Pontifícia Universidade Católica do Rio de Janeiro

**Cristiano Saad Travassos do Carmo**

# A hybrid solution using stochastic and neural networks modeling for the consideration of safety uncertainties in construction planning methods

**Doctoral thesis**

Thesis presented to the Department of Civil and Environmental Engineering of Pontifical Catholic University of Rio de Janeiro (DEC/PUC-Rio) in partial fulfillment of the requirements for the degree of Doctor of Civil Engineering

Advisor: Prof. Elisa Dominguez Sotelino

Rio de Janeiro
December 2023

Pontifícia Universidade Católica do Rio de Janeiro

Cristiano Saad Travassos do Carmo

# A hybrid solution using stochastic and neural networks modeling for the consideration of safety uncertainties in construction planning methods

Thesis presented to the Department of Civil and Environmental Engineering of Pontifical Catholic University of Rio de Janeiro (DEC/PUC-Rio) in partial fulfillment of the requirements for the degree of Doctor of Civil Engineering

**Prof. Elisa Dominguez Sotelino**
Advisor
Department of Civil and Environmental Engineering – PUC-Rio

**Prof. Daniel Carlos Taissum Cardoso**
Department of Civil and Environmental Engineering – PUC-Rio

**Prof. Fernanda Araujo Baião**
Department of Industrial Engineering – PUC-Rio

**Prof. Renata Gonçalves Faisca**
Department of Civil Engineering – UFF

**Prof. Sergio Luiz Braga França**
Department of Civil Engineering – UFF

Rio de Janeiro, December 1st, 2023.

**Cristiano Saad Travassos do Carmo**

M.Sc. Degree in Civil Engineering at Pontifical Catholic University of Rio de Janeiro (PUC-Rio) in 2019. Adjunct professor of the Civil Engineering course at Pontifical Catholic University of Rio de Janeiro (PUC-Rio) since 2022.

Bibliographic data

Carmo, Cristiano Saad Travassos do

A hybrid solution using stochastic and neural networks modeling for the consideration of safety uncertainties in construction planning methods / Cristiano Saad Travassos do Carmo ; advisor: Elisa Dominguez Sotelino. – 2023.
201 f. : il. color. ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Civil e Ambiental, 2023.
Inclui bibliografia

1. Engenharia Civil e Ambiental - Teses. 2. Método de planejamento de obras. 3. Incertezas. 4. Ocorrências de segurança. 5. Processo estocástico. 6. Redes neurais. I. Sotelino, Elisa Dominguez. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Civil e Ambiental. III. Título.

CDD: 624

I dedicate this thesis to my guardian
angel Maria Alice Oliveira de
Souza.

# Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPEs) – Finance Code 001.

I would like to thank the Pontifical Catholic University of Rio de Janeiro for providing me with adequate structure and partial funding over the past four years as I have developed my studies. Also, Prof. Elisa Sotelino deserves special mention for her technical assistance during my PhD journey. I'd also like to thank her for all of the opportunities and lessons she's given me, always guiding me in the best academic path. I also want to express my gratitude to the professors who agreed to support this study and take part in the thesis presentation.

During my PhD journey, I lost one of the most important people in my life, a non-biological mother who had cared for me since I was a baby. She taught me how to read and write, as well as how to be simple and happy, and for that I am grateful to Maria Alice Oliveira de Souza. You will never die in our hearts, no matter where you are.

Also, I'd like to thank my wife for encouraging me to pursue my dream of becoming a professor. Even when I was very stressed and bored with my progress, she gave me reasons to believe that I could do it.

Special thanks to my mother for being my hero. My father, for being an example of happiness, empathy, and a healthy dose of craziness. My brother, for setting a good example of professionalism and friendship. Nana, my nanny, for being a warrior role model for me.

I'd like to express my gratitude to my family for all of their emotional support and moments of relaxation during this time, especially during the months of lockdown caused by the COVID-19 pandemic.

I'd like to thank my RioBIM colleagues for listening to me many times during the week meetings and assisting me in providing ideas and solutions for better work.

Finally, I'd like to remember all of the workers who died while building the "Ponte Rio-Niterói" bridge, including Domicio Barbosa Lima, Luiz Carlos Couto, Wilson Touso, José Machado, Nilkom Vianna, Gilberto Leal Costa, José Machado, Raul Arajo Arendes, and others.

## Abstract

Carmo, Cristiano Saad Travassos do; Sotelino, Elisa Dominguez (Advisor). **A hybrid solution using stochastic and neural networks modeling for the consideration of safety uncertainties in construction planning methods**. Rio de Janeiro, 2023. 201 p. Doctoral thesis - Department do Civil and Environmental Engineering, Pontifical Catholic University of Rio de Janeiro.

The construction industry, known for its dynamic and chaotic nature, often experiences work accidents. Existing planning methods addressing uncertainties, however, frequently overlook safety variables, and the relevant literature is scarce. This study introduces a novel construction planning method focused on investigating the impact of safety incidents on project duration, specifically in energy infrastructure construction projects. The main hypothesis is that safety events during construction significantly affect project duration, leading to deficient schedules when not considered in the planning process. Utilizing stochastic process theory, particularly the quasi birth and death process, the study explores how safety states influence delay states. Neural network models complement the stochastic model for forecasting bivariate time series derived from safety and delay stochastic states. Real-life project data demonstrates that safety events, assuming planned delay events, are over double the delay states' value. Applying the stochastic model to a real project with a planned 8-day delay indicates a most probable safety state of 19. Long short-term memory models outperform statistical methods in bivariate time series forecasting, with a significantly smaller root mean square estimation metric. The proposed hybrid construction planning approach proves suitable for both pre-construction and construction phases, offering improved decision-making indicators and supporting reactive safety management.

## Keywords

Construction planning method; Uncertainties; Safety events; Stochastic process; Neural networks; Energy infrastructure construction projects.

# Resumo


Carmo, Cristiano Saad Travassos do; Sotelino, Elisa Dominguez (Orientadora). **Uma solução híbrida utilizando modelagem estocástica e de redes neurais para a consideração de incertezas de segurança em métodos de planejamento de construção**. Rio de Janeiro, 2023. 201 p. Tese de Doutorado – Departamento de Engenharia Civil e Ambiental, Pontifícia Universidade Católica do Rio de Janeiro.

Na indústria da construção, conhecida por sua natureza dinâmica e caótica, muitas vezes há acidentes de trabalho. Os métodos de planejamento existentes que abordam incertezas, no entanto, frequentemente ignoram as variáveis de segurança, e a literatura relevante é escassa. Este estudo introduz um novo método de planejamento de obras focado na influência de ocorrências de segurança na duração do projeto, especificamente em projetos de construção de usinas de energia. A principal hipótese é que eventos de segurança durante a construção afetam significativamente a duração do projeto, levando a cronogramas deficientes quando não considerados no processo de planejamento. Utilizando a teoria de processos estocásticos, particularmente o processo de quase-nascimento e morte, o estudo explora como os estados de segurança influenciam os estados de atraso. Modelos de redes neurais complementam o modelo estocástico para previsão de séries temporais bivariadas derivadas dos estados estocásticos. Dados reais de projetos demonstram que os eventos de segurança, supondo eventos de atraso planejados, são mais do que o dobro do valor dos estados de atraso. A aplicação do modelo estocástico a um projeto real com um atraso planejado de 8 dias indica um estado de segurança mais provável de 19. Os modelos de memória de curto prazo de longo prazo superam os métodos estatísticos na previsão de séries temporais bivariadas, com uma métrica de estimação quadrática média raiz significativamente menor. A abordagem de planejamento de construção híbrida proposta mostra-se adequada para as fases de pré-construção e construção, oferecendo melhores indicadores de tomada de decisão e apoiando a gestão de segurança reativa.



# Palavras-chave

Método de planejamento de obras; Incertezas; Ocorrências de segurança; Processo estocástico; Redes neurais; Projetos de construção de usinas de energia.

## Summary

## List of figures

## List of tables

# List of equations

# List of abbreviations and symbols

| Acronym | Description |
|---|---|
| ADF | Augmented Dickey Fuller |
| AI | Artificial Intelligence |
| ARIMA | Auto Regression Integrated Moving Average |
| BDP | Birth and Death Processes |
| BIM | Building Information Modeling |
| CNN | Convolutional Neural Network |
| COVID-19 | Coronavirus Disease 2019 |
| CPM | Critical Path Method |
| DES | Discrete Event Simulation |
| DSM | Distributed Scheduling Model |
| EDA | Exploratory Data Analysis |
| FL | Fuzzy Logic |
| FTP | Full Time Project |
| GA | Genetic Algorithms |
| GDP | Gross domestic product |
| GDPR | General Data Protection Regulation |
| GERT | Graphical Evaluation and Review Technique |
| KDD | Knowledge Discovery in Database |
| KDE | Kernel Density Estimation |
| ksN | Number of samples for the Kolmogorov-Smirnov test |
| LATAM | Latin America |
| LOB | Line of Balance |
| LSTM | Long Short-Term Memory |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| MC | Monte Carlo |
| MFTN | Modified Fault Tree Networking |
| ML | Machine Learning |
| MLE | Maximum Likelihood Estimation |
| OSH | Occupational Health and Safety |
| PDF | Probability Density Function |
| PERT | Program Evaluation and Review Technique |
| PR-AUC | Precision-Recall Area Under Curve |
| p-value | Critical value that rejects or not the null hypothesis |
| QBDP | Quasi Birth and Death Processes |
| ReLU | Rectified Linear Unit |
| RMSE | Root Mean Square Error |
| SCPM | Stochastic Critical Path Method |
| SLR | Systematic Literature Review |
| SQL | Structured Query Language |
| SVM | Support Vector Machine |
| VAR | Vector Auto Regression |
| VARIMA | Vector Autoregressive Integrating Moving Average |
| VDC | Virtual Design and Construction |
| XML | Extensible Markup Language |
| $\Upsilon$ | New domain |

| | |
|---|---|
| $\Omega$ | Sample space |
| $n(\,)$ | Stochastic process map function |
| $t$ | Time variable |
| $\boldsymbol{A}$ | Generator matrix for the birth and death process |
| $\boldsymbol{L}, \boldsymbol{F}, \boldsymbol{B}$ | Submatrices included in the generator matrix for the QBDP |
| $\boldsymbol{MS}$ | Matrix similarity score |
| $\boldsymbol{P}$ | Transition probabilities matrix |
| $\boldsymbol{Q}$ | Generator matrix for the quasi birth and death process |
| $\boldsymbol{R}$, $\boldsymbol{G}$, $\boldsymbol{U}$ | Fundamental matrices for the QBDP |
| $\boldsymbol{v}$ | Transition vector |
| $\delta$ | Time increment |
| $\lambda$ | Birth rate |
| $\mu$ | Death rate |
| $S_i$ | Stochastic states |
| $p_0$ | Initial parameter guess for the stochastic birth and death process |
| $p_{bounds}$ | Parameter bounds for the stochastic birth and death process |
| $p_{ij}(\,)$ | Transition probability from state $i$ to state $j$ |
| $x_i$ | Experiment observations |
| $\gamma_{ij}(t)$ | Safety event rate from state $i$ to state $j$ at time $t$ |
| $\lambda_{ij}(t)$ | Delay rate from state $i$ to state $j$ at time $t$ |
| $\mu_{ij}(t)$ | Advance rate from state $i$ to state $j$ at time $t$ |
| $\psi_{ij}(t)$ | Combined safety and advance events rate from state $i$ to state $j$ at time $t$ |
| $\phi_{ij}(t)$ | Combined safety and delay events rate from state $i$ to state $j$ at time $t$ |

