Towards Automated Quality Control in Industrial Systems: Developing Markov Decision Process Model for Optimized Decision-Making

Katerina Mitkovska-Trendova[1], Robert Minovski[2], Verica Bakeva[3], Simeon Trendov[4] and Dimitar Bogatinov[5]
[1]*Faculty of Engineering Technology and Science, Higher Colleges of Technology, Al Mizn - Baniyas North, 20020 Abu Dhabi, UAE*
[2]*Faculty of Mechanical Engineering, Ss. Cyril and Methodius University, Rugjer Boshkovikj 18, 1000 Skopje, N. Macedonia*
[3]*Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, Rudzer Boshkovikj 16, 1000 Skopje, N. Macedonia*
[4]*Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany*
[5]*Military Academy Gen. Mihailo Apostolski, st. Vasko Karangelevski n.n., 1000 Skopje, N. Macedonia*
ktrendova@hct.ac.ae, robert.minovski@mf.edu.mk, verica.bakeva@finki.ukim.mk, simeon.trendov@hs-anhalt.de, dimitar.bogatinov@morm.gov.mk

Keywords: Markov Decision Processes, Performance Measurement Systems, Quality Control, Percent of Scrap, Costs Minimization, Automatization.

Abstract: In the context of rapidly evolving industrial environments, optimizing decision-making for quality control is crucial. This paper develops a Markov Decision Process (MDP) model aimed at enhancing automated quality control and reducing scrap in manufacturing systems, addressing challenges posed by complex and uncertain decision scenarios. The study focuses on improving the sub-key element of quality-accuracy within a Performance Measurement System (PMS) framework, specifically targeting scrap minimization and cost reduction. The research employs a mathematical model that integrates vector random processes, each representing critical factors such as machine condition, operator behaviour, tools, and materials. These factors are modeled as individual one-dimensional MDPs, which are combined to create a multi-dimensional MDP capable of monitoring and offering optimal policy for minimizing scrap rates and costs. The research methodology leverages advanced data analytics, statistical modeling, and real-time monitoring to accurately estimate transition probabilities and optimize policies. Different MDP models and methods are explored to enhance adaptability and iterative learning, allowing for optimal policy refinement over time. The proposed model is validated through its application to a real-world printing enterprise identified critical element, demonstrating a reduction in scrap and costs. This improvement underscores the model’s effectiveness in practical settings, offering structured, subsystem-specific interventions that enhance manufacturing quality control. The results hold both theoretical and practical significance. Theoretically, the study contributes to the body of knowledge on MDP modeling for industrial quality control, providing a scalable approach that addresses complex interdependencies and decision-making under uncertainty. Practically, the model offers a robust tool for optimizing manufacturing processes, supported by modern IT systems, integration of advanced technologies, predictive maintenance, and data-driven decision-making. This integrated approach enables manufacturers to proactively identify and mitigate quality issues, enhancing operational efficiency, reducing waste, and driving continuous improvement in industrial systems.

1 INTRODUCTION

Enterprises today are compelled to evolve due to a convergence of technological advancements, globalization, shifting market demands, economic fluctuations, regulatory pressures, and internal transformations such as workforce and cultural changes. These factors collectively influence enterprise performance and require agility and innovation to maintain competitiveness in a dynamic environment. Various innovative approaches have been introduced to address these changes, particularly in improving enterprise functionality and adaptability.

The belief that "You cannot improve what you cannot measure" reflects the prevailing mindset among today's business leaders. Performance Measurement Systems (PMS) emerged as a central focus in the early 1990s, becoming a key concept for guiding enterprises in a dynamic environment towards improvement and competitive advantage. Over time, numerous PMS concepts have been developed, all centered around a common goal: to segment the enterprise's operations into distinct areas and monitor performance by evaluating these areas.

Initially in our study, we aimed to improve a PMS by introducing and applying mathematical models to support decision-making in enterprise restructuring while coping with rapid changes. The PMS under consideration is referred to as COMPASS (COmpany's Management Purpose ASSistance) [1]. COMPASS focuses on key areas of success, including time, quality, costs, flexibility, and productivity. Due to their broad nature, these areas are further refined into sub-key elements of success (subKEs), resulting in a total of 18 subKEs. The extensive range of elements within PMSs poses challenges when it comes to generating actionable improvements, known as Success Factors. Each subKE may be influenced by multiple factors, and each factor can be enhanced through various actions. Given the complexity and interdependencies, action generation is typically approached heuristically, but modern systems increasingly leverage data analytics, real-time monitoring, and automation to support the decision making.

Therefore, in this study, when we refer to new approaches, we primarily focus on mathematical modeling. However, it is essential to consider related IT advancements supported by mathematical modeling which are further observed. Regarding mathematical modeling specifically, a diverse range of models and tools have been implemented in recent years to enhance organizational performance.

Markov Decision Processes (MDPs) are utilized to model specific aspects of enterprise operations due to their proven capabilities and advantages. MDPs are supported by a well-developed theory and have become a mature modeling tool. Their successful application is largely due to the availability of efficient algorithms for finding optimal solutions. Furthermore, MDPs provide a flexible framework for solving optimization problems across a wide range of fields and are particularly valuable in sequential planning applications where accounting for process uncertainty is critical [2].

Applying MDP to solve a restructuring, decision-making, or planning problem within an enterprise, while managing the dimensionality of the problem, yields an optimal policy based on a specified optimality criterion. If this criterion aligns with a key performance metric for a specific subKE (e.g., percent of scrap for the subKE quality-accuracy), it implies that the subKE is being optimized. This approach enables the generation of success factors or improvement actions for various elements of the PMS using quantitative methods supported by software, rather than relying on heuristic approaches.

