IMPLEMENTATION AND ANALYSIS OF A DIGITAL TWIN FOR PROPYLENE GLYCOL PRODUCTION: DYNAMIC SIMULATION AND PROCESS STATISTICAL CONTROL

Gabriela Gonçalves Monteiro do Rêgo¹
Francisco Lucas de Lima Carneiro²
Juliana Rosa Leite Araújo Pereira³
Micael Pereira⁴
Antonio Tavernard Pereria Neto⁵
Heleno Bispo da Silva Júnior⁶

ABSTRACT

Purpose: Development of a dynamic simulation and analysis system for Propylene Glycol production through the hydrolysis of Propylene Oxide, utilizing market-available tools including a Digital Twin developed in the Unisim/Honeywell Simulation integrated to the PI System AVEVA/OSIsoft. This system aims to provide a realistic industrial environment for subsequent analysis and statical quality control.

Theoretical framework: The manufacturing of Propylene Glycol involves complex reactive systems that require precise control and optimization. The integration of advanced simulation tools and statistical control charts, such as Shewhart Control Charts, is pivotal in monitoring and enhancing the quality of the process. This approach addresses the need for efficient process management in the chemical industry.

Method/design/approach: The project employs a decoupled simulation structure, where the reactive system and the statistical control system operate simultaneously, exchanging information through the PI System supervisory software (AVEVA/OSIsoft) and a Python programming environment. The implementation of Shewhart Control Charts within the simulation provides real-time process quality monitoring.

Results and conclusion: The outcomes demonstrate the feasibility of using a combination of software tools for daily industrial event simulation, as well as for training and acclimatizing engineers and operators. The analysis of statistical control charts allows for the detection of special cause variations, offering valuable insights for efficient process intervention, thereby enhancing the quality of both the process and the final product.

Research implications: This study contributes to a deeper understanding of reactive systems in chemical engineering and the effective integration of simulation and statistical process control techniques.

Originality/value: The development of a simulation framework that combines Digital Twin technology with statistical process control for the efficient and reliable production of Propylene Glycol.

¹ Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brazil. E-mail: gabriela.goncalves@eq.ufcg.edu.br Orcid: https://orcid.org/0009-0005-6229-2760
² Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brazil. E-mail: heleno.bispo@eq.ufcg.edu.br Orcid: https://orcid.org/0000-0002-7570-303X
³ Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brazil. E-mail: leite.jrlap@gmail.com Orcid: https://orcid.org/0009-0005-6552-9030
⁴ Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brazil. E-mail: apmicael@hotmail.com Orcid: https://orcid.org/0009-0003-4715-7697
⁵ Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brazil. E-mail: tavernard@eq.ufcg.edu.br Orcid: https://orcid.org/0000-0002-2938-2140
⁶ Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brazil. E-mail: heleno.bispo@eq.ufcg.edu.br Orcid: https://orcid.org/0000-0002-2138-124X
Implementation and Analysis of a Digital Twin for Propylene Glycol Production: Dynamic Simulation and Process Statistical Control

Keywords: Propylene Glycol, Reactive System Simulation, Digital Twin, Statistical Process Control, Shewhart Control Charts.

IMPLEMENTAÇÃO E ANÁLISE DE UM DIGITAL TWIN PARA A PRODUÇÃO DE PROPILENO GLICOL: SIMULAÇÃO DINÂMICA E CONTROLE ESTATÍSTICO DE PROCESSO

RESUMO

Objetivo: Desenvolvimento de um sistema de simulação e análise dinâmica para a produção de Propileno Glicol através da hidrólise de Óxido de Propileno, utilizando ferramentas disponíveis no mercado, incluindo um Digital Twin desenvolvido no software comercial de simulação Unisim/Honeywell integrado ao PI System da AVEVA/OSIsoft. Este sistema visa proporcionar um ambiente industrial realista para análises e controle estatístico de processos.