*"I am no longer accepting the things I cannot change. I am changing the things I cannot accept."*

Angela Davis

# 1
# Introduction

It is well-known that construction projects are of a unique nature (TAM; ZENG; DENG, 2004), which differently from other industries can turn the workplace dangerous (ZHOU; IRIZARRY; LI, 2013). According to the Brazilian Statistical Yearbook of Works Accidents in 2021, the construction industry had the sixth highest number of accidents compared to all economic activities in the country. At the same time, however, it is one of the oldest industries (40,000 ~12,000 B.C.) and with safety management through the regulations written by the King of Hammurabi (~ 2,200 B.C.) (PÉREZGONZÁLEZ, 2005). As an example, law 229 written in this code says:

> **229** – *If a builder builds a house for someone, and does not construct it properly, and the house which he built falls in and kills its owner, then that builder shall be put to death*.

Despite this, scientific publications in the research field of accident prevention has only grown in the last decade (HUANG et al., 2022). Most studies related to construction project management generally focus on the classical performance indicators: time and cost, and do not consider safety issues as critical for the project success (CARMO; SOTELINO, 2023).

The current study aims to investigate the impact of safety incidents on project duration in energy infrastructure construction projects. The main hypothesis is that a safety event that occurs during the construction phase affects quantitatively the construction duration, and when it is not considered in the planning method, it results in deficient construction schedules. Therefore, this work is based on the research question "How much does a construction accident cost in terms of delay days?". Usually, the construction managers do not measure this safety impact, but the present work infers that safety events, when numerous, can result in many delay events in the project duration.

## 1.1. Motivation

The Brazilian construction industry accounts for an average of 5.1% of gross value added, according to data from the last 22 years from the Brazilian Statistics Institute (IBGE, 2022). Behind this number are several effects on the local economy, such as job creation, reduction of the housing deficit, among others. According to the same sources, in the year 2019, almost 8 million of people were employed by this industry. In view of this, the study of the problems of civil construction and, consequently, of the possible solutions becomes paramount, since it impacts not only the country's economy, but also the lives of workers.

Indeed, Soltani and Fernando (2004) demonstrated that the safest worker path on a construction site can save nearly 30% of the activity cycle and process time. Furthermore, according to Koehn and Musser (1983), safety regulations can reduce construction costs from 2.8 percent to 1.4 percent. However, previous studies did not address the influence of safety events (accidents) in traditional construction planning methods, as Carmo and Sotelino (2023) pointed out. The scientific motivation is, thus, to determine whether planners are ignoring the safety issue in their planning methods and possibly carrying out unrealistic construction schedules.

Given this context, the current work proposes a new construction planning method for considering uncertainties related to safety events, with the goal of improving decision-making in terms of the main project indicator: time. The proposed method is unique in that it employs Quasi Birth and Death Processes (QBDP), which are based on the fundamental stochastic process theory (KARLIN; MCGREGOR, 1955; KENDALL, 1948) to capture random events associated with a safety occurrence. Furthermore, the proposed method considers the use of artificial neural network techniques to deal with the limitations of the QBDP during a real-time construction plan.

## 1.2. Objectives

The primary goal of this study is to create a stochastic model for accounting for the uncertainties associated with safety and delay events in energy infrastructure construction projects. This model is also used in conjunction with Artificial Intelligence (AI) techniques for bivariate time-series forecasting.

The specific goals are as follows: (1) propose a Markov transition diagram that represents the evolution of safety and delay events in a construction project; (2) use computational methods to calculate the stationary and transition probabilities based on stochastic process theory; (3) apply the Knowledge Discovery in Database (KDD) process to construction datasets that are typically unstructured and scarce; (4) investigate the effect of the safety random variable on the delay random variable using real-world project datasets; and (5) employ a hybrid solution integrating the stochastic model with AI techniques to overcome the shortcomings of complex mathematical models derived from the stochastic theory.

## 1.3. Organization

This thesis' organizational structure adheres to the manuscript format, in which standard thesis chapters are replaced by manuscripts that have been published or submitted for publication in peer-reviewed international journals. Chapter 2 provides a literature review on construction safety management. Chapter 3 presents the fundamental concepts related to the stochastic processes. Chapter 4 is the manuscript titled "Planning for the unexpected in construction projects: a review" that was published in a peer-reviewed international journal, which presents a literature review related to construction planning methods considering uncertainties. This chapter discusses the current state of the art in terms of the uncertainties that are typically considered in construction planning methods and, thus, highlight the importance of the current study. Chapter 5 consists of the manuscript entitled "A stochastic pure birth model for predicting and help prevent accidents in energy infrastructure construction projects" which has been submitted to the Safety Science Journal. This chapter proposes a new method for predicting safety occurrences, based on the pure-birth process, which derives from the birth and death processes, both stochastic processes that can generate the QBDP when combined. Therefore, Chapter 5 addresses the objective (3) and partially (1), (2), and (4). Chapter 6 presents the manuscript "A quasi birth and death process to understand the effects of safety occurrences into construction project delays" to be submitted to Safety Science Journal. This chapter presents a new construction planning method based on a hybrid solution that combines QBDP and neural network modeling, with a focus on the influence of safety events on delay events. Thus, Chapter 6 addresses the research objectives (1), (2), (4), and (5). Chapter 7 describes the practical application of the proposed method in the pre-construction and construction phases of an energy infrastructure construction project, which will be submitted as case study in Safety Science Journal. This chapter shows how to apply the proposed planning method in real-world projects, following the objective (4), through a simulation and by comparing with traditional approach. Finally, Chapter 8 provides a summary of the main conclusions, limitations, and future work suggestions.

## 1.4. Scientific categorization

In terms of the classification of scientific research, this study employed the inductive method, with a focus on applied research. Therefore, this study aimed to derive generalizations from specific observations and concentrated on addressing practical issues in the realm of construction safety management.

Also, the research objectives were exploratory in nature, seeking an in-depth understanding and the identification of new perspectives in construction safety management. To meet these objectives, a qualitative-quantitative data treatment approach was adopted, enabling a comprehensive analysis of the collected information.

Related to the data sampling, the current study adopted a probabilistic approach to ensure data representativeness and statistical validity. The research strategy incorporated diverse methods, including a literature review for theoretical grounding, experimental research for the practical evaluation of proposed interventions, and a case study for an in-depth understanding of specific contexts.

Finally, in relation to data collection, this work was comprehensive, utilizing simulation for controlled scenarios and participant observation to capture nuances and complexities in real construction environments. This multifaceted approach facilitated a holistic analysis of aspects related to construction safety, providing significant insights for the formulation of conclusions and recommendations.

# 2
# State-of-art-review of construction safety management

This chapter presents a literature review related to construction safety management. Due to the relevance of such topic in the last decade, many review articles were found in a brief search in the Scopus database, which is a well-known scientific database for studies related to construction management. Therefore, the current study adopted the snowballing technique to cover studies that are related to the reference ones.

The snowballing analysis involves the researcher using a reference list or citations to well-known articles on the subject (WOHLIN, 2014). So, it is essentially a thorough dive into a certain papers' references. This approach focuses on the key articles in a given subject, and with these papers, the researchers use inclusion and exclusion criteria to filter the reference list, and then begin the analysis, summary, results, and reports similarly to the Systematic Literature Review (SLR). Snowballing can be done in two ways: forward or backward. The references in the important publications are used in the former, while the citation to the key papers is used in the latter.

According to Jalali and Wohlin (2012), the snowballing analysis is simple to understand and replicate, as opposed to the traditional SLR, which contains more difficult procedures for a rookie researcher. However, as a disadvantage, "the lack of randomized: representativeness" in the snowballing study can result in biased conclusions, according to Geissdoerfer et al. (2017).

The key paper selected in the current study was developed by Zhou et al. (2015) and to support the snowballing analysis, the "*connectpapers*" platform was used to map the citations (prior works) and the papers that cited the key article (derivative works). The graph presented in Figure 1 shows in the center the key paper (ZHOU; GOH; LI, 2015) with a purple border and related works are around it. The circles are organized as follows: the darker the green circle, the more recent the article; the closer the circles, more similarity between the works; the bigger the circle, more citations; the thicker the line, more connection between the works.