To accurately assess the performance of subKEs, a specific measure must be assigned to each subKE. The primary objective of monitoring and recording these measures is to identify issues and generate actions for continuous improvement. Additionally, it is essential to pinpoint the influential factors contributing to a given situation.

Taking all of this into consideration, the primary research challenge of this study was to develop a mathematical model based on MDP to enhance one aspect of the PMS, quality-accuracy, by managing its key measure: the percent of scrap.

Further we look at the MDPs extensive utilization in real-world scenarios that require decision-making under uncertainty, highlighting the pivotal role of modern IT systems. Advanced IT tools facilitate data collection, processing, and real-time analysis, enabling more precise modeling and optimization of complex decision processes, thereby enhancing the practical implementation of MDPs. They are particularly valuable in enhancing the automation of industrial processes by improving efficiency, adaptability, and reliability. MDPs can optimize various aspects, including process optimization and control, quality control and inspection, predictive maintenance, inventory, supply chain management, logistics, scheduling and resource allocation, adaptive process control in dynamic environments, etc. By leveraging MDPs, automated industrial systems can effectively manage complex, interdependent decisions, striking a balance between short-term costs and long-term efficiency and sustainability. This results in highly efficient, resilient, and sustainable enterprise environments.

2 LITERATURE REVIEW

The evolution of PMS has been driven by the need to better understand, measure, and improve organizational performance beyond traditional financial metrics. Early systems primarily focused on financial indicators, which often failed to capture the complexity and multidimensional nature of modern business operations [3]. This limitation paved the way for the development of more holistic frameworks, such as the Balanced Scorecard by Kaplan and Norton [4], which integrated financial measures with non-financial perspectives, including customer satisfaction, internal processes, and organizational learning and growth. Concurrent with the development of the Balanced Scorecard, Total Quality Management (TQM) approaches, championed by Deming [5] and Juran [6], emphasized continuous improvement, customer-centric performance metrics, and process optimization. These methodologies highlighted the need for aligning strategic goals with operational measures, laying a foundation for comprehensive PMS approaches. As businesses faced rapid globalization, technological advances, and increased complexity, PMS evolved further to incorporate dynamic and flexible measurement tools such as the implementing MDPs. Modern systems increasingly leverage data analytics, real-time monitoring, and automation to provide more granular insights and facilitate agile decision-making [7]. With advances in data collection, artificial intelligence, and business intelligence systems, PMS has become more sophisticated, enabling organizations to optimize their performance and respond effectively to market changes [8]. Today, PMS is recognized as a critical tool for guiding strategic initiatives, measuring success, and fostering continuous improvement and competitiveness.

On the other hand, advancements in MDPs have allowed their adoption across a wide range of applications, greatly supported by the latest developments in Information Technology (IT). MDPs are a mathematical framework for modeling decision-making in environments where outcomes are partly random and partly under the control of a decision-maker. The foundation of MDPs dates back to the work of Andrey Markov in the early 20th century, who introduced the concept of Markov chains to model stochastic processes [9]. MDPs extend these concepts by incorporating actions and rewards, allowing for optimization of long-term outcomes through sequential decision-making. The formalization of MDPs for decision problems was developed in the 1950s by Richard Bellman, who introduced the principle of dynamic programming as a method to solve MDPs and coined the term "Bellman equation" to describe the recursive decomposition of value functions [10]. This framework has since become foundational in operations research, artificial intelligence, and control theory, providing a structured way to handle uncertainty in decision-making. Throughout the late 20th century, MDPs were further refined with the development of exact algorithms such as value iteration and policy iteration [2]. However, many real-world applications present challenges such as large or continuous state spaces, which led to advancements in approximation methods and reinforcement learning algorithms, including Q-learning and policy gradient methods [11]. For this study it was important to examine the development of application of MDPs in quality control. For instance, Markov Chains to model and simulate transitions between different product quality states in manufacturing processes were used in [12]. Their approach identifies influential factors and proposes measures for continuous quality improvement, highlighting the value of stochastic modeling in industrial quality management. Today, MDPs are widely used in fields such as robotics, automated control systems, finance, healthcare decision-making, and artificial intelligence for complex planning and optimization problems under uncertainty [13]. A methodology for determining the transition probabilities of MDP for quality accuracy improvement inside a PMS framework was proposed in [14]. Modern applications continue to push the boundaries of scalability and efficiency in different MDP models and solutions through techniques like Monte Carlo methods, approximate dynamic programming, deep reinforcement learning, Q-learning, etc. Some of the reviewed references like [15], [16], [17], [18], [19], [20] were found insightful for this work.

3 METHODOLOGIES

As mentioned before, the PMS methodology COMPASS served as the foundational framework within which the research was conducted. The approach leverages mathematical modeling to address real-world challenges, utilizing operations research models and methods, with a particular focus on MDP models and the policy iteration optimization technique. Extensive literature review was conducted in order to select the right model for the modeled quality control problem. During the modeling and the application of the mathematical model to a specific enterprise problem, various management techniques were employed to analyze the organization, complemented by statistical methods for data collection and processing. To identify and address the causes of scrap, quality management tools such as data collection lists, process flowcharts, Ishikawa diagrams, histograms, scatter plots, and control charts were utilized. This research developed methodologies to calculate transition probability matrices and revenue (cost) matrices for each factor individually, as well as collectively, to monitor the percentage and cost of scrap for each influential factor and overall [14]. Custom software is developed to perform the calculations with the provided data.