Referencial teórico: A fabricação de Propileno Glicol envolve sistemas reativos complexos que requerem controle e otimização precisos. A integração de ferramentas avançadas de simulação e cartas de controle estatístico, como as Cartas de Controle de Shewhart, é fundamental no monitoramento e melhoria da qualidade do processo. Esta abordagem atende à necessidade de gestão eficiente de processos na indústria química.

Método: O projeto emprega uma estrutura de simulação desacoplada, onde o sistema reativo e o sistema de controle estatístico operam simultaneamente, trocando informações através do software de supervisão PI System (AVEVA/OSIsoft) e um ambiente de programação em Python. A implementação das Cartas de Controle de Shewhart na simulação proporciona um monitoramento em tempo real da qualidade do processo.

Resultados e conclusão: Os resultados demonstram a viabilidade do uso combinado de ferramentas de software para simulação de eventos industriais diários, bem como para o treinamento e ambientação de engenheiros e operadores. A análise das cartas de controle estatístico permite a detecção de variações de causas especiais, oferecendo insights valiosos para intervenções eficientes no processo, melhorando assim a qualidade do processo e do produto final.

Implicações da pesquisa: Este estudo contribui para um entendimento mais profundo dos sistemas reativos em engenharia química e a integração eficaz de técnicas de simulação e controle estatístico de processo.

Originalidade/valor: O desenvolvimento de uma estrutura de simulação que combina a tecnologia Digital Twin com controle de processo estatístico para a produção eficiente e confiável de Propileno Glicol.


RGSA adota a Licença de Atribuição CC BY do Creative Commons (https://creativecommons.org/licenses/by/4.0/).

1 INTRODUCTION

The focus on product quality in the plant production is closely connected with managing process variability. In the manufacturing area, variability arises from various sources, including data inconsistencies, raw materials, and measurement methods. This variability is broadly classified into natural and special causes. Natural cause variations are consistent and inherent to the process, with a variability in a range, while special cause variations are erratic, arising from unforeseen disturbances that necessitate immediate attention for process integrity. On this way, the establishment of statistical control is pivotal, where the process is adjusted until only natural cause variations remain. For this the need on the use of Statistical Process Control (SPC), a methodology that employs a range of statistical tools to maintain and improve process
quality. Central to SPC is the development and application of control charts, particularly Shewhart control charts, which are instrumental in monitoring quality parameters and managing process deviations in real time.

In the current era of digital transformation, typified by Industry 4.0 and the Industrial Internet of Things (IIoT), the role of data analysis in process management has become more pronounced. Advanced data processing technologies now play as mandatory in process evaluation and decision-making. Thus, the integration of sophisticated software like Python with systems like the PI System (AVEVA/OSIsoft) offers a robust basis for automated statistical control. This project aims to use these technologies to create a Python-based tool that uses data from the PI System to automatically generate Shewhart control charts. This initiative extends beyond mere process control, encompassing comprehensive data management, real-time analysis, and support for strategic decision-making in a highly dynamic manufacturing environment.

2 THEORETICAL REFERENCE

In the production chain, there is a concern regarding the quality level of the target product, which is inversely proportional to the level of variation in the manufacturing process. Since variability can only be described in statistical terms, its reduction in a process, and the consequent quality improvement, positions statistical methods as central to the monitoring and control of stages in industrial development (Montgomery, 2009). As highlighted by Ramos (2000), all real processes exhibit variability due to inconsistencies among workers, batches of raw materials, measuring instruments, among others. However, the causes of variation are divided into two groups: common and special.

Variations from common causes are sources of variation that affect all individual values of a process and tend to form patterns, being always present and cannot be reduced without changes in the process design. Special cause variations stem from factors that affect the process behavior unpredictably: they are disturbances or perturbations and require immediate correction to prevent further damage to the process. A process is said to be statistically stable when only common causes are present. To achieve this, special causes must be identified and eliminated until the process is back under statistical control.