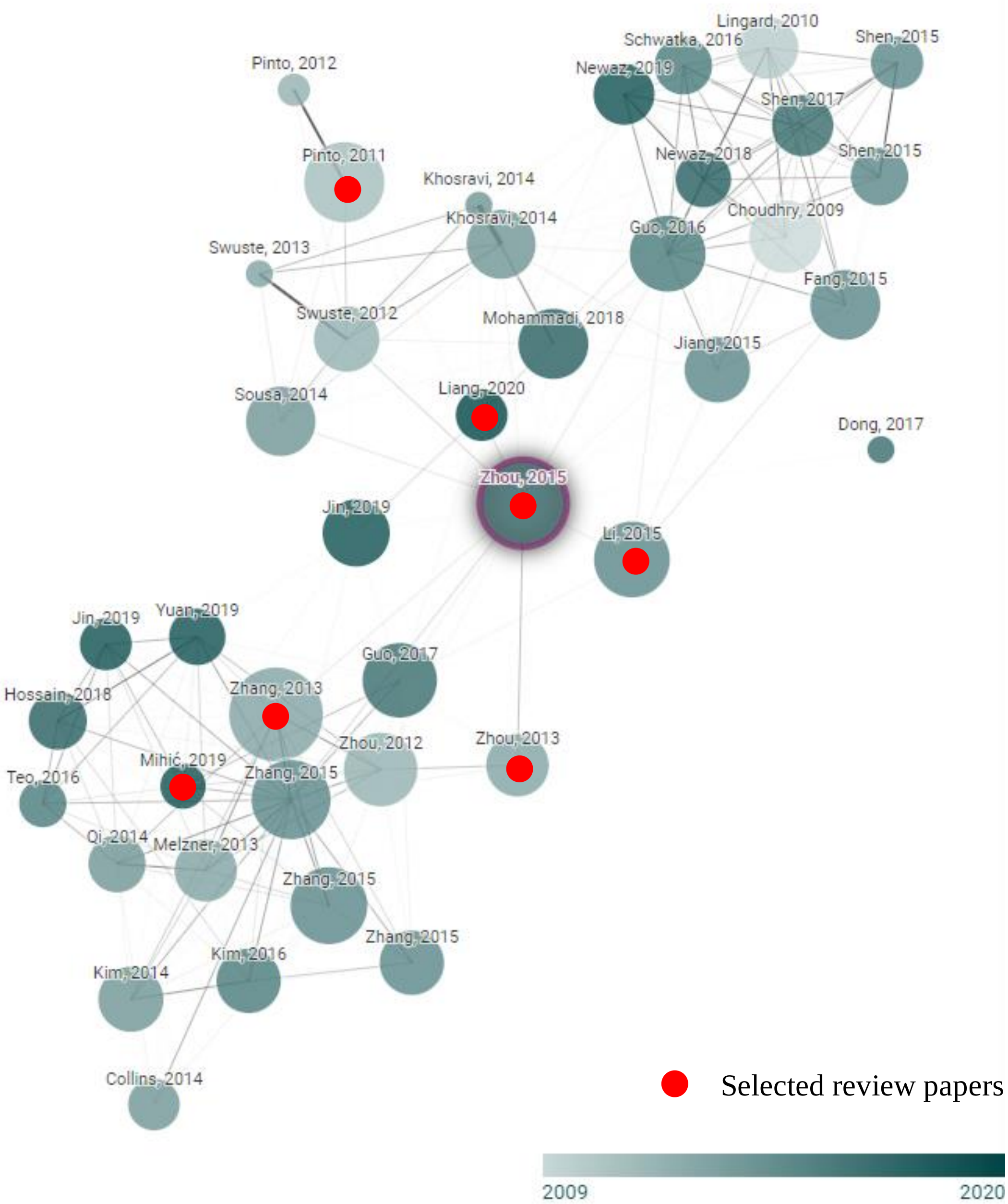



Figure 1 - Selected review papers during the snowballing analysis. (Adapted from *connectedpapers*[1])

## 2.1. Research fields

There are two main research fields studying the safety management in the construction industry: one related to management issues, normally associated with

[1] Retrieved November 14, 2023, from https://www.connectedpapers.com/main/7df7f2d87a7db9ec5d490f05f01ababaea97461a/graph?utm_source=share_popup&utm_medium=copy_link&utm_campaign=share_graph

cognitive studies (e.g., safety culture and human perceptions); and the other related to technological issues, usually associated with robotics and automation (ZHOU; GOH; LI, 2015; ZHOU; IRIZARRY; LI, 2013). Beyond the research fields, Zhou et al. (2015) defined three more frequent topics covered by the literature: the first one involving safety management process, the second one involving the influence of individual and group aspects on the safety management process, and the third one related to accident and incident data. With a different approach, Liang et al. (2020) defined seven other frequent topics, namely: “safety-specific industry practices”, “safety strategies and outcomes”, “accident statistics and analysis”, “behavior-driven management”, “technology-driven management”, “risk identification and assessment”, and “design for safety”. These topics can be related in some manner and together they represent an overview of the main research areas developed so far (see Figure 2).

As shown in Figure 2, the number of research areas increase over time in derivative fields. For instance, the fields of “strategies and outcomes” and “risk identification and assessment” described by Liang et al. (2020) are derived from the “safety management process” area described by Zhou et al., (2015), which was derived from the research field “safety management issues” defined by Zhou et al., (2013).

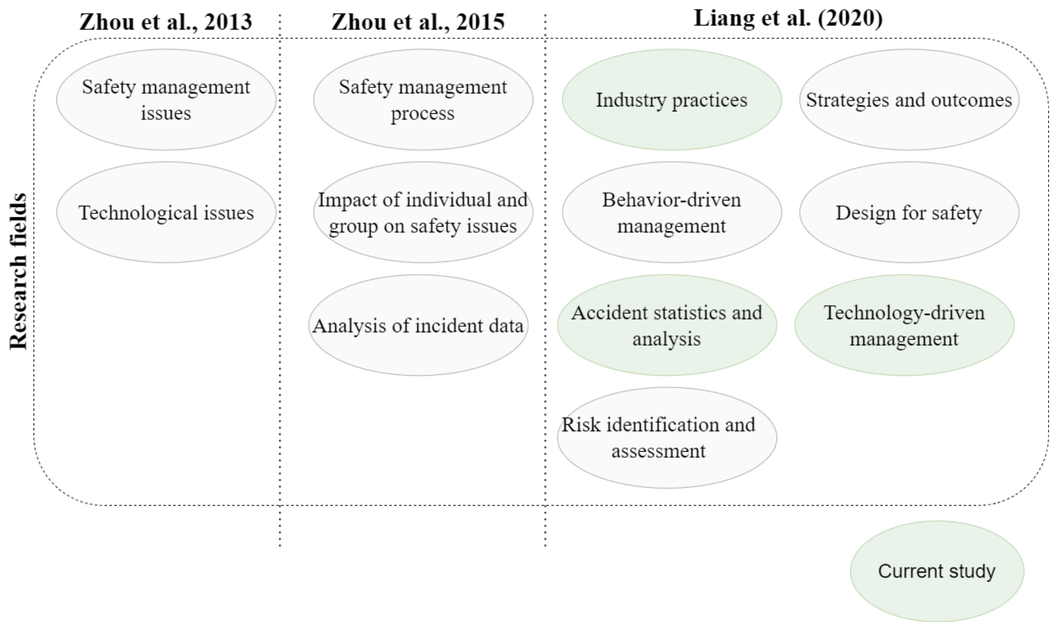



Figure 2 - Identified research fields in the existing literature.

The current study focuses on the areas of “accident statistics and analysis” by using the safety records to create statistic, stochastic and machine learning models;

“technology-driven management” by applying sophisticated technological methods in the safety and construction planning methods; and “industry practices” by analyzing the traditional construction planning method used in energy infrastructure projects. It should be noted that this type of project was not found in the reviews carried out by Zhou et al. (2013) and Zhou et al. (2015). The present study overcomes this limitation by carrying out such a review related specifically to renewable power plants.

It is also important to understand that Zhou et al. (2013) and Zhou et al. (2015) observed that the most publications were from developed countries. Developing countries, like Brazil, had almost no publication in this research area – 1% of the total located studies related to safety management in the construction industry. Therefore, the present paper fills partially this scientific gap, mainly with real data from construction projects developed in developing countries.

## 2.2.
## Research gaps identified in the literature

One of the most important objectives of carrying out a literature review is the finding of research gaps. Through them, new studies are oriented to fill these gaps even if partially. The following paragraphs describe the main research gaps identified in the snowballing analysis.

Zhou et al. (2013) reviewed the applied technologies in safety management and concluded that more studies should be done to consider the total project life cycle, to analyze the cost impact related to the implementation of new technologies, to consider the risk associated with the technology, to develop practical applications of technology in construction safety, and to assess legal aspects that can be affected by the applied technology. In fact, Liang et al. (2020) verified that in the period between 2011 and 2016, innovative technologies were frequently used in studies to improve on-site safety management.

Similarly with the review carried out by Zhou et al. (2013), Mihic et al. (2019) reviewed the applied technologies in construction health and safety research area. They reported that more studies are needed to develop a universal approach aiming to identify hazards in construction projects, not only in the construction phase and not limited to building projects, but also with the preconstruction phases and for infrastructure projects. In somehow, Mihic et al. (2019) reinforced the research gaps described by Zhou et al. (2015), which means that studies are still important and necessary in nowadays.

Zhou et al. (2015) reviewed studies related to safety management in the construction industry and concluded that more research should be done to monitor unsafe behavior, to understand how to predict accidents applying safety climate (normative ideals, beliefs, and behaviors), to define the organization's impact in the safety performance during the construction phase, to investigate the construction activities providing better solutions for safe tasks, to study projects that are not related with buildings, such as infrastructure projects, and to bring the academic innovative and technological methods into practical cases. However, Liang et al. (2020) emphasized that the sample reviewed by Zhou et al. (2015) is limited to 10 publication sources and, thus, they provided a more complete sample using bibliometric analysis.

In terms of construction safety management, the literature review carried out by Liang et al. (2020) concluded that more studies should be carried out to consider the entire project lifecycle and not only design and construction phases. Also, Liang et al. (2020) suggested the adoption of innovative technologies with actual data and a better integration of the safety behaviors in the construction planning approaches, like the conclusions of Zhou et al. (2013) and Zhou et al. (2015). In fact, they concluded that construction planning methods should consider the safety behaviors at a workgroup level (e.g., social relationships and the safety climate), when usually is considered at an organizational level (e.g., safety trainings and rewards for good safety performance). It is important to point out that this review is limited to articles published until 2016.

In fact, other studies have already reinforced these research gaps: Goh and Askar (2016) developed a planning method to model construction activities considering safety issues, but they emphasize that their framework should be improved with actual data; and Zhou et al. (2015) showed that most studies in the literature are related only to the construction phase.

The current study seeks to address some of the following research gaps:

- To study projects that are not related with buildings, such as infrastructure projects.

- To put academic innovative and technological methods related to construction safety management into practice.

## 2.3. Emerging topics and trends

It should be noted that the studies involving safety management in the construction industry are relatively recent, as pointed by Zhou et al. (2015). According to their literature review, the number of publications started to increase in 2002. In contrast, Liang et al. (2020) showed that the publications had in fact only increased from 2009. Related to the use of technologies in the safety management, the topic is even more recent in the construction industry, as stated by Zhou et al. (2013). Only after 2008, similarly with the results pointed out by Liang et al. (2020), this topic started to call researchers attention.

