigma and Industry 4.0 implementation framework for operational excellence: A case study. *Total Qual. Manag. Bus. Excell.* **2024**, *35*, 1015–1053. [\[CrossRef\]](#)
- Jamwal, A.; Agrawal, R.; Sharma, M. Deep learning for manufacturing sustainability: Models, applications in Industry 4.0 and implications. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100107. [\[CrossRef\]](#)
- Singh, A.B.; Gaurav, S.P.; Dangayach, G.S.; Meena, M.L. Present, Past and Future of Lean Six Sigma Applications: From Evolution to the Era of Artificial Intelligence. *Recent Pat. Eng.* **2024**, *18*, 2–17. [\[CrossRef\]](#)
- Sordan, J.E.; Andersson, R.; Antony, R.; Pimenta, M.L.; Oprime, P.C. How Industry 4.0, artificial intelligence and augmented reality can boost Digital Lean Six Sigma. *Total Qual. Manag. Bus. Excell.* **2024**, *35*, 1542–1566. [\[CrossRef\]](#)
- Bertocci, F.; Grandoni, A.; Fidanza, M.; Berni, R. A Guideline for Implementing a Robust Optimization of a Complex Multi-Stage Manufacturing Process. *Appl. Sci.* **2021**, *11*, 1418. [\[CrossRef\]](#)
- Perera, A.; Jayamaha, N.P.; Grigg, N.P.; Tunnicliffe, M.; Singh, A. The Application of Machine Learning to Consolidate Critical Success Factors of Lean Six Sigma. *IEEE Access* **2021**, *9*, 112411–112424. [\[CrossRef\]](#)
- Vargas, M.; Mosquera, R.; Fuertes, G.; Alfaro, M.; Perez Vergara, I.G. Process Optimization in a Condiment SME through Improved Lean Six Sigma with a Surface Tension Neural Network. *Processes* **2024**, *12*, 2001. [\[CrossRef\]](#)

15. Nader, J. Advanced Analytics Tools for Process Improvement: A Case Study in a Brewery. In Proceedings of the International Conference on Interdisciplinary Research in Technology and Management, IRTM 2022, Kolkata, India, 24–26 February 2022; Chakrabarti, S., Sakib, A.A., Singh, R., Banerji, P.K., Poddar, S., Bhattacharya, A., Gangopadhyaya, M., Chakraborty, S., Eds.; CRC Press: Abingdon, UK, 2024; p. 180156.
16. Sharma, A.; Singh, B.J. Evolution of modified LSS 4.0 model for sustainable Indian textile industry: A narrative review. *Int. J. Interact. Des. Manufacturing IJIDEM* **2023**, *1*, 1–27. [\[CrossRef\]](#)
17. Hegazi, E.H.; Yang, L.; Huang, J. A Convolutional Neural Network Algorithm for Soil Moisture Prediction from Sentinel-1 SAR Images. *Remote Sens.* **2021**, *13*, 4964. [\[CrossRef\]](#)
18. Gajewski, T.; Grabski, J.K.; Cornaggia, A.; Garbowski, T. On the use of artificial intelligence in predicting the compressive strength of various cardboard packaging. *Packag. Technol. Sci.* **2024**, *37*, 97–105. [\[CrossRef\]](#)
19. Garbowski, T. Revolutionizing Corrugated Board Production and Optimization with Artificial Intelligence. *Bioresources* **2024**, *19*, 2003–2006. [\[CrossRef\]](#)
20. Yang, H.; Prahalad, R.; Simpson, T.; Lu, Y.; Witherell, P.; Nassar, A.R.; Reutzel, E.; Kumar, S. Six-Sigma Quality Management of Additive Manufacturing. *Proc. IEEE* **2021**, *109*, 347–376. [\[CrossRef\]](#)