Once the cause of variability is identified as common or special, it becomes possible to decide on how to proceed with production. To identify disturbances harmful to the process, Statistical Process Control (SPC) can be applied. SPC consists of a set of statistical tools that, operating as a sampling inspection system, allow for the systematic reduction of the unnatural variability of the main quality characteristics of the target product, resulting in improvements in productivity, reliability, and cost of the processed items. Notably, problems eliminated before affecting the final product imply a reduction in the cost associated with reprocessing time and analysis of batches that are out of specification.

Among the available tools for statistical control, the "Magnificent Seven" stand out: histograms or stem-and-leaf plots, control charts, Pareto charts, cause-and-effect diagrams, defect concentration diagrams, scatter diagrams, and control charts, with a particular focus here on the development of Shewhart control charts. The control chart is one of the tools available in SPC and enables effective control of a particular quality characteristic in real time. It allows for the detection of deviations from representative process parameters, reducing the number of products out of specification and production costs, and determining process capability, among others. To define the type of chart, it is important to know the process characteristics, such as the size of the sample subgroup and the magnitude of the deviation to be detected.
The construction of each chart takes into account specific characteristics and generally assumes that the process is statistically stable, that is, there are no special causes acting on the process. Thus, the use of Control Charts enables the training of operators/users to act preventively in the process, correcting possible quality deviations online. The basic structure of control charts consists of a central line, between the upper and lower control limits, as depicted by Figure 1.

![Control Chart](typical_representation_of_a_control_chart.png)

**Figure 1:** Typical Representation of a Control Chart.

**Source:** Author's own work

Characteristic values of the variable of interest are plotted on the chart over time. When these values remain within the control limits, without exhibiting any trend, the process is considered under statistical control. However, if the points fall outside the control limits or exhibit a predictable arrangement or any kind of trend, the process is deemed out of control (Montgomery, 2009). The joint use of software tools (Python and PI System) then enables the application of SPC through control charts for the detection of special causes (disturbances). Case studies were conducted using process data stored in the PI System, owing to its widespread industrial use as a tool for storing, analyzing, and distributing data and information.

The PI System comprises three basic components: the first is the PI Server – software operating in a Windows Server environment – a second component, the PI Interface, which can be installed on the same machine where the PI Server is installed or on different machines, and the third component is the client tools (PI Clients) – visualization tools that display in real time the information of interest received by the previous two components in the form of tables, charts, and other visual resources for the end-user – used for monitoring and investigating the data. The infrastructure of these components is detailed in Figure 2.
In the context of Industry 4.0 and the Industrial Internet of Things (IIoT), a variety of analysis and correlation technologies have emerged to predict situations based on data collected from multiple sources. Using a data infrastructure, it is possible to develop applications that use this information for process evaluation, making operational, economic, or strategic decisions, as well as for the development of machine learning structures and phenomenological models with numerous applications (Roffel and Betlem, 2006), and in many cases being considered as a Process Digital Twin. Decision-making in this scenario depends primarily on the monitoring and in-depth analysis of an extensive series of data and information generated at every moment of the process. To filter, store, structure, interpret, and properly allocate this conglomerate of information, the use of PI System (AVEVA/OSIsoft) and Python software is proposed to enable automatic statistical control.

Thus, this work aims to develop a Python-based computational tool and via OPC Communication that, utilizing data provided through the PI System (AVEVA/OSIsoft) for the automatic construction of Shewhart control charts (Montgomery, 2009; Wheeler, 2004). This includes developing functionalities for initial SPC implementation in Python, establishing calculation routines for Shewhart Control Charts, and integrating Python with the PI System for asset TAG creation and data communication. Additionally, the project focuses on automating the PI Asset Framework for data management, creating monitoring interfaces for result analysis, assessing outcomes online through PI Vision. With this, it will facilitate the evaluation of ongoing production processes, as well as comparative studies of results from the literature, enabling real-time analysis and the development of structures for operational, economic, and/or strategic decisions.