All stages of the research were synthesized into a model for generating optimal decision policies and success factors, specifically targeting the management of the critical quality-accuracy element within the selected PMS. This model is designed to facilitate practical implementation within a real enterprise, enhancing the functionality and effectiveness of its PMS.

The methodology introduced in [14] has influenced subsequent research by providing a robust framework for applying MDP models in quality management, facilitating better understanding and control of key performance factors through precise probabilistic modeling. It has contributed the way for more accurate and data-driven decision-making strategies across various applications in quality and operational management [15].

4 MATHEMATICAL MODELING OF A REAL SYSTEM USING FOUR-DIMENSIONAL MDP

The model for managing quality-accuracy, specifically focusing on scrap and scrap cost management, serves as a decision-making support tool for stochastic, multi-stage planning processes. This model represents a system comprising a single job station, consisting of one machine and one operator. Due to prolonged usage, both the machine and its tools experience deterioration, resulting in decreased quality and increased scrap production. However, additional factors can contribute to scrap generation. SubKE quality-accuracy can be influenced by various elements, such as the machine, operator, tools, materials, environment, and methods. To strike a balance between model complexity and realism, this study focuses on the most significant factors—machine, operator, tools, and materials. However, with advancements in modern IT, the model can be further expanded, addressing previous limitations and enhancing its capabilities. The model aims to provide a detailed breakdown of scrap production by cause, as well as overall scrap levels. To achieve this, stochastic processes represented by random variables are defined to describe the conditions of the machine, operator, tools, and materials, specifically in terms of their contributions to scrap production. Subsequently, a vector random process is constructed, comprising these four individual stochastic processes to capture both the individual and cumulative scrap production for the system. At the conclusion of each production cycle, the conditions of the four factors, in terms of the percentage of scrap they generate, are recorded and classified into a finite number of states, which represent the values of the random processes. Historical data was used to determine the transition probabilities for each possible state change between production cycles, for each influential factor. One of the major challenges today lies in the need for extensive data to implement these models, which is difficult to obtain using traditional methods. However, advancements in modern IT provide powerful tools for capturing, recording, processing, and utilizing data effectively in determining the decision-making policy. Since the transition probabilities are independent of the states from previous cycles, these stochastic processes can be modeled as discrete-time, finite, homogeneous Markov chains. Finite action spaces are defined for each Markov chain, representing available decision alternatives. The revenue structure associated with each process yields matrices corresponding to all possible transitions, with the revenue function reflecting gains or losses through scrap percentages and costs for each transition step. As a result, four one-dimensional MDPs are derived, which are then combined into a four-dimensional MDP represented as a vector random process, with a specifically designed action space and revenue structure based on the one-dimensional MDPs. The first one-dimensional MDP is described by the random process "the condition of the machine after every run", as one of the most important influence factors or cause for scrap identified in quality-accuracy management, and for quality measure the percent of scrap is chosen. For the random variable X_n^1 which is the condition of the machine in a discrete moment n, it is assumed that the stochastic process $\{X_n^1 | n \in \mathbb{N}\}$ is homogeneous Markov chain. At any point of time n, the condition of the machine can be classified in one of several possible states and the random variable X_n^1 in a given moment n, takes values from the defined state space for the condition of the machine. It is assumed that in every discrete moment of time the random variable X_n^1, takes values from the same state space, and further for simpler notation, this random process will be denoted only by X_1 for every transition moment and this is valid for the other stochastic processes. It is assumed that all random variables describing the given Markov chains are mutually independent. The second random process X_2, i.e. the second Markov chain, which is "the condition of the operator after every run of the work place", expressed by the caused percent of scrap from the operator. The third one-dimensional MDP is the random process X_3, which is "the condition of the machine tools after every run of the work place", expressed by the caused percent of scrap from the tools. The fourth one-dimensional MDP is the random process X_4, which is "the condition of the materials after each run of the work place", expressed by the caused percent of scrap from the materials. Let the random process X_l, $l \in \{1, 2, 3, 4\}$, takes $|R_{X_l}| = n_l$ values. For example, let the sets of values for these stochastic processes are $R_{X_l} = \{x_1^l, x_2^l, x_3^l\}$, $n_l = 3$. To simplify, the same notations for the states of the one-dimensional MDPs and the values of associated random variables are used. A_l, denote the sets of primary actions (decisions) for the one-dimensional MDPs and $|A_l| = m_l$ are their numbers, for $l \in \{1, 2, 3, 4\}$. For example, let $A_l = \{a_1^l, a_2^l, a_3^l\}$, $|A_l| = m_l = 3$.