The present study reproduced the same term search strategy of the studies developed by Zhou et al. (2015) in the Scopus database – search terms “construction” and “safety” must appear in the article title and only papers written in English and peer-reviewed were selected. The results indicate another peak of publications related to construction safety management in 2020 (see Figure 3). This rapid growth can be caused by many reasons, like the COVID-19 pandemic, which imposed restrictions to construction projects. In fact, “COVID” term appeared as research topic in 30 publications between 2020 and 2023.

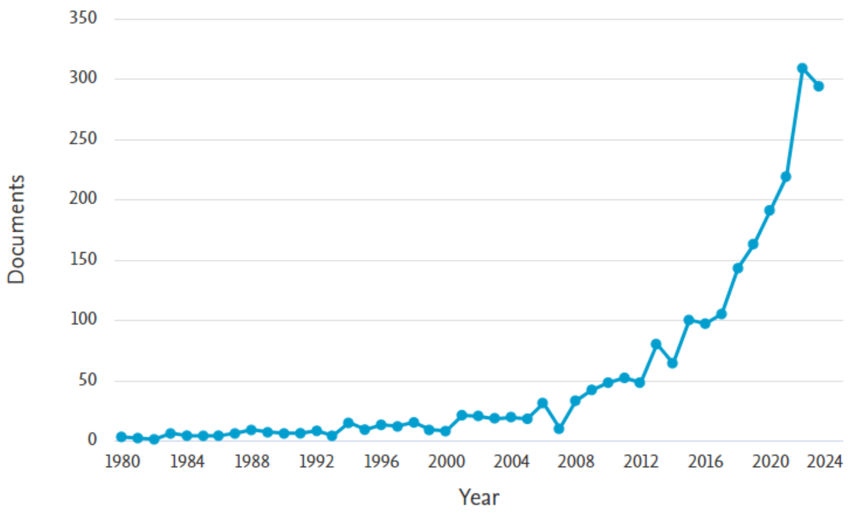



Figure 3 - Number of publications related to construction safety throughout the years. (Extracted from *Scopus* database[2])

[2] Retrieved November 14, 2023, from https://www.scopus.com/term/analyzer.uri?...

Moreover, Liang et al. (2020) noted that some research trends appeared in the literature, such as the use of a variety of research topics (such as sustainability issues and social network analysis), innovation technologies, and considerations about safety behavior issues in the construction safety management. In fact, in the research field of accident analysis, Huang et al. (2022) defined four trends that are somehow related to the three previously mentioned, which are: “daily accident prevention”, “model-based research”, “system analysis and accident prediction”, and “occupational safety and public health research”. Specifically in the research field related to accident prevention, according to Huang et al. (2022), the last decade, from 2011 to 2021, was characterized by a rapid development stage of publications.

As a result, the current study was carried out in the research trends related to innovation technologies using model-based systems for accident prevention and prediction.

## 2.4. Technologies and approaches applied in safety construction management

One of the common technologies applied in safety management, according to Zhou et al. (2013), are the mathematical models that can handle large dataset and predict safety occurrences. Between 2005 and 2010, Liang et al. (2020) highlighted that mathematical models were frequently used to assess safety risk. Yet, Zhou et al. (2015) highlighted the use of sensors, virtual reality, computer-aided design, and 4D technologies. Li et al. (2016) stated that many safety models were developed with a mathematical background.

Some research works use data mining techniques, for example Rivas et al. (2011) used Bayesian networks, classification trees, and decision rules techniques to predict workplace accident. Their methodology was based on feature selection, cross validation, and data interpretation to understand the causes of safety occurrences.

The real-time strategy adopted by Li et al. (2016) is based on construction accessories, such as smart helmets, that track the location of each worker. However, to complement the literature in the context of developing countries, as suggested by Zhou et al. (2013) and Zhou et al. (2015), this technology may be not available or poorly used in the construction site. Therefore, this paper adopted a hybrid approach, using safety records issued weekly to produce a quasi-live planning method, eliminating the need for technological gadgets.

Other studies (LI, Heng et al., 2016) do not detail the preprocessing stages, which can hinder the use of these planning methods by other studies. Therefore, the present study adopted the Knowledge Discovery in Databases (KDD) process detailing all steps until the new knowledge discovery. This helps in providing insights mainly related to relations and conditions inside the dataset, improving the overall KDD process (AMARAL; BAIÃO; GUIZZARDI, 2021).

Though the work developed by Li et al. (2016) did not mention foundational ontologies, their understanding and description of hazard regions and safety states contain many relations and connections between data. This paper used the ontologies to formalize the relations between data. In fact, as reported by Mihic et al. (2019), the combination of ontology and natural language processing is an innovative technology that has been appearing the literature related to construction health and safety.

Related to the adopted approach in the literature in terms of safety construction management, Zhou et al. (2013) highlighted that past studies were more oriented to reactive approaches, such as cause analysis, and suggested that more studies are necessary to develop proactive safety planning methods. In fact, some studies have already done some efforts in that direction. Li et al. (2016) proposed a live construction planning using real time location systems and Markovian stochastic process. Teizer et al. (2010) proposed a proactive safety management system using radio frequency technology and tested it in real construction cases with focus on equipment flow.

According to a state-of-the-art related to occupational risk assessment presented by Pinto et al. (2011), there are three groups of traditional methods: deterministic, probabilistic, and hybrid. In fact, the authors discussed that the usual uncertainties in the available information for the risk assessment methods are not handled by the traditional probabilistic methods. Also, Marhavilas et al. (2013) stated that there are two groups of quantitative accident forecasting: one based on time-series, and another based on causality models. The current work focuses on the first approach.

According to the literature review, probabilistic solutions may be ineffective when dealing with on-site construction safety data. To deal with the high level of uncertainty, however, hybrid solutions that leverage traditional developments while incorporating innovative technologies as enhancer techniques are required. In the present work stochastic processes are adopted. Their basic theory is described in the following chapter.

In conclusion, and in accordance with the literature, the current study proposes a mathematical model combined with AI techniques to account for both safety and delay events in the construction planning method. To treat and process the actual dataset, the proposal includes stochastic process theory, data mining techniques, and neural networks modeling.

# 3
# Theoretical background

## 3.1.
## Stochastic processes

A stochastic process is a family of functions that, using a parameter, map each instance of a sample space to a new domain. This map of functions, also known as a random process, is usually related to a time parameter, which can be discrete or continuous and, in turn, depends on how the experiment's observations are made. Because the observed instances can be discrete or continuous, stochastic processes with continuous or discrete parameters are defined (ALBUQUERQUE, 2017). When observing work accidents in construction sites, for example, the instance (accident) is discrete because it will always be in unit increments, but the time observed is continuous, the time difference between accidents does not follow a standard increment, and an instance can occur at any time instant.

Figure 4 depicts two examples of stochastic processes – $n(x_1, t)$ and $n(x_2, t)$ – found in the literature. The first is a discrete stochastic process with a continuous parameter that refers to the process of birth and death of any animal species ($x_1$). The second example is the number of calls received at a telemarketing center. This process, like the first, is discrete with a continuous parameter, but the amplitude is always non-negative.

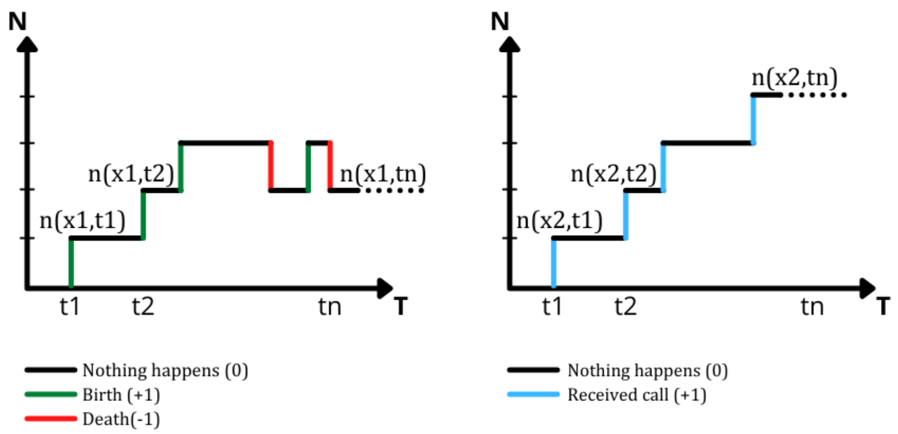



Figure 4 - Examples of stochastic processes in the existing literature.

### 3.1.1. Birth and death process

Kendall (1948) defines the birth and death process as a stochastic process $n(t)$ with non-negative integer values ($S_n$ states) and birth and death rates as a function of time, i.e., $\lambda(t)$ and $\mu\ (t)$. Furthermore, Karlin and McGregor (1955) define the birth and death process as a random walk process with a continuous or discrete time parameter.

The model's possible states and transition probabilities are listed in Table 1 and illustrated in the states diagram in Figure 5. The diagram can be used to identify possible transitions between states represented by arrows with birth or death rates. The arrows indicate a state advance when directed to the right, a state retreat when directed to the left, and a state permanence when the origin and destination of the arrow are in the same state (circles). It should be noted that the diagram depicts state transitions at time $t$, i.e., all possible state changes when time equals a certain value.

Table 1 - Events, states, and probabilities that describe the traditional birth and death process.

| Events | States | Probabilities |
|---|---|---|
| **Birth** | $n(t+\delta) = n(t) + 1$ | $p_{ij\vert j=i+1}(dt) = \lambda_i(t)dt + o(dt)$ |
| **Death** | $n(t+\delta) = n(t) - 1$ | $p_{ij\vert j=i-1}(dt) = \mu_i(t)dt + o(dt)$ |
| **Nothing** | $n(t+\delta) = n(t)$ | $p_{ij\vert j=i}(dt) = 1 - \{\lambda_i(t) + \mu_i(t)\}dt + o(dt)$ |

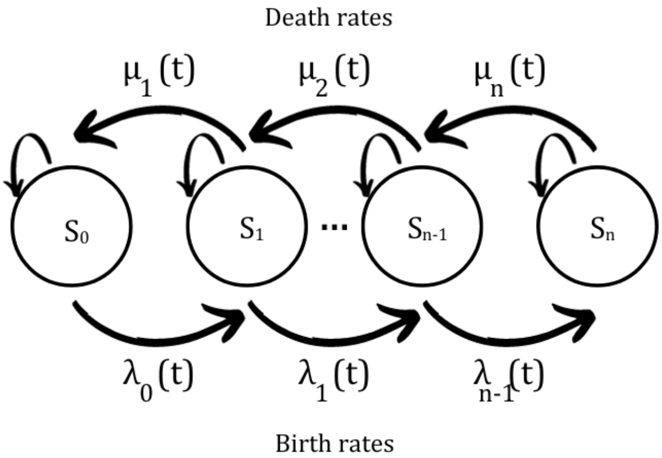



Figure 5 - States diagram that represents a typical birth and death process.