3 METHODOLOGY

During this study, an economic indicator for the Propylene Glycol production process (Fogler, 1999), the cooling water consuming has been analyzed from a process simulation using the UniSim/Honeywell software interconnected to PI System AVEVA/OSIsoft by OPC protocol. This process involves the hydrolysis of Propylene Oxide, a key reaction in the chemical industry, as depicted by Figure 3, and it could be considered as a representation of a Process Digital Twin.
The construction of control charts adhered to the Shewhart methodology, a fundamental approach in statistical process control. These charts were developed within the Python environment, utilizing the standard parameters (Montgomery, 2009), some of such data has been detailed in Table 1. Such a table outlines the factors for the construction of control charts for variables, providing a comprehensive framework for assessing the process performance.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>X-bar Chart</th>
<th>R Chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>size, n</td>
<td>A2</td>
<td>D3</td>
</tr>
<tr>
<td>2</td>
<td>1.880</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1.023</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0.729</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0.577</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0.483</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0.419</td>
<td>0.076</td>
</tr>
<tr>
<td>8</td>
<td>0.373</td>
<td>0.136</td>
</tr>
<tr>
<td>9</td>
<td>0.337</td>
<td>0.184</td>
</tr>
<tr>
<td>10</td>
<td>0.308</td>
<td>0.223</td>
</tr>
</tbody>
</table>

The data analysis can be conducted online, using the calculated control limits of the charts in the python environment and also through web-based monitoring using PI Vision. This dual approach underscores the integration of traditional process control with modern, digital solutions. The integration of this project with Industry 4.0 principles is evident in its use of digital tools and online platforms for real-time analysis and monitoring.

Based on this, the fusion of advanced technologies like the Internet of Things (IoT), big data analytics, and cloud computing, aiming to enhance connectivity, efficiency, and automation in industrial processes, can be reached. In this context, the use of Python and PI Vision represents an advance to a more interconnected, and data-driven manufacturing processes. By leveraging these tools, the project not only facilitates the efficient monitoring and control of the chemical production process but also aligns with the broader objectives of Industry 4.0: enhancing process efficiency, reducing downtime, and enabling more informed decision-making through real-time data analysis.
About the reactive process, it is related to an exothermic reaction, demanding a temperature control system to maintain the process within the specified operational conditions. The chosen economic indicator for this study is closely linked to the consumption of the coolant used in the cooling jacket of the reactor. An efficient management of this coolant consumption is important once it directly influences the operational cost and safety of the process. The communication and data processing has been made using the Spyder software (Anaconda) directly integrated with the PI System (AVEVA/OSIsoft). The algorithmic routine established for this purpose includes: the importing of necessary libraries, data acquisition, computing the mean and range of the samples, and limits calculation. Such computed values are then transmitted to designated tags in the PI server, which is responsible for storage and distribution data.

The Control charts has been constructed by sampling specific values indicative of water consumption in the process. Each chart is composed of 25 subgroups, with 3 samples for subgroup. The data collection and storage system reads the variable signals from the supervisory system, and the developed code, Figure 4, captures values from these PI System pre-stored data, organizing them into a dataframe. In this dataframe, the number of rows correlates to the number of subgroups, and the columns correspond to the number of samples in each subgroup. After filling in the water consumption data, X-bar and R charts limits, for the variable, can be calculated. Such limits are then sent to the PI System storage environment, the PI Server, from where they can be retrieved for display and process monitoring in the web by PI Vision.

```python
# Configurações
TAG_NAME = 'MPI - GreenScope.Economic Indicator.c7dd2c30-094b-476a-80dd-362421823793' #Tag que armazena os dados do processo de interesse
INTERVALS = [
    ('2023-04-27 06:00 AM', '2023-04-27 07:30 AM'),
    ('2023-04-27 08:00 AM', '2023-04-27 09:30 AM'),
    ('2023-04-27 10:15 AM', '2023-04-27 11:45 AM'),
]
SAMPLING_INTERVAL = '4m' #Intervalo entre cada amostra coletada

# Função para buscar valores interpolados
def fetch_interpolated_values(point, start_time, end_time, sampling_interval):
    return [value for time in point.interpolated_values[start_time:end_time, sampling_interval]

# Função para calcular média e amplitude
def calculate_metrics(row):
    mean = np.mean(row)
    amplitude = np.max(row) - np.min(row)
    return pd.Series({mean, amplitude})

Figure 4: Python code to data acquisition.
Source: Author's own work
```