One significant limitation of the MDP model is the challenge in accurately determining transition probabilities. Typically, historical data are used to estimate these probabilities, but such data are often unavailable, hard to collect, or outdated. Additionally, transition probabilities can be influenced by a variety of changing factors in the environment, making them susceptible to fluctuation over time. As a result, the values of these probabilities may shift, leading to potential inaccuracies in the model and impacting the reliability of the decision-making process. In an automated environment, systematically tracking causes, and recording and categorizing failure data can help address this challenge. By continuously collecting accurate data on failure patterns and updating the MDP model accordingly, the system can adapt more effectively to changes in the operational environment, improving the accuracy of transition probabilities over time. This model assumes that transition probabilities do not change over time, i.e. Markov chains are homogenous. The transition probabilities are denoted with $p_{ij}^k(l)$, and they are the conditional probabilities that the random variable X_l takes value x_j^l if its previous value was x_i^l, under the influence of the action a_k^l, $i \in \{1, 2, 3\}$, $j \in \{1, 2, 3\}$, $k \in \{1, 2, 3\}$, $l \in \{1, 2, 3, 4\}$. For the given example, the matrices of the transition probabilities are:

a_k^l	x_1^l	x_2^l	x_3^l
x_1^l	$p_{11}^k(l)$	$p_{12}^k(l)$	$p_{13}^k(l)$
x_2^l	$p_{21}^k(l)$	$p_{22}^k(l)$	$p_{23}^k(l)$
x_3^l	$p_{31}^k(l)$	$p_{32}^k(l)$	$p_{33}^k(l)$

Each transition probability matrix is followed by revenue or cost matrix.

a_k^l	x_1^l	x_2^l	x_3^l
x_1^l	$c_{11}^k(l)$	$c_{12}^k(l)$	$c_{13}^k(l)$
x_2^l	$c_{21}^k(l)$	$c_{22}^k(l)$	$c_{23}^k(l)$
x_3^l	$c_{31}^k(l)$	$c_{32}^k(l)$	$c_{33}^k(l)$

The state space $\mathcal{S}$ for the four-dimensional MDP is defined as the set of all ordered quadruplets formed by the elements of the value sets of X_1, X_2, X_3, and X_4, and that is $\mathcal{S} = \{(x_i^1, x_j^2, x_k^3, x_l^4), i \in \{1, 2, \ldots, |R_{X_1}|\}, j \in \{1, 2, \ldots, |R_{X_2}|\}, k \in \{1, 2, \ldots, |R_{X_3}|\}, l \in \{1, 2, \ldots, |R_{X_4}|\}\}$ and it consists of $|\mathcal{S}| = |R_{X_1}| \cdot |R_{X_2}| \cdot |R_{X_3}| \cdot |R_{X_4}| = n_1 \cdot n_2 \cdot n_3 \cdot n_4$ states. The number of all possible transitions between the states of the system is calculated by $|\mathcal{S}|^2$. The action space is defined similarly as the state space $\mathcal{A} = \{(a_i^1, a_j^2, a_k^3, a_l^4), i \in \{1,2, \ldots, |A_1|\}, j \in \{1,2, \ldots, |A_2|\}, k \in \{1,2, \ldots, |A_3|\}, l \in \{1,2, \ldots, |A_4|\}\}$ and it consists of $|\mathcal{A}| = |A_1| \cdot |A_2| \cdot |A_3| \cdot |A_4| = m_1 \cdot m_2 \cdot m_3 \cdot m_4$ actions. Using the fact that the random processes are independent, a method for calculating the joint transition probabilities is proposed, knowing those transition probabilities in the one-dimensional MDPs. The number of transition probabilities is $|\mathcal{S}|^2 \cdot |\mathcal{A}|$. For the transition $(x_{i_1}^1, x_{i_2}^2, x_{i_3}^3, x_{i_4}^4) \xrightarrow{(a_{k_1}^1, a_{k_2}^2, a_{k_3}^3, a_{k_4}^4)} (x_{j_1}^1, x_{j_2}^2, x_{j_3}^3, x_{j_4}^4)$, where $i_1, j_1 \in \{1, 2, \ldots, |R_{X_1}|\}$, $i_2, j_2 \in \{1, 2, \ldots, |R_{X_2}|\}$, $i_3, j_3 \in \{1, 2, \ldots, |R_{X_3}|\}$, $i_4, j_4 \in \{1, 2, \ldots, |R_{X_4}|\}$, $k_1 \in \{1,2, \ldots, |A_1|\}$, $k_2 \in \{1,2, \ldots, |A_2|\}$, $k_3 \in \{1,2, \ldots, |A_3|\}$, $k_4 \in \{1,2, \ldots, |A_4|\}$, the transition probability is calculated by $p_{i_1 j_1}^{k_1}(1) \cdot p_{i_2 j_2}^{k_2}(2) \cdot p_{i_3 j_3}^{k_3}(3) \cdot p_{i_4 j_4}^{k_4}(4)$, and the corresponding revenue is calculated by $c_{i_1 j_1}^{k_1}(1) + c_{i_2 j_2}^{k_2}(2) + c_{i_3 j_3}^{k_3}(3) + c_{i_4 j_4}^{k_4}(4)$.

The primary transition matrices are stochastic, $\sum_{j_l} p_{i_l j_l}^{k_l}(l) = 1$, so $\sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_4} p_{i_1 j_1}^{k_1}(1)\, p_{i_2 j_2}^{k_2}(2) p_{i_3 j_3}^{k_3}(3) p_{i_4 j_4}^{k_4}(4) = \sum_{j_1} p_{i_1 j_1}^{k_1}(1) \sum_{j_2} p_{i_2 j_2}^{k_2}(2) \sum_{j_3} p_{i_3 j_3}^{k_3}(3) \sum_{j_4} p_{i_4 j_4}^{k_4}(4) = 1$, i.e. the new matrices are also stochastic. For example, if $n_1 = n_2 = n_3 = n_4 = 3$, $m_1 = m_2 = m_3 = m_4 = 3$, the system has 12 primary transition matrices 3x3, and that is $9 \cdot 12 = 108$ primary transition probabilities. The number of primary revenues is the same. The state space has $3^4 = 81$ states, and the number of actions in the action space is the same. The number of all possible transitions between the states of the system is $3^8 = 6561$, and $|\mathcal{S}|^2 \cdot |\mathcal{A}| = 3^8 \cdot 3^4 = 3^{12} = 531441$ is the number of the joint transition probabilities. The number of the joint revenues is the same as the number of the joint transition probabilities. Exhaustive enumeration of all stationary policies is only practical in problems with small dimensions. In this model the number of all stationary policies is $|\mathcal{A}|^{|\mathcal{S}|}$ and it is a very big number. For the example, this number is $81^{81} \approx 3{,}87 \cdot 10^{154}$. So, the ranking of all stationary policies is not practical and the focus is on finding the optimal decision i.e. planning policy using some optimality method that can handle these dimensions. Later the policy iteration method is chosen (with and without discount rate). Convergence and optimal solution existence are considered in the research for the real data, collected in concrete enterprise. Table 1 gives the summary of the state space and the action space for the created four-dimensional MDP, for the MDP real example.