21. Rojek, I.; Dostatni, E.; Kopowski, J.; Macko, M.; Mikołajewski, D. AI-Based Support System for Monitoring the Quality of a Product within Industry 4.0 Paradigm. *Sensors* **2022**, *22*, 8107. [\[CrossRef\]](#)
22. Zhang, X.; Chu, D.; Zhao, X.; Gao, C.; Lu, L.; He, Y.; Bai, W. Machine learning-driven 3D printing: A review. *Appl. Mater. Today* **2024**, *39*, 102360. [\[CrossRef\]](#)
23. Hassan, M.; Misra, M.; Taylor, G.W.; Mohanty, A.K. A Review of AI for optimization of 3D Printing of Sustainable Polymers and Composites. *Compos. Part C Open Access* **2024**, *1*, 100513. [\[CrossRef\]](#)
24. Ulkir, O.; Bayraklılar, M.S.; Kuncan, M. Raster Angle Prediction of Additive Manufacturing Process Using Machine Learning Algorithm. *Appl. Sci.* **2024**, *14*, 2046. [\[CrossRef\]](#)
25. Živković, M.; Žujović, M.; Milošević, J. Architectural 3D-Printed Structures Created Using Artificial Intelligence: A Review of Techniques and Applications. *Appl. Sci.* **2023**, *13*, 10671. [\[CrossRef\]](#)
26. Zgodavova, K.; Bober, P.; Majstorovic, V.; Monkova, K.; Santos, G.; Juhaszova, D. Innovative Methods for Small Mixed Batches Production System Improvement: The Case of a Bakery Machine Manufacturer. *Sustainability* **2020**, *12*, 6266. [\[CrossRef\]](#)
27. Henry, W.; Mensaco, M.; Takada, H. *New Product Development and Testing*, 1st ed.; Lexington Books, Macmillan Inc.: New York, NY, USA, 1989; pp. 71–98.
28. Juran, J.M.; Godfrey, A.B. *Juran's Quality Handbook*, 5th ed.; McGraw-Hill: New York, NY, USA, 1998; pp. 61–64.
29. Riedl, R.; Neumann, D.; Teubner, J. *Tehnologie des Offsetdruck*, 1st ed.; VEB Fachbuchverlag: Leipzig, Germany, 1989; pp. 30–32, 97–100, 230–231.
30. Walenski, W. *Der Offsetdruck: Eine Einführung in Theorie und Praxis*, 1st ed.; DuMont: Köln, Germany, 1991; pp. 112–117.
31. Walenski, W. *Polygraph-Handbuch: Offsetdruck*, 1st ed.; Polygraph Verlag: Frankfurt, Germany, 1991; pp. 119–136, 150–155, 176–179.
32. Helbig, T.; Bosse, R. *Druckqualität: Grundlagen der Qualitätsbewertung im Offsetdruck*, 1st ed.; Polygraph Verlag: Frankfurt, Germany, 1993; pp. 78–122.
33. Walenski, W. *Der Rollenoffsetdruck*, 1st ed.; Fachschriften Verlag: Fellbach, Germany, 1995; pp. 76–92, 183–201.
34. Teschner, H. *Offsetdruck Technik: Technologien und Werkstoffe in der Druckindustrie*, 1st ed.; Fachschriften Verlag: Fellbach, Germany, 1997; pp. 10/142–10/146, 12/17–12/23, 14/8–14/13, 11/20–11/27, 16/1–16/8.
35. MacPhee, J. *Fundamentals of Lithographic Printing Vol, 1: Mechanics of Printing*, 1st ed.; PIA GATF Press: Sewickley, PA, USA, 1998; pp. 9–48, 147–162, 227–301.
36. Kipphan, H. *Handbook of Print Media: Technologies and Production Methods*, 1st ed.; Springer: Berlin, Germany, 2001; pp. 206–226, 284–258, 300–319.
37. DeJidas, L.P.; Destree, T.M. *Sheetfed Offset Press Operating*, 3rd ed.; PIA GATF Press: Sewickley, PA, USA, 2005; pp. 29–33, 84–112, 371–392, 448–470.
38. Breede, H.M. *Handbook of Graphic Arts Equations*, 2nd ed.; PIA/GATF Press: Pittsburgh, PA, USA, 2006; pp. 77–83, 101–105.
39. Bann, D. *The All New Print Production Handbook*, 1st ed.; Watson-Guptil Publications: New York, NY, USA, 2007; pp. 62–113.