The data acquisition procedure involves the use of an interpolating function for the defined time interval and the selected variable, followed by the calculation of the mean and range of the acquired sample. This methodical approach not only ensures accurate monitoring and control of the exothermic process but also provides critical insights into the efficiency and economic aspects of coolant usage, thereby aligning with the principles of efficient resource management in process engineering. The samples collected need to be concatenated into a dataframe along with their computed means and ranges. This structured compilation of data is a crucial step in preparing for the subsequent calculation of control limits for the Control Charts. Leveraging the factors outlined in Table 1 and the previously stored means and ranges, the limits for the Control Charts can be calculated and sent to PI System, Figure 5 and 6, respectively.
Implementation and Analysis of a Digital Twin for Propylene Glycol Production: Dynamic Simulation and Process Statistical Control

Figure 5: Python code to chart limits calculation.
Source: Author's own work

```
def limits(mean_of_means, mean_of_amplitudes, n, len(INTERVALS)):  # Tabela com fatores para cálculo dos limites de controle
    A2 = fators[A2]
    D3 = fators[D3]
    D4 = fators[D4]
    if n <= 2:
        # Cálculo dos limites
        ucl_X = mean_of_means + A2*n*mean_of_amplitudes
        cl_X = mean_of_means
        lcl_X = mean_of_means - A2*n*mean_of_amplitudes
        ucl_R = D4*n*mean_of_amplitudes
        cl_R = mean_of_amplitudes
        lcl_R = D4*n*mean_of_amplitudes
        limits = ([ucl_X, cl_X, lcl_X, ucl_R, cl_R, lcl_R])
```

Figure 6: Python code to transmission to PI System environment.
Source: Author’s own work

```
# Envio dos limites calculados para o PI System
ucl_X = server.search('PGRx.OUT_OUT63')[0]  # As tags utilizadas devem ser previamente criadas
ucl_X.update_value(limits[0], datetime.now(), UpdateMode.NO_REPLACE, BufferModeUFFER_IF_POSSIBLE)
cl_X = server.search('PGRx.OUT_OUT64')
cl_X.update_value(limits[1], datetime.now(), UpdateMode.NO_REPLACE, BufferModeUFFER_IF_POSSIBLE)
lcl_X = server.search('PGRx.OUT_OUT65')
lcl_X.update_value(limits[2], datetime.now(), UpdateMode.NO_REPLACE, BufferModeUFFER_IF_POSSIBLE)
ucl_R = server.search('PGRx.OUT_OUT66')
ucl_R.update_value(limits[3], datetime.now(), UpdateMode.NO_REPLACE, BufferModeUFFER_IF_POSSIBLE)
cl_R = server.search('PGRx.OUT_OUT67')
cl_R.update_value(limits[4], datetime.now(), UpdateMode.NO_REPLACE, BufferModeUFFER_IF_POSSIBLE)
lcl_R = server.search('PGRx.OUT_OUT68')
lcl_R.update_value(limits[5], datetime.now(), UpdateMode.NO_REPLACE, BufferModeUFFER_IF_POSSIBLE)
```

This approach, integrating data storage with analytical processing, exemplifies the precision and efficiency required in modern process monitoring. The use of a dataframe for aggregating samples and calculated statistics allows for a streamlined process where data visualization and analysis are closely interconnected. This method ensures that the monitoring of the process is not only thorough but also adaptable, allowing for real-time adjustments and insights. Such a methodology is in line with modern process engineering practices and needs, where data-driven decision-making is paramount for optimal process management and quality control.