Table 1: State and Action spaces for the MDP example.

State Number	State	Action Number	Action
1	$(x_1^1, x_1^2, x_1^3, x_1^4)$	1	$(a_1^1, a_1^2, a_1^3, a_1^4)$
2	$(x_1^1, x_1^2, x_1^3, x_2^4)$	2	$(a_1^1, a_1^2, a_1^3, a_2^4)$
…		…	…
22	$(x_1^1, x_3^2, x_2^3, x_1^4)$	22	$(a_1^1, a_3^2, a_2^3, a_1^4)$
…		…	…
27	$(x_1^1, x_3^2, x_3^3, x_3^4)$	27	$(a_1^1, a_3^2, a_3^3, a_3^4)$
28	$(x_2^1, x_1^2, x_1^3, x_1^4)$	28	$(a_2^1, a_1^2, a_1^3, a_1^4)$
29	$(x_2^1, x_1^2, x_1^3, x_2^4)$	29	$(a_2^1, a_1^2, a_1^3, a_2^4)$
…		…	…
49	$(x_2^1, x_3^2, x_2^3, x_1^4)$	49	$(a_2^1, a_3^2, a_2^3, a_1^4)$
…		…	…
54	$(x_2^1, x_3^2, x_3^3, x_3^4)$	54	$(a_2^1, a_3^2, a_3^3, a_3^4)$
55	$(x_3^1, x_1^2, x_1^3, x_1^4)$	55	$(a_3^1, a_1^2, a_1^3, a_1^4)$
56	$(x_3^1, x_1^2, x_1^3, x_2^4)$	56	$(a_3^1, a_1^2, a_1^3, a_2^4)$
…		…	…
76	$(x_3^1, x_3^2, x_2^3, x_1^4)$	76	$(a_3^1, a_3^2, a_2^3, a_1^4)$
…		…	...
81	$(x_3^1, x_3^2, x_3^3, x_3^4)$		$(a_3^1, a_3^2, a_3^3, a_3^4)$

The calculated values of the primary transition probabilities and the primary revenues are input for the software designed to calculate the joint transition probabilities and the associated revenues for the four-dimensional MDP. Because of the relatively small number of states and actions in the example and the relatively short time of finding the optimal solution, the discounted policy iteration method to solve the MDP is chosen. It gives the optimal decision policy and the respective state-value functions for every state for the optimal policy, i.e. the average expected returns for every state. The optimal policy, determines the associate matrix of transition probabilities P and the matrix of revenues R. The vector $X = (x_1, x_2, \dots, x_{81})^T$ of the long-run stationary probabilities for the optimal policy is determined by solving the system of linear equations obtained from the matrix equation $P \cdot X = X$ and the equation $x_1 + x_2 + \dots + x_{81} = 1$. Further vectors $v = diag(P \cdot R^T)$ and $E = X^T \cdot v$ are calculated. The value of E is the expected revenue of the optimal policy per transition step and later it reflects the improvement of the condition of the system [14].

5 REAL APPLICATION AND RESULTS

The mathematical model, developed as part of the methodology for generating optimal decision policies and success factors, was applied to a specific enterprise within the printing industry to address a real-world quality-accuracy management issue focused on minimizing scrap and reducing associated costs. According to their PMS, the sub-key element of success quality-accuracy is located as critical. The opinion of the experts from the company was that the importance of the sub-key element quality-accuracy in this company is 0.7. The performance of this element was measured by the indicator percent of scrap. The average value of percent of scrap for the selected sample was 13.59%, which was grade 4 as per the scale for the performance-axis in the I/P (Importance/Performance) matrix. The developed mathematical model proposed optimal or suboptimal decision policy to improve the performance of the located critical sub-key element.

The printing machine is linked to the design studio through specialized software and includes functionality to record total scrap production. However, it does not categorize scrap by individual causes. Instead, operators manually document these causes in detailed daily reports, which are subsequently processed into specific forms to track scrap sources. This system provides a clear method for identifying the root causes of scrap, but as discussed earlier, this can be enabled with the new IT development. A portion of these records was made available for this research. The collected data for scrap production referred to 396 consecutive orders (printing machine runs), possible causes for scrap, appropriate corrective actions and associate costs. This information was instrumental in modeling the real problem. Considering the expert's opinion and the analysis of two measures, percent of scrap and number of scrap sheets, it is concluded that it is more appropriate to use the measure (the indicator) percent of scrap and the average of percents of scrap in defining the states of MDPs. Real data of available corrective actions were used to define action spaces for MDPs. Clearly, the available sets of states and corrective actions are too big and complex and the model needed to simplify the state and action sets in order to avoid the problem of dimensionality.