Where,

- $\delta$ is any time increment;
- $p_{ij}(dt)$ is the probability of transitioning from state $i$ to state $j$ $(i, j = 0,1,2, \ldots)$, given the increment $dt$;
- $o(dt)$ is a function that meets the following criteria $\lim_{dt \to 0} \frac{0(dt)}{dt} = 0$.

It has also been established that moving backwards from a state equal to zero or forward from states prior to zero and the last state $S_n$ is not permitted. These assumptions, that are listed below, are critical in order to keep the process non-negative, implying that it is never possible for a construction project to have negative duration.

Assumptions:

- $\mu_0(t) = 0$, which means that it is impossible to die from the number 0;
- $\lambda_{-1}(t) = 0$, indicating that a birth cannot occur from the quantity $-1$;
- $\lambda_{n+1}(t) = 0$, indicating that a birth cannot occur from the quantity $n + 1$.

With this in mind, some generic states can be used to develop the mathematical expressions that result in the transition probabilities based on the formal definition of the birth and death process. The population size is then assumed to be equivalent to a value $a$ at time $t_0 = 0$ and to a value $b$ at time $t_1$, where $t_1 > t_0$. As a result, given a time $t$ and an increment $dt$, the transition probability $(a \to b)$, taking into account the continuous time parameter and unit transitions (unit increments), can be calculated from the state probabilities described in Table 1 and expressed by the differential equation:

$$p_{ab}(t + dt) = \lambda_{b-1}(t)p_{a,b-1}(t)dt + \{1 - [\lambda_b(t) + \mu_b(t)]\}p_{ab}(t)dt + \mu_{b+1}(t)p_{a,b+1}(t)dt + o(dt) \quad (1)$$

In relation to the current study's scope, the first part of the right-hand side of the expression represents progress towards state $b$ – a work delay; the second part represents remaining in state $b$; and the third part represents regression to state $b + 1$ – a potential schedule advance. It can then be seen that if $a = b$, the probability $p_{ab}(t = 0)$ will be equal to 1, and that subtracting from both sides $p_{ab}(t)$ and then deriving both sides by $t$ results in the following expression:

$$\frac{\partial p_{ab}(t)}{\partial t} = \lambda_{b-1}(t)p_{a,b-1}(t) - [\lambda_b(t) + \mu_b(t)]p_{ab}(t) + \mu_{b+1}(t)p_{a,b+1}(t) \tag{2}$$

Equation 2 represents how the transition probabilities change over time and is also known as the Kolmogorov forward equations. However, it is critical to understand that the stochastic process in this work is stationary in the strict sense, which means that the probability density function of any order does not vary over time for the state probabilities. This assumption is critical for the mathematical formulations presented in this work, but additional manipulations and definitions can be found in Albuquerque et al. (2008).

Also shown in the format of Equation 2 is the transition probability matrix (KARLIN; MCGREGOR, 1955), which represents all transition probabilities in matrix format.

$$\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{A} \tag{3}$$

Where,

- The generator matrix is represented by $\boldsymbol{A}$, which has the following positions:

$$\mathbf{A} = \begin{bmatrix} -\lambda_0(t) & \lambda_0(t) & \cdots & 0 \quad 0 & \\ \mu_1(t) & -[\lambda_1(t) + \mu_1(t)] & & 0 \quad 0 & \\ & \vdots & \ddots & \vdots & \\ & 0 \quad 0 & \cdots & -[\lambda_{n-1}(t) + \mu_{n-1}(t)] & \lambda_{n-1}(t) \\ & 0 \quad 0 & & \mu_n(t) & -[\lambda_n(t) + \mu_n(t)] \end{bmatrix} \tag{4}$$

This tridiagonal, non-negative matrix satisfies the following properties and conditions:

I. $\boldsymbol{P}(0) = \boldsymbol{I}$

II. $p_{ij}(t) \geq 0$

III. $\sum_{j=0}^{\infty} p_{ij}(t) \leq 1$

IV. $\boldsymbol{P}(t+s) = \boldsymbol{P}(t)\boldsymbol{P}(s)$

### 3.1.2. Numerical example

As an example, consider the following construction site scenario: delay states range from 0 to 34 days and are observed in random order with unit increments over time (for example, from 27 to 26 at time $t_1$ and 26 to 27 at time $t_2$). The number of delays in this hypothetical example is always greater than the number of schedule advances. In fact, in Brazil, approximately 69% of public works do not meet the contractual deadline set for the construction phase (ALVARENGA et al., 2021).

As a result, in the middle of the construction phase ($t_{normalized} = 0.5$), the state diagram will show the delay and advance rates shown in Figure 6.

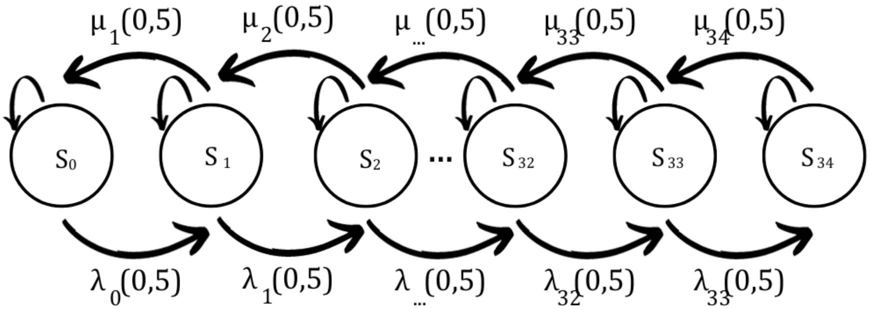



Figure 6 - States diagram that represents a hypothetical example related to a construction project.

Thus, at the normalized time of analysis ($t_{norm} = 0.5$), the following matrix expression can be used to calculate the transition probabilities given an increment $dt$, following Equation 3.

$$\boldsymbol{P}'(t)$$

$$= \begin{bmatrix} p_{0,0}(0,5) & p_{0,1}(0,5) & \cdots & 0 & 0 \\ p_{1,0}(0,5) & p_{1,1}(0,5) & & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \cdots & p_{33,33}(0,5) & p_{33,34}(0,5) \\ 0 & 0 & & p_{34,33}(0,5) & p_{34,34}(0,5) \end{bmatrix}$$

$$\times \begin{bmatrix} -\lambda_0(0{,}5) & \lambda_0(0{,}5) & \cdots & 0 & 0 \\ \mu_1(0{,}5) & -[\lambda_1(0{,}5)+\mu_1(0{,}5)] & & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \cdots & -[\lambda_{33}(0{,}5)+\mu_{33}(0{,}5)] & \lambda_{33}(0{,}5) \\ 0 & 0 & & \mu_{34}(0{,}5) & -[\lambda_{34}(0{,}5)+\mu_{34}(0{,}5)] \end{bmatrix}$$

$$= \begin{bmatrix} -\lambda_0(0{,}5)p_{0,0}(0,5) + \mu_1(0{,}5)p_{0,1}(0,5) & \lambda_0(0{,}5)p_{0,0}(0,5) - [\lambda_1(0{,}5)+\mu_1(0{,}5)]p_{0,1}(0,5) & \cdots \\ -\lambda_0(0{,}5)p_{1,0}(0,5) + \mu_1(0{,}5)p_{1,1}(0,5) & \lambda_0(0{,}5)p_{1,0}(0,5) - [\lambda_1(0{,}5)+\mu_1(0{,}5)]p_{1,1}(0,5) + \mu_2(0{,}5)p_{1,2}(0,5) & \\ & \vdots & \ddots \end{bmatrix}$$

It is possible to infer the most common transitions at a given time $t$ using this matrix and, thus, establish the most critical states associated with these transitions. For example, if the delay rate $\lambda_3(t)$ is equal to 0.9, it means that when the number of delays accumulates to 3 units (days, weeks, or any other measure), there is a 90% chance of moving to state 4 of delays at time $t$ of the work. If, on the other hand, $\lambda_{34}(t)$ equals 0.05, the probability of a transition to a state one unit higher is low.

### 3.1.3. Markov chain

This matrix expression can also be used to understand the Markov property, which is an important feature of the stochastic processes used in this study. A Markov chain occurs when the probability of an event occurring depends only on the recent past and not on the distant past (BOLCH et al., 2006). In other words, a Markov process has the property of being memoryless, which can be expressed mathematically as:

$$\begin{aligned} p\big(X_{t_{n+1}} \leq s_{n+1} \,\big|\, X_{t_n} = s_n, X_{t_{n-1}} = s_{n-1}, \dots, X_{t_0} = s_0\big) \\ = p\big(X_{t_{n+1}} \leq s_{n+1} \,\big|\, X_{t_n} = s_n\big) \end{aligned} \quad (5)$$

This property is easily observed in the numerical and hypothetical example of delays used by the matrix format, in which the transition probabilities depend only on the immediately preceding and following states. The same holds true in the case of construction site accidents. The process is Markovian because the probability of an accident occurring at the start of the project has no statistically significant effect on the probability of accidents occurring at the end of the project. In other words, once unitary and independent increments are considered, the study of the influence of accidents can be done locally.

Still on the subject of Markov characteristics, the semigroup property can be defined using property IV and the Chapman-Kolmogoroff equation (KARLIN; MCGREGOR, 1955):

$$p_{ij}(t+s) = \sum_{k=0}^{\infty} p_{ik}(t) p_{kj}(s) \quad (6)$$

Because of the semigroup property, if one knows the transition probabilities from state $i$ to $k$ and $k$ to $j$ at times $t$ and $s$, one also knows the transition probability from $i$ to $j$ at time $t + s$. Thus, using the semigroup property, it is possible to estimate transition probabilities in the future based on past records, or to create a predictive accident model as long as the history of previous works is available.