4 RESULTS AND DISCUSSION

To enable variable of interest monitoring and to effectively track each step executed by the code, the algorithm was developed with a feature to print the obtained results step by step in the python, as seen in Figure 7. With this is possible to evaluate specific stages in the script where modifications might be necessary, especially if errors occur during program execution. The verification of data transmission is confirmed by comparing the stored values in the designated tag to those calculated. Figure 8 highlights the PI tags used for storing in PI System environment for each limit after the conducted tests.
Figure 7: Data evaluation in Python environment.  
**Source:** Author's own work

Figure 8: PI System data evaluation.  
**Source:** Author's own work

Consistency tests has been performed post-chart construction to simulate special causes and assess whether the charts could detect induced disturbances characterized as special causes. The initial tests focused on common process noise, or the natural causes. As expected, and demonstrated in Figure 9, all points remained within the control limits in both charts, indicating a random behavior consistent with a process operating under statistical control. Subsequently, a disturbance was simulated in the process, characterized by an increase in the reactor's inlet flow, mimicking a special cause. As seen in Figure 10, one of the sample points exceeded the upper control limit in both charts, modifying specially the mean for the R Chart.
This disturbance, caused by a sudden increase in feed flow from 30% to 60%, directly impacted the reactor's level and residence time, and in consequence the specified reaction conversion. This change affects the heat demand from the chemical reaction and consequently from the heat exchanger, leading to an increased consumption of the coolant and fluctuations in the indicator under study. The change in residence time impacts the chemical conversion of the process, highlighting the importance of residence time as an optimization variable for the analyzed reaction.
Figure 10: Python Control Chart evaluation after disturbance (special cause).
Source: Author’s own work
Figure 11: PI Vision Dashboard to on-line key variable evaluation and control charts.

Source: Author’s own work

The monitoring screen set up for the proposed structure utilizes the data stored in the PI System during the limit transmission phase and can be illustrated in Figure 10. This setup not only facilitates real-time observation and response to process changes but also underscores the
significance of integrating advanced data management and analysis tools in modern process engineering.

5 CONCLUSION

The integration of Statistical Process Control (SPC) methods with Industry 4.0 technologies marks a significant advance in the field of chemical engineering. By combining traditional SPC tools with modern software like Python and the PI System, this approach realizes a paradigm shift towards smarter manufacturing processes. The use of these tools for real-time data monitoring and analysis is perfectly aligned with the core objectives of Industry 4.0, which emphasize digitalization, interconnectivity, and streamlined decision-making processes.

The architecture of the PI System, encompassing server, interface, and client components, is a demonstration to the layered and interconnected structure of contemporary smart manufacturing systems, ensuring that data is efficiently channeled from operational mechanisms to decision-making entities. Additionally, the employment of Python for data analysis and the implementation of SPC represents the growing significance of versatile, robust programming languages in industrial settings. Python’s compatibility with various data sources and its comprehensive libraries for statistical analysis make it exceptionally suitable for deploying control charts and evaluating process variability within the Industry 4.0 framework. This integration not only reinforces process monitoring and control but also supports broader Industry 4.0 goals such as enhanced automation, optimized data utilization, and proactive maintenance strategies. These advancements enable chemical engineering processes to become more efficient, agile, and resilient, which are critical attributes in the competitive, technologically advanced industrial landscape.

Furthermore, the use of commercial software based on phenomenological models in this context underscores the commitment to analytics and informed decision-making, integral aspects of Industry 4.0. The use of OPC Communication for data transfer exemplifies the interoperability and seamless connectivity that Industry 4.0 encourages. This holistic approach, which incorporates advanced data analytics, digital twin, and Industrial Internet of Things (IIoT) technologies, culminates in a comprehensive tool for process monitoring, analysis, and decision support. It not only enhances operational efficiency and strategic planning but also aligns chemical engineering practices with the innovative and dynamic spirit of Industry 4.0, paving the way for future advancements in smart manufacturing.

REFERENCES