According to the opinion of the experts from the company and the collected real data for 396 consequent series with different number of sheets printed on the monitored machine, it is decided that it's appropriate to consider three intervals of percents of scrap for the scrap cause – the machine. This defined the states for the random process "condition of machine" based on percent of scrap it caused for the considered sample. The limits are determined according to the average of percents of scrap for all series. The percents of scrap caused by the machine for the monitored sample were provided from real records. The operator as factor of influence is most difficult to evaluate, because of its complex and unpredictable behavior in different situations. Therefore, the systems in which the influence of the operator is greater are more difficult for modeling and analyzing. The modeling and the analysis in this paper mainly rely on collected real data. The condition of the operator is defined in terms of the induced percent of scrap, for which the cause is the operator. The condition of the tools is defined in terms of the induced percent of scrap, for which the cause are the tools. The condition of the materials is defined in terms of the induced percent of scrap, for which the cause are the materials. Intervals of percents of scrap to define the states of all four causes are determined based on real data, the average of percents of scrap of the considered sample for this cause and experts' opinion.

Based on corrective actions taken in reality, three types of actions were selected for each cause. To make the model more realistic in the future, the spaces of states and actions should be more detailed and more comprehensive, but large size problems require more complex algorithms and software for solving. However, the mathematical model is open in this sense and has opportunity to explore larger issues.

The collected real data were used to calculate the transition probabilities with the formula $p_{ij}^k(l) = \frac{N_{ij}^k(l)}{N_i^k(l)}$, where $N_i^k(l)$ denotes the number of times the one-dimensional MDP described by the random process l was in state with index i, under the influence of primary action with index k, and $N_{ij}^k(l)$ denotes the number of times it made transition from state with index i to state with index j, under the influence of primary action with index k. The notations of the primary states and actions are simplified identifying them with their indexes. For the given example, $i \in \{1,2,3\}$, $j \in \{1,2,3\}$, $k \in \{1,2,3\}$, $l \in \{1,2,3,4\}$.

Real data give information of all transitions from one state to another under the influence of certain action. Thus, the transition probabilities were determined [14].

The enterprise uses special price lists and software to calculate the price for an order. These information were used to calculate the elements of costs matrices. The revenue matrices that follow the percent of scrap from the collected real data of percent of scrap divided by cause were obtained. They were calculated separately for each cause. First the differences between the percents of scrap for all consecutive transitions from one state to another under the influence of some action were calculated. Then they were divided in 18 columns representing the feasible transitions. Those representing the same transition under the same action were put in the same column. After that, averages for each column were calculated, for the considered sample, and for every cause the average percent of scrap's increase or decrease for every transition under the influence of every action was obtained. These values were denoted by $r_{ij}^k(l)$, $i, j, k \in \{1,2,3\}$, $l = 1,2,3,4$. Clearly, for a larger sample, more accurate results should be obtained. Thus, the percent of scrap is followed individually for each cause and totally for the whole system. Also, the optimal decision policy was obtained, minimizing the percent of scrap. For all unfeasible transitions, for $l = 1,2,3,4$, $r_{21}^1(l) = r_{31}^1(l) = r_{32}^1(l) = r_{12}^2(l) = r_{13}^2(l) = r_{23}^2(l) = r_{12}^3(l) = r_{13}^3(l) = r_{23}^3(l) = 0$.

For the considered model the output from the designed software was 81 transition probabilities matrices with dimensions 81x81 and 81 revenue matrices with the same dimensions. Applying policy iteration method, for different discount rates, different optimal policies were obtained represented in the form of two vector columns with dimensions 81x1, which means that for every state of the system (numerated from 1 to 81 in Table 3), optimal policy suggests the corrective action numerated in Table 4. The other output vector column consists of state-value functions i.e. the average expected return for every state. Optimal policy determines transition probability matrix P and revenue matrix R. Vector $X=(x_1,x_2,\dots,x_{81})^T$ of the stationary probabilities is solution to the system of linear equations obtained from $P\cdot X=X$ and $x_1+x_2+\dots+x_{81}=1$. Further, vectors $v=(v_1,v_2,\dots,v_{81})^T$, $v=diag(P\cdot R^T)$ are determined and $E=X^T\cdot v$ is obtained. The obtained value is $E=5.2107\%$ which is the expected percent of scrap for the optimal policy per transition step and grade 8. The results for discount rate 0.99 are given in Table 2. The optimal decision policy can be compared to any other decision policy, or any two stationary policies can be compared. Figure 1 shows the revised I/P matrix.

Table 2: Results from the optimization of the percent of scrap for discount factor 0.99.

State i	Optimal policy	V *1.0e+03	x_i	v_i
1	81	1.2577	0.0164	-36.5635
2	81	1.2580	0.0116	-30.7814
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
14	78	1.2601	0.0009	7.4408
15	78	1.2624	0.0019	8.2197
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
80	81	1.2831	0.0705	30.6179
81	81	1.2831	0.0705	30.6179

The average value of percent of scrap for the selected sample is 13.5928%, and with the optimal policy it decreased to 5.2107%. This improved value for the performance of the sub-key element quality-accuracy is entered in the revised I/P matrix, which shows the improvement of the condition of the system.

In order to examine the sensitivity of the model to changes in transition matrices, the performed sensitivity analysis led to the conclusion that the model is not very sensitive to such changes.