40. Johansson, K.; Lundberg, P.; Ryberg, R. *A Guide to Graphic Print Production*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2011; pp. 303–317.
41. Zjakić, I. *Upravljanje Kvalitetom Ofsetnog Tiska*, 1st ed.; Hrvatska sveučilišna naklada: Zagreb, Croatia, 2007; pp. 39–53, 71–88, 92–117.
42. Englund, C.; Verikas, A. Ink flow control by multiple models in an offset lithographic printing process. *Comput. Ind. Eng.* **2008**, *55*, 592–605. [\[CrossRef\]](#)

43. Leach, R.H.; Pierce, R.J. *The Printing Ink Manual*, 5th ed.; Springer: Dordrecht, The Netherlands, 2008; pp. 74–80.
44. Liu, F.; Shen, W. Forced wetting and dewetting of liquids on solid surfaces and their role in offset printing. *Colloids Surf. A Physicochem. Eng. Asp.* **2008**, *316*, 62–69. [\[CrossRef\]](#)
45. Hird, K.F.; Finley, C.E. *Offset Lithographic Technology*, 4th ed.; Tinley Park: Goodheart-Willcox Co.: Tinley Park, IL, USA, 2009; pp. 224–239.
46. Sharma, A.; Starr, B. Evaluation of Ink Optimization Technology in Offset Color Printing. *J. Imaging Sci. Technol.* **2010**, *54*, art00013. [\[CrossRef\]](#)
47. Amon-Tran, I.; Anayath, R.; Pai, A.S.; Kamath, N.; Shenoy, S.; Harikrishnan, A. An approach to minimize carbon footprint for an environmental friendly printing by optimizing an offset machine in a printing facility. *Procedia Soc. Behav. Sci.* **2012**, *37*, 514–527. [\[CrossRef\]](#)
48. Varepo, L.G.; Brazhnikov, A.Y.; Nagornova, I.V.; Novoselskaya, O.A. Methods of Measurement the Quality Metrics in a Printing System. *Metrol. Standardization Qual. Theory Pract.* **2017**, *998*, 012040. [\[CrossRef\]](#)
49. Varepo, L.G.; Brazhnikov, A.Y.; Volinsky, A.A.; Nagornova, I.V.; Kondratov, A.P. Control of the offset printing image quality indices. *J. Phys. Conf. Ser.* **2017**, *858*, 012038. [\[CrossRef\]](#)
50. Zolotukhina, K.; Khadzhyanova, S.; Velychko, O.; Kushlyk, B.; Kushlyk-Dyvulska, O. Researching the interaction of different printed materials types with liquids. *East. Eur. J. Enterp. Technol.* **2019**, *3*, 26–34. [\[CrossRef\]](#)
51. Aydemir, C.; Yenidogan, S. The influence of surface tension on wetting in dampening solution and sustainability of printing: A review. *J. Graph. Eng. Des.* **2019**, *10*, 5–10. [\[CrossRef\]](#)
52. Zorenko, Y.; Skyba, V. Stabilization of imprint's optical parameters in offset printing. In Proceedings of the 15th International Conference on Correlation Optics, Chernivtsi, Ukraine, 13–16 September 2021; Angelsky, O.V., Ed.; SPIE: Chernivtsi, Ukraine, 2021; p. 121260T.
53. Sajek, D.; Havenko, S. Quality assessment of offset thermosensitive printing plates. *J. Graph. Eng. Des.* **2022**, *13*, 29–36. [\[CrossRef\]](#)
54. Abdunazarov, M.; Babahanova, H. *Offset Printing Process Parameters for Lightweight Paper*, 1st ed.; Our Knowledge Publishing: Port Louis, Mauritius, 2024; pp. 19–36.
55. Mahajan, M.; Arulmozhi, A. Multivariate Model for Characterising Physical Dot Gain, Chroma and Lightness in Offset Lithography. *Prog. Color Color. Coat.* **2024**, *17*, 365–380. [\[CrossRef\]](#)
56. Kumar, D. *Six Sigma Best Practices: A Guide to Business Process Excellence for Diverse Industries*, 1st ed.; J. Ross Publishing: Fort Lauderdale, FL, USA, 2006; pp. 344–346, 380–411.