### 3.1.4. Pure birth process

The study of occupational safety occurrences can be extracted from the birth and death model described for the study of delays by using the pure birth model. In the pure birth process, the mortality rate $\mu(t)$ is zero for any state (see Figure 7). In this study, this model is used to represent construction safety occurrences, where the birth represents a safety occurrence, which can be an accident, a near miss event (when an accident almost occurs), or a safety observation (when a near miss event almost occurs), i.e., construction events that cause or nearly cause work accidents.

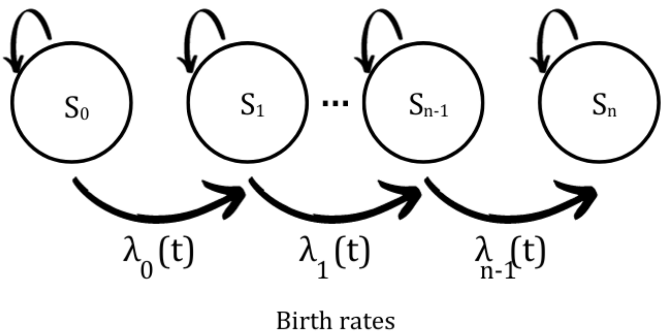



Figure 7 - States diagram that represents a typical pure birth process.

The following additional assumptions are made in this model:

V. $n(0) = 0$: the starting point will always be zero births (accidents);

VI. $n(t_2) - n(t_1)$ : the number of occurrences in the interval $]t_1, t_2]\ \forall\ 0 \leq t_1 < t_2$;

VII. $n(t)$ has independent increments, which means that an increase in the number of accidents in $(t_2 - t_1)$ is unrelated to another in $(t_4 - t_3)$;

VIII. $n(t_2) - n(t_1)$ : a Poisson random variable (or vector) with parameter $\lambda(t_2 - t_1)$, where $\lambda$ is the birth rate equivalent to the occurrence rate.

The transition matrix and state probabilities for this model can be easily calculated by zeroing the mortality rate in Equation 1, Equation 2, and Equation 4, resulting in:

$$p_{ab}(t+\partial t) = \lambda_{b-1}(t)p_{a,b-1}(t)\partial t + [1-\lambda_b(t)]p_{ab}(t)\partial t + o(\partial t) \tag{7}$$

$$\frac{\partial p_{ab}(t)}{\partial t} = \lambda_{b-1}(t)p_{a,b-1}(t) - \lambda_b(t)p_{ab}(t) \tag{8}$$

$$\mathbf{A} = \begin{bmatrix} -\lambda_0(t) & \lambda_0(t) & \cdots & 0 & 0 \\ 0 & -\lambda_1(t) & & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ 0 \quad 0 & & \cdots & -\lambda_{n-1}(t) & \lambda_{n-1}(t) \\ 0 \quad 0 & & & 0 & -\lambda_n(t) \end{bmatrix} \tag{9}$$

### 3.1.5.
### Poisson model

Previous research (CHUA; GOH, 2005) indicates that the distribution of accidents can be described by the Poisson model and, thus, the equations of the pure birth model can be simplified even further based on the following condition that indicates the constancy of the birth rate:

$\lambda(t) = \lambda$, at any time $t$.

Thus, the pure birth model's description of the accident evolution problem can also be understood as a Poisson stochastic model with the following probability distribution:

$$p_{n(t)}(N) = \sum_{m=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^m}{m!}\delta(N-m) \mid \delta(N-m) = \begin{cases} 1, if\ m = 0,1,2,\dots \\ 0,\ otherwise \end{cases} \tag{10}$$

Where,

$\delta(N-m)$ is the Dirac delta.

Since the increments are statistically independent, the joint probability density function (PDF) between them is the multiplication of the PDFs of the increments. This means that the probability of $N_2$ safety occurrences between times $t_3$ and $t_4$ given $N_1$ occurrences between times $t_1$ and $t_2$ can be calculated using the individual probabilities of each increment, i.e.:

$$P_{i_1,i_2,\dots,i_m}(N_1, N_2, \dots, N_m) = P_{i_1}(N_1)P_{i_2}(N_2)\dots\ P_{i_m}(N_m) \tag{11}$$

Where,

- $i_1 = n(t_2) - n(\ t_1), i_2 = n(t_4) - n(t_3), \dots, i_m = n(t_{2m}) - n(t_{2m-1});$
- $0 \le t_1 < t_2 < t_3 < t_4 < \cdots < t_{2m-1} < t_{2m}$ .

Also,

$$\begin{aligned} p_{n(t_2)-n(t_1)}(N) &= \sum_{m=0}^{\infty} \frac{e^{-\lambda(t_2-t_1)}(\lambda(t_2-t_1))^m}{m!}\delta(N-m) \\ &= p_{n(t_2-t_1)}(N) \end{aligned} \tag{12}$$

And by doing,

$y(t) = n(t+h) - n(t)$

It leads to:

$p_{y(t)}(N) = p_{n(t+h)-n(t)}(N) = p_{n(h)}(N)$

$p_{y(t+c)}(N) = p_{n(t+h+c)-n(t+c)}(N) = p_{n(h)}(N)$

$$p_{y(t+c)}(N) = p_{y(t)}(N) \tag{13}$$

This means that the number of accidents $n(t)$ exhibits first order increments of stationarity, i.e., it is a stationary stochastic process in which the first order probability density function does not change as time passes. Furthermore, because the increments are independent, the stochastic process described is strictly stationary, as can be seen for any integer value of $m$. In other words, the PDFs do not change over time.

Next, an analysis of the time between incidents is presented to help understand the applicability of the stochastic process described above. The random variable under analysis, $t_1$, is assumed to represent the time of the first incident (accident, first aid, or safety observation). There were no incidents prior to $t_1$, i.e.: $n(t < t_1) = 0$.

Applying the cumulative probability function,

$$P(\{t < t_1\}) = P(n(t) = 0) = \frac{e^{-\lambda t}(\lambda t)^m}{m!}\bigg|_{m=0} = e^{-\lambda t}$$

Therefore,

$$P(\{t_1 \leq t\}) = 1 - e^{-\lambda t}$$

In other words, the cumulative density function is defined as:

$$F_{t_1}(t) = P(\{t_1 \leq t\}) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (14)$$

As a result, the time of the first occupational safety event on site follows an exponential distribution, as shown in Figure 8. Projects with a history of higher accident rates (higher $\lambda$) are more likely to have new accidents at the beginning of the project (lower $t$), as expected.

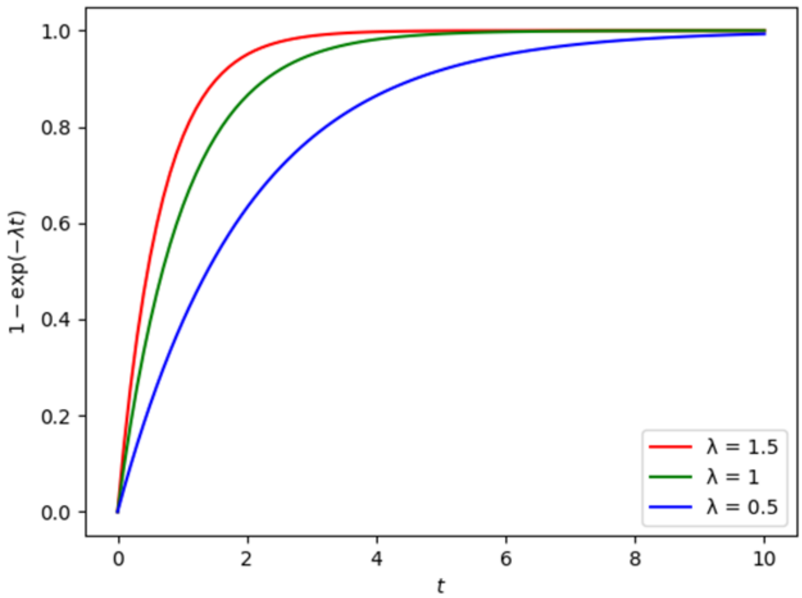



Figure 8 - Cumulative density function for the Poisson model with birth rate variation.

The same reasoning applies to an increment $\Delta t$:

$$F_{t_1}(\Delta t) = P(\{\Delta t_1 \leq \Delta t\}) = \begin{cases} 1 - e^{-\lambda \Delta t}, & \Delta t \geq 0 \\ 0, & \Delta t < 0 \end{cases} \tag{15}$$

In other words, any time interval between occurrences follows the exponential distribution in addition to being equally independent of each other, and because they are independent of each other, the probability density function can be obtained by multiplying exponential distributions.

$$F_{t_1,t_2,\ldots,t_m}(\Delta t_1, \Delta t_2, \ldots \Delta t_m) = P(\{\Delta t_1 \leq \Delta t\}).P(\{\Delta t_2 \leq \Delta t\}) \ldots P(\{\Delta t_m \leq \Delta t\})$$
$$= \begin{cases} \left(1 - e^{-\lambda \Delta t_1}\right)\left(1 - e^{-\lambda \Delta t_2}\right) \ldots \left(1 - e^{-\lambda \Delta t_m}\right), & \Delta t_1 \geq 0, \Delta t_2 \geq 0, \ldots \Delta t_m \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

## 3.2.
## Neural networks

Neural networks models are composed of Machine Learning (ML) algorithms, which are a kind of Artificial Intelligence (AI) technique. In fact, the concept of intelligence associated with computer machines emerged from “The Imitation Game” proposed by Turing (1950). Although he did not mention the term “Artificial Intelligence”, which was created by McCarthy et al. (1955), the Turing test reflects the origin of a learning machine. It is important to highlight, however, that ML is a subset of AI scientific area.

Thus, to “teach” a computer machine aiming to create an artificial intelligence, the neural networks is one of the most recognized models used. It is inspired by the human brain which is formed by neurons that communicate each other through impulses. Thus, NN models can solve complex tasks thanks to its powerful structure, which usually consists of an input layer, processing units and activation function (hidden layers), and output layer.

Following the artificial neuron proposed by Haykin (2009), a processing unit can be understood as a mathematical model, in which there are $n$ input nodes with respective weights and one output node. The input nodes are combined linearly by a sum function, considering the associated weights, and the result is used by an activation function. The sum operation includes a parameter, named bias, to increase or decrease the input value in the activation value. It should be noted that the activation function is responsible for limiting the output by a finite value.