Determination of cost matrices was also needed to find the optimal decision policy minimizing costs. In these matrices costs are with negative sign, and the revenue with positive sign. For all unfeasible transitions, for $l=1,2,3,4$, $c^1_{21}(l)=c^1_{31}(l)=c^1_{32}(l)=c^2_{12}(l)=c^2_{13}(l)=c^2_{23}(l)=c^3_{12}(l)=c^3_{13}(l)=c^3_{23}(l)=0$. Other costs are computed with the $c^k_{ij}(l)=-t^k_l+\frac{r^k_{ij}(l)\cdot 1470\cdot 12.7}{100}$, $i,j\in\{1,2,3\}$, $l=1,2,3,4$, $k\in\{2,3\}$, and for k=1, $c^1_{ij}(l)=-\frac{r^1_{ij}(l)\cdot 1470\cdot 12.7}{100}$, where 1470 is the average of the sum of the order size and the number of scrap sheets for the considered sample. 12.7 is the average price of one finished sheet calculated according the methodology of the enterprise. The research included analysis of costs t^k_l for the corrective actions for each cause, which gives Table 3, illustrated with matrix.

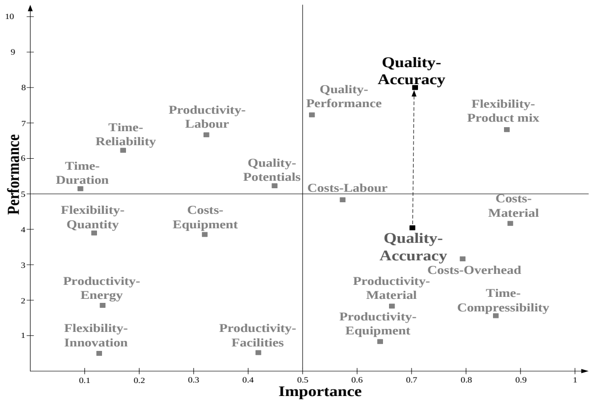

Figure 1: Revised I/P matrix.

Table 3: Costs for the corrective actions for each cause.

t^k_l	Action 1	Action 2	Action 3
Machine	0	250	1000
Operator	0	150	230
Tool	0	200	1473
Material	0	300	1395

The calculated average costs induced with the transitions between the states, under the influence of the primary actions for all the MDPs, were calculated and presented with cost matrices and joined table, similarly as for the transition probabilities in [14].

Table 4: Results from the optimization of the scrap costs for discount factor 0.99.

State i	Optimal policy	V *1.0e+05	x_i	v_i *1.0e+03
1	49	1.3159	0.0527	-4.7413
2	50	1.3151	0.0080	-4.7913
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
14	50	1.3200	0.0017	0.8922
15	50	1.3233	0.0010	-0.1339
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
80	50	1.3558	0.0031	-0.3511
81	50	1.3558	0.0031	-0.3511

Same methodology as for determining the optimal decision policy for minimization of percent of scrap is used to get the optimal decision policy for minimization of costs of scrap and the computed value was $E = -546.0880$ which is the expected cost for scrap for the optimal policy, per transition step. Table 4 shows the results from the calculations with the respective currency.

To compare, in the same way the expected costs per transition step are calculated, for the optimal policy obtained previously minimizing the percent of scrap, and the value is $E = -1189$.

6 EMERGING CHALLENGES AND FUTURE PROSPECTS

Markov Decision Processes (MDPs) are highly suitable for automating manufacturing processes due to their ability to model decision-making in complex and uncertain environments, where current actions influence future states and outcomes. Automated quality control systems benefit from MDPs as they provide a framework for optimizing inspection frequencies and methods, thereby ensuring consistent product quality. Key advantages of incorporating MDPs in manufacturing automation include data-driven decision-making, adaptability to changing conditions, operational efficiency and cost savings, and enhanced predictive capabilities. Within automated quality control systems, various MDP models are particularly effective in managing uncertainty and adapting policies based on quality metrics and outcomes.

However, the successful application of MDPs, especially in automated quality control systems, is subject to several challenges and risks that can limit their effectiveness in real-world industrial settings. Key challenges include inaccurate transition probabilities, oversimplified system assumptions, the need for a stationary environment, data quality and availability issues, the exploration versus exploitation trade-off, model overfitting, computational complexity, and human factors. To address these challenges, industries must prioritize robust model validation and verification, continuously update transition probabilities with real-time data, and incorporate adaptive mechanisms to respond to evolving conditions. Regular testing and refinement of models based on actual operational data are essential for maintaining the accuracy and efficiency of automated quality control systems and reducing scrap in manufacturing.

To further enhance flexibility and adaptability, MDPs can be extended through vector random processes, enabling the simultaneous management of multiple quality and process metrics within a single system. In some models, independent state variables—representing different subsystems such as machines, operators, tools, and materials—simplify the MDP structure by allowing separate consideration of each subsystem's dynamics and transitions. This approach enables a multi-dimensional MDP framework in which each subsystem functions as a one-dimensional MDP, focusing on minimizing scrap by addressing significant contributors individually. Monitoring each subsystem's state facilitates targeted, subsystem-specific interventions, optimizing quality control in manufacturing environments.

Reinforcement Learning (RL)-enhanced MDPs and hierarchical MDPs offer additional adaptability in multi-dimensional contexts, allowing coordinated control across subsystems. These models are particularly promising for reducing scrap in complex manufacturing setups, as they learn and refine optimal policies over time. Nonetheless, determining accurate transition probabilities from historical data remains a challenge and requires ongoing attention to data integrity and validation.