57. Moore, D.S.; McCabe, G.P.; Craig, B.A. *Introduction to the Practice of Statistics*, 9th ed.; Macmillan Learning: New York, NY, USA, 2017; pp. 22–37.
58. Snee, R.D. Increasing newspaper accuracy using Six Sigma methodology. In *World Class Applications of Six Sigma: Real World Examples of Success*, 1st ed.; Antony, J., Banuelas, R., Kumar, A., Eds.; Routledge: London, UK, 2006; Volume 1, pp. 211–223.
59. Wu, C.W.; Pearn, W.L.; Kotz, S. An overview of theory and practice on process capability indices for quality assurance. *Int. J. Prod. Econ.* **2009**, *117*, 338–359. [\[CrossRef\]](#)
60. Juran, J.M.; Gryna, F.M. *Quality Planning and Analysis*, 1st ed.; McGraw-Hill: New York, NY, USA, 1993; pp. 377–418.
61. Oakland, J.S. *Statistical Process Control*, 6th ed.; Elsevier: Oxford, UK, 2008; pp. 257–273.
62. Kane, V.E. Process capability indices. *J. Qual. Technol.* **1986**, *18*, 41–52. [\[CrossRef\]](#)
63. Kotz, S.; Lovelace, C.R. *Process Capability Indices in Theory and Practice*, 1st ed.; Hodder Education Publishers: London, UK, 1998; pp. 19–27.
64. Newbold, P.; Carlson, W.L.; Thorne, B. *Statistics for Business and Economics*, 9th ed.; Pearson Prentice Hall: Hoboken, NJ, USA, 2020; pp. 242–251.
65. Kleijnen, J.P.C. *Design and Analysis of Simulation Experiments*, 1st ed.; Springer: New York, NY, USA, 2008; pp. 46–65.
66. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E. *Multivariate Data Analysis: A Global Perspective*, 7th ed.; Pearson: Boston, MA, USA, 2010; pp. 76–77.
67. Wengang, Z.; Chongzhi, W.; Haiyi, Z.; Yongqin, L.; Lin, W. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geosci. Front.* **2021**, *12*, 469–477. [\[CrossRef\]](#)
68. Ahmed, I.; Raihan, A.S. Machine learning techniques for sustainable industrial process control. In *Computational Intelligence Techniques for Sustainable Supply Chain Management*, 1st ed.; Sanjoy, K.P., Sandeep, K., Eds.; Elsevier: Amsterdam, The Netherlands, 2024; Volume 1, pp. 141–176.
69. Kabir, M.M.; Roy, S.K.; Alam, F.; Nam, S.Y.; Im, K.S.; Tijjing, L.S.; Ho, K. Machine learning-based prediction and optimization of green hydrogen production technologies from water industries for a circular economy. *Desalination* **2023**, *567*, 116992. [\[CrossRef\]](#)
70. Sun, L.; Ji, Y.; Zhu, X.; Peng, T. Process knowledge-based random forest regression for model predictive control on a nonlinear production process with multiple working conditions. *Adv. Eng. Inform.* **2022**, *52*, 101561. [\[CrossRef\]](#)
71. Chen, Y.; Li, F.; Zhou, S.; Zhang, X.; Zhang, S.; Zhang, Q.; Su, Y. Bayesian optimization based random forest and extreme gradient boosting for the pavement density prediction in GPR detection. *Constr. Build. Mater.* **2023**, *387*, 131564. [\[CrossRef\]](#)

72. Ertel, W.; Black, N.T. *Introduction to Artificial Intelligence (Undergraduate Topics in Computer Science)*, 2nd ed.; Springer: London, UK, 2017; pp. 198–213.
73. Sheppard, C. *Tree-Based Machine Learning Algorithms: Decision Trees, Random Forest and Boosting*, 2nd ed.; CreateSpace Independent Publishing Platform: Austin, TX, USA, 2019; pp. 75–79.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.