Therefore, the activation function can be selected depending on the model application. Usually, nonlinear functions are used, such as hyperbolic tangent and sigmoid, to capture complex behaviors in the neural networks model that are not captured with linear functions (MÜLLER; GUIDO, 2016). Table 2 represents the hyperbolic tangent and sigmoid nonlinear functions used normally as activation functions:

Table 2 - Examples of nonlinear activation functions.

| Activation function | Formulation |
|---|---|
| hyperbolic tangent | $\varphi(v_j) = \frac{1}{1 + e^{-v_j}}$ |
| sigmoid | $\varphi(v_j) = \frac{e^{(v_j)} - e^{-(v_j)}}{e^{(v_j)} + e^{-(v_j)}}$ |

Where,

- $\varphi(v_j)$ represents the activation function according to the parameter bias $v$ at the neuron index $j$.

### 3.2.1. Architecture

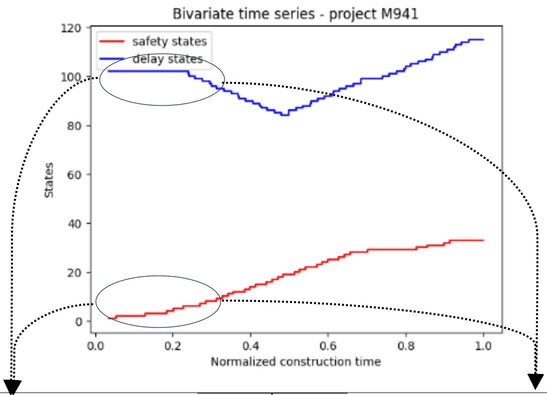
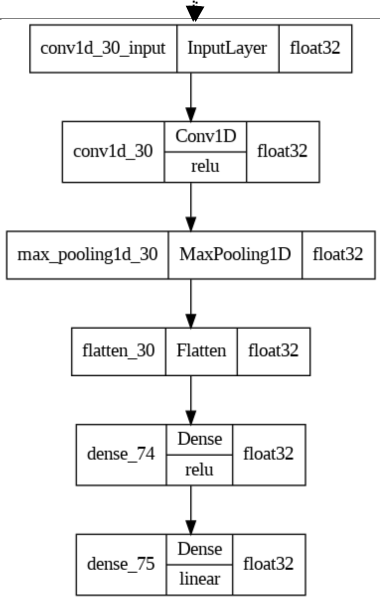
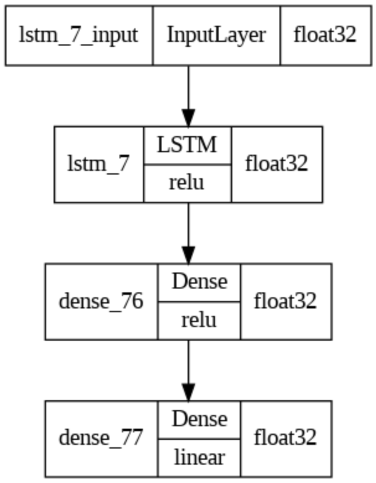
When combining the artificial neurons, a neural network is created. Thus, the architecture can be defined according to the connection between the neurons and are usually classified as feed-forward or recurrent. The former is defined when the neurons are connected and organized in layers and the information flow is in one direction – from the input layer to the output layer. The latter architecture occurs when the information flow is possible in two directions – from the input layer to the output layer (feed-forward) or vice versa (feed-backward)

The Multilayer Perceptron (MLP) is one example of neural networks architecture classified as feed-forward. In this architecture, there are one input layer, one or more hidden layers, and one output layer, in this sequence. The processing nucleus is in the hidden layers and so, the number of nodes in the hidden layer is an important parameter to be defined (MÜLLER; GUIDO, 2016).

The current study adopted two architectures: Convolutional (CNN) and Long Short-Term Memory (LSTM), as shown in Table 3. The CNN is a feed-forward architecture inspired by the structure of a visual cortex, in which the hidden layers are characterized by one or more layers that run convolutional operations. The

LSTM is however a recurrent architecture, that is useful to deal with time sequence and is formed by hidden layers that learn when remember or forget the information.

Table 3 - Adopted neural networks architectures.

<table>
<tr><th>CNN</th><th>LSTM</th></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

### 3.2.2. Stop criteria and performance metrics

According to Haykin (2009), although there are not clear convergence criteria to stop the backpropagation algorithms, some stop criteria can be highlighted. The first one refers to an absolute rate based on the variation in the mean squared error. For example, when the absolute rate per epoch is close 1%, it can be considered sufficiently low, and the algorithm stops. However, it could interrupt the routine prematurely during the learning process.

The second criteria mentioned by Haykin (2009) is based on the general network performance. The generalization refers to the capacity of a neural network in predicting the results using an input data that was not used in the training model. When the performance is considered satisfactory, the algorithm is therefore stopped.

The performance can be evaluated using metrics, such as the mean absolute error (MAE). In this study, three performance metrics were adopted: MAE, mean absolute percentage error (MAPE), and root-mean-square error (RMSE). The absolute and percentual errors were important to understand how close or far is the prediction for the actual values. But, when there are few data to train and test the model, it can return in large MAE and MAPE metrics, due to wrong estimation. The RMSE metric is calculated based on the square root of the average of squared errors and is sensitive to outliers. Table 4 shows the equations that defined the adopted metrics.

Table 4 – Adopted performance metrics for neural networks evaluation.

| **Performance metrics** | **Formulation** |
|---|---|
| MAE | $\frac{\sum_{i=1}^{n}\lvert error_i \rvert}{n}$ |
| MAPE | $\frac{\sum_{i=1}^{n}\left\lvert \frac{error_i}{actual\_value_i} \right\rvert}{n}$ |
| RMSE | $\sqrt{\frac{\sum_{i=1}^{n}(error_i)^2}{n}}$ |

# 4
# Planning for the unexpected in construction projects: a review

Paper published by Cristiano S. T. do Carmo and Elisa D. Sotelino, in peer-reviewed international journal[3].

Global crises, such as pandemic and wars, bring to light how construction projects can be impacted by unexpected events that are typically overlooked by planning teams. Therefore, the goal of this study is to review the literature to understand how uncertainties are being considered in construction planning methods and, what are the next steps to face new crises. By doing so, the authors mapped the traditional variables that are included as uncertainties in planning methods, such as project time and cost, as well as the unusual variables that are not typically included as uncertainties in the methods, such as safety and sustainability issues. The state-of-the-art of planning methods with uncertainties entailed a thorough reading of 103 journal articles found through an adapted systematic literature review, which included, in addition to traditional processes, a scientometric study and a snowballing analysis. As a result, it was discovered that the main uncertainties considered are related to time, cost, and resources. Furthermore, it was possible to observe that there is no single consolidated technique for incorporating uncertainties in planning methods, but rather a combination of different techniques, ranging from the most traditional with analytical analysis to the most contemporary with artificial intelligence algorithms.

## 4.1. Introduction

During the current global crisis caused by the COVID-19 pandemic and the Ukraine war, the construction industry turns on the lights to improve predictability and risk contingency in contracts. Commodities experienced rapid price fluctuations, prompting financial investors to shift their portfolios into low-risk

[3] DOI: 10.47456/bjpe.v9i4.42244

industries. However, as is well known, the construction industry is characterized by randomness and uncertainty, increasing the level of vulnerability to unforeseeable events, which is unappealing to investors.

Many factors contribute to the inherent uncertainties that arise during construction work. For example, Laufer and Cohenca (1990) conducted a survey and found that completion of the design phase, previous experiences, labor supply, weather conditions, and planner subjectivity are all factors that have a high impact on construction planning results. As a result, it is almost mandatory in today's turbulent times to consider uncertainties in construction planning methods to include the variability of that industry more explicitly and, thus, attract more investment.

Defining uncertainty is a difficult task. Uncertainty is the lack of knowledge about a situation in which one does not understand the values, possible ranges, or whether the outcome will be positive or negative (ZHENG; CARVALHO, 2016). Unlike risk, which is commonly associated with negative known scenarios (threats), uncertainties occur before risks (FENG et al., 2018) and can lead to threats or opportunities (ZHENG; CARVALHO, 2016). In other words, one only knows the risks if uncertainties are understood beforehand. While risk is equivalent to numerical variability, uncertainty can be associated with chaos and without any control over the probabilistic events (DE MEYER; LOCH; PICH, 2002).

Note that there are different levels of uncertainty. According to Walker et al. (2013), it is divided into five levels: the lowest level (1) refers to a single system model that guide to only one direction; and highest level (5) occurs when there is no known system model or even known outcomes. The latter is also known as deep uncertainty, as defined by Lempert et al. (2003), and is the one applicable to construction planning of unusual projects, such as infrastructure projects, according to Feng et al. (2022). This is because in these projects other contexts beyond the technical issues such as human, political, social, and environmental issues must also be considered, thus, increasing the difficulty in defining uncertainties and the choice of model to be used.

So, how do construction planners deal with uncertainties when planning? This is the central question for the current study, which is divided into five sections. Thus, the objective of the paper is to provide a comprehensive overview of the state of art on construction planning with uncertainties and to identify gaps of knowledge. The research context is presented in the first section and a theorical background is presented in the second one. The research methodology is explained in the third section. Following, the fourth section contains the main discussions on uncertainties in construction planning methods. The fifth section then summarizes the study's development as well as the main contributions and limitations.

## 4.2. Theorical background

### 4.2.1 Construction industry and its challenges with planning methods

According to the McKinsey report (Ribeirinho et al., 2020), the construction industry accounts for 13% of worldwide Gross domestic product (GDP), and according to the Brazilian statistical report (CBIC, 2023), an average of 5.3% in Brazil. Behind these figures are societal effects such as job creation and income generating. Furthermore, the building business has been hit hard by worldwide crises such as the COVID-19 pandemic. According to the Brazilian Statistics Institute (IBGE, 2023), COVID-19 caused a 2% drop in GDP construction participation. Then, as indicated in the McKinsey global construction reports (Ribeirinho et al., 2020), the implications of these occurrences are typically industry demands for new technology aimed primarily at increasing efficiency. Indeed, as stated by Shibani et al. (2020) and Edmund et al. (2018), during the worst global crises in history, new technologies “appeared” in the building industry to enhance productivity and, as a result, pull the economic recovery.