In MDP-based manufacturing quality control models, subsystems are treated as stochastic processes governed by different probability distributions, which capture the uncertainty and variability in state transitions. These distributions form the basis for modeling scenarios such as machine failure, tool wear, material defects, and operator errors. By incorporating these distributions into MDP frameworks, manufacturers can make informed decisions on process adjustments to minimize defects and improve overall accuracy. Specialized software tools like MATLAB/Simulink, AnyLogic, Arena, Python, @Risk, R, and Simio facilitate the simulation and optimization of MDP models, allowing tailored probability distributions to enhance decision-making and production efficiency.

To further enhance quality control, integrated technologies such as sensors, wearable devices, predictive maintenance systems, and data analytics platforms can be used for failure detection and responsibility assignment within subsystems. These tools, when combined within unified systems like Manufacturing Execution Systems (MES), enable precise tracking of defects and their causes, enhancing accountability and proactive mitigation efforts. Emerging technologies, including machine learning (ML), artificial intelligence (AI), predictive analytics, IoT, and cloud-based solutions, are shaping the future of automated quality control by enabling more accurate, proactive defect detection and process optimization.

7 CONCLUSIONS

This study introduces a comprehensive methodology for integrating MDPs into industrial quality control systems, emphasizing automation and decision-making under uncertainty. By developing a four-dimensional MDP model within a PMS framework, the research effectively minimizes scrap and associated costs. Application in a real-world printing enterprise validated the model's practical value, achieving significant scrap reductions and cost savings while providing a structured approach to addressing inefficiencies.

Key contributions include methodologies for transition probability estimation and revenue matrices, alongside policy iteration optimization to generate actionable decisions. The study highlights the adaptability and scalability of MDPs, accommodating additional factors and larger datasets to tackle complex manufacturing challenges.

Integrating advanced IT tools, such as real-time monitoring and data analytics, enhances model accuracy and responsiveness, improving efficiency, product quality, and sustainability. Future work should expand state-action spaces, leverage machine learning and IoT, and address data challenges to advance automation and industrial competitiveness.

ACKNOWLEDGEMENT

We acknowledge support by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and the Open Access Publishing Fund of Anhalt University of Applied Sciences.

REFERENCES

[1] R. Minovski, D. Jovanoski, and K.-P. Zeh, "Ein universelles Modell zur Unternehmensstrukturierung," Compass, REFA – Nachrichten, Heft 5, pp. 53-58, Oct. 2002.

[2] M. Puterman, "Markov Decision Processes: Discrete Stochastic Dynamic Programming," 1st ed., John Wiley & Sons Inc., 1994.

[3] A. Neely, "The performance measurement revolution: why now and what next?" Int. J. Oper. Prod. Manag., 1999.

[4] R. S. Kaplan and D. P. Norton, "The Balanced Scorecard—Measures That Drive Performance," Harv. Bus. Rev., 1992.

[5] W. E. Deming, "Out of the Crisis," MIT Press, 1986.

[6] J. M. Juran, "Juran on Planning for Quality," Free Press, 1988.

[7] U. S. Bititci, P. Garengo, V. Dörfler, and S. S. Nudurupati, "Performance measurement: Challenges for tomorrow," Int. J. Manag. Rev., 2012.

[8] B. Marr, "Key Performance Indicators: The 75+ Measures Every Manager Needs to Know," Pearson UK, 2015.

[9] A. A. Markov, "Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain," 1906.

[10] R. Bellman, "Dynamic Programming," Princeton Univ. Press, 1957.

[11] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction," MIT Press, 1998.

[12] I. Janicevic, J. Filipovic, and J. Miscevic, "Using a Markov Chain for Product Quality Improvement Simulation," UPB Sci. Bull., Ser. D: Mech. Eng., vol. 76, no. 1, pp. 227-242, 2014.

[13] W. B. Powell, "Approximate Dynamic Programming: Solving the Curses of Dimensionality," Wiley-Interscience, 2007.

[14] K. Mitkovska-Trendova, R. Minovski, and D. Boshkovski, "Methodology for transition probabilities determination in a Markov decision processes model for quality-accuracy management," J. Eng. Manag. Competitiveness (JEMC), vol. 4, no. 2, pp. 59-67, 2014.

[15] E. Tolstaya, A. Koppel, E. Stump, and A. Ribeiro, "Nonparametric Stochastic Compositional Gradient Descent for Q-Learning in Continuous Markov Decision Problems," ACC, pp. 6608-6615, 2018.

[16] R. J. Boucherie and N. M. van Dijk, "Markov Decision Processes in Practice," Springer Int. Publ., 2017.

[17] Y. Yang, J. Wu, X. Song, D. Wu, L. Su, and L. Tang, "A Novel Black Box Process Quality Optimization Approach based on Hit Rate," arXiv preprint arXiv:2305.20003, 2023.

[18] C. Reisinger and J. Tam, "Markov Decision Processes with Observation Costs: Framework and Computation with a Penalty Scheme," arXiv preprint arXiv:2201.07908, 2023.

[19] A. Nasir, S. Mekid, Z. Sawlan, and O. Alsawafy, "Optimized Task Assignment and Predictive Maintenance for Industrial Machines using Markov Decision Process," arXiv preprint arXiv:2402.00042, 2024.

[20] S. Chen, D. Simchi-Levi, and C. Wang, "Experimenting on Markov Decision Processes with Local Treatments," arXiv preprint arXiv:2407.19618, 2024.