However, there is sometimes a harmful aftereffect from the never-ending quest for better productivity. According to Enshassi et al. (2009), there was an increase in work accidents in the construction industry during the same time periods. Even if developing countries have a lower construction representativeness, the negative effects of global crises may be stronger in developing countries than in wealthy ones. As a result, construction planners play a critical role in taking these issues and uncertainties into account in the planning methods and, thus, during the construction phase.

One of the early studies on uncertainty in construction planning was conducted by the Fleet Ballistic Missile program in the 1950s, as described by Touran (1986) and Williams (1999). In this initiative, the method Program

Evaluation and Review Technique (PERT) was proposed by Malcolm et al. (1959) to introduce probabilistic input variables into the Critical Path Method (CPM), which is the most used technique for construction planning, but deterministic in nature. It is worth noting that CPM was idealized almost at the same time by Kelley and Walker (1959), who were inspired by the Gannt chart proposed by Clark in 1922 and later adapted with the predecessor network concept by Fondahl in 1962, but always maintaining its deterministic nature, as stated by Hadipriono (1988). Following the PERT/CPM method, other derived techniques for dealing with uncertainties emerged, such as the Graphical Evaluation and Review Technique (GERT) proposed by Moore and Clayton (1976). The predecessor network in this approach works with deterministic and probabilistic nodes with additional operators to the PERT technique. As a result, Kavanagh (1985) considered GERT to be a complex and refined methodology, whose use is challenging to planners who do not have experience with statistical analysis and derived topics.

Later, in 1988, Hadipriono proposed a deductive method based on fault tree analysis for considering uncertainties in construction planning, which is applicable to both deterministic and non-deterministic analyses. The Modified Fault Tree Networking (MFTN) method adds to CPM causal interrelationships between events that can cause scheduling problems. This approach follows the recommendation made by Caron et al. (1998) regarding the planning process flow, which is to begin at the end. For example, you can first establish the deadlines, and then define the deliverables and procurement processes based on those dates. Using this logic, Hadipriono (1988) argues that MFTN is a construction planning method useful for identifying activities and construction sequences that are most likely to contribute to schedule delays.

Many construction planning techniques were developed during the aforementioned periods to optimize the processes of repetitive construction and to deal with cost and risk estimation (RUSSELL; WONG, 1993). Ock and Han (2010), for example, proposed a fuzzy-based method to calculate the risks associated with uncertainties in other methods. The Line of Balance (LOB) approach is another approach focused on repetitive construction that allows productivity and production rates to be considered alongside the PERT approach, but in a deterministic manner (KAVANAGH, 1985). The problem with this deterministic analysis is that it normally results in optimistic estimations because it does not incorporate uncertainties and random variables into the construction plan (TOURAN, 1986). In fact, according to Bacon et al. (1996), when paired with inadequate risk analysis, this optimistic scenario can render huge projects unaffordable in infrastructure projects built in developing countries.

Indeed, as stated by Lee et al. (2009), in infrastructure projects involving new technologies, project managers typically apply a large margin compared to building projects to cover many uncertainties that traditional planning methods do not take into account. As a result, for that scope of project, contingency estimation is critical (TSENG; ZHAO; FU, 2009), and computer simulations are commonly used to overcome the drawbacks of traditional planning methods. In this context, it is important to mention Monte Carlo (MC) simulation, which according to Woolery and Crandall (1983), is an acceptable technique for performing stochastic analysis in network models of large and complex projects, and, thus, applicable for construction activities. However, Manik et al. (2008) emphasized that MC simulations necessitate a significant amount of computational time and power, and Lee (2005) claimed that these simulations are a useful supplement to traditional planning methods. Similarly, some stochastic methods are based on deterministic methods (e.g., CPM) as demonstrated by the work of Kokkaew and Chiara (2010). The authors proposed the Stochastic Critical Path Method (SCPM) in that study, which combines the critical path with MC simulations and enveloping analysis. By doing so, the authors argue that it is possible to account for the manager's subjectivism in schedule estimation. According to Tseng et al. (2009), another feature of MC simulation is the requirement for better data history and maintenance, which can include pre-processing and data mining techniques. Furthermore, according to Du et al. (2016), MC simulation maximizes the benefits of Markov chain models, which explains why many studies combine MC with stochastic analysis (HASSAN; EL-RAYES; ATTALLA, 2023; HOSNY; NIK-BAKHT; MOSELHI, 2022; KAMMOUH et al., 2022; ZHONG et al., 2016).

It should be noted that simulation techniques are not limited to MC; for example, Halpin (1977) proposed the CYCLONE system, which is based on Discrete Event Simulation (DES), to study construction operations. Indeed, DES is an important technique that is being used by several studies in the construction planning area to integrate with building information models (ABBASI;

TAGHIZADE; NOORZAI, 2020), fuzzy analysis (SZCZESNY; KÖNIG, 2015), probabilistic approaches (FENG et al., 2022), among others.

Furthermore, construction planning methods frequently employ probabilistic analysis to estimate activity duration based on historical data. Naturally, there are inherent uncertainties in that historical database that can affect the estimation results. Probability Density Functions (PDF) are commonly used in this type of approach, as evidenced by the studies of AbouRizk and Halpin (1992) and Lee (2005). Although AbouRizk and Halpin (1992) suggested using the Beta distribution to estimate earth movement activity durations, PDFs are difficult to determine. Furthermore, as stated by Touran (1986), general users typically lack the necessary statistical analysis knowledge to incorporate probabilistic techniques into planning methods.

Indeed, according to Jaśkowski and Sobotka (2006), time and cost are the most used decision-making indicators, which explains why these two variables are frequently used in construction planning methods that account for uncertainties. Also, according to Ock and Han (2010), the success of a construction project is related to three factors: time, cost, and quality, but the “risk path”, as the authors refer to it, may comprehend activities that are not included in the critical path and are related to other areas. However, several authors (KAVANAGH, 1985; OCK; HAN, 2010; OZDEMIR; KUMRAL, 2017; YANG; CHANG, 2005) criticize traditional methods for ignoring resources or believing that resources are limitless. As a matter of fact, resources are the means to an end (VAZIRI; CARR; NOZICK, 2007), which can be time and cost variables, but their omission in construction planning results in unrealistic scenarios. As a result, schedule estimates can be overly optimistic (KAVANAGH, 1985), causing practical issues on the job site such as contractual schedule adjustments. Therefore, CPM, PERT, bar charts, and other traditional methods are insufficient to solve the problem of resource allocation on construction sites, and the subject is widely discussed in construction site planning studies, which propose everything from linear programming to genetic algorithms (YANG; CHANG, 2005).

In this line of reasoning, the worker or labor is introduced, which is regarded as the most critical type of resource by Vaziri et al. (2007) when compared to equipment and materials. However, labor sizing has traditionally been done by relying on past experiences rather than using specific tools or methods. Based on this finding, Elhakeem and Hegazy (2005) proposed the Distributed Scheduling Model (DSM), which is a model based on CPM concepts, progress rates and work crew estimation, and has as its goal the optimization of resource allocation in construction and maintenance operations. Unlike DSM, which employs abacus calculations and deterministic formulations, other authors propose more advanced methods, such as the study by Zahraie and Tavakolan (2009), which employs genetic algorithms associated with fuzzy logic to optimize time and cost while also allocating and placing labor. Similarly, Tomczak et al. (2019) proposed a conceptual mathematical model of multi-criteria optimization with nonlinear processes under deterministic conditions, with the goal of minimizing team downtime and total project duration.

Even more specific is the issue of workplace safety, which is frequently overlooked by construction planning methods but can undeniably interfere with the main indicators of time and cost. The impact of workplace safety on the duration of construction activities is investigated in the study developed by Francis (2019), and a method is proposed that considers both themes concurrently to avoid errors in decision making. It is worth noting that the simultaneous consideration of variables in construction planning methods has already been studied by several authors, including Isidore et al. (2001), who proposes the integration of time and cost simulations to understand the correlation between them.

When compared to vertical building projects, the concern with occupational safety and health issues is even more important in the realm of infrastructure works because they typically involve heavier equipment and a larger number of workers in the field. According to Elhakeem and Hegazy (2005), there are three key decisions in this type of construction: the number of available teams; the construction method used in each activity; and the order of execution of the activities in each space. Certainly, activity prioritization is a study that has piqued the interest of researchers such as Bruni et al. (2011), who propose new prioritization rules based on heuristic programming and statistical analysis, with the goal of taking uncertainties and resource constraints into account in construction planning. Unlike other planning methods with uncertainty, the authors' method was designed to have an easy-to-use and friendly graphical interface, which has aided in the spread of stochastic methods in the construction industry.

Moreover, logistics and equipment sizing can have a significant impact on the risk of accidents on the construction site in infrastructure projects. With this, fleet sizing is a problem that has drawn the attention of stochastic process scholars in construction planning. Ozdemir and Kumral (2017) proposed the use of stochastic processes in the deterministic Match Factor method to consider risks over time, as well as Monte Carlo simulations to understand equipment availability over time. The authors were able to demonstrate through case studies that traditional methods generate exaggerated or pessimistic estimates, whereas the proposed stochastic method generates more realistic, but not necessarily optimistic, scenarios.

To summarize, planners began to consider uncertainties in construction using variations of the CPM approach and were primarily concerned with the time issue (construction duration). The planning network model was then improved by logical methods, which enabled the creation of process maps to simulate the construction sequence. However, the simulation itself was only possible due to advances in computer processing power, which enabled Monte Carlo simulation models and, as a result, the consideration of multiple variables beyond time. Furthermore, specific challenges in construction projects, such as repetitive construction, prompted the development of new techniques that were not entirely based on CPM and were more closely related to statistical analysis. In doing so, the theoretical background revealed that there are many construction planning methods that originated over the last decades and due to technological evolution but understanding the different levels of uncertainty appears to keep this topic at a superficial level of implementation.

### 4.2.2
### Literature review approaches

Following Denyer and Tranfield (2009), the Systematic Literature Review (SLR) is an approach to discovering findings and research gaps in scientific areas in an impartial and objective manner. Although the SLR method strives for objectivity, the researcher's parameters, such as search terms and filters, may be biased due to personal perspective, experience, and knowledge. As a result, this paper proposes the use of adapted systematic reviews, as suggested by He et al. (2017), including term cooccurrence analysis to better understand the topics covered by many papers without requiring a thorough reading and a snowballing analysis as an additional step. By doing so, it is possible to gain a better understanding of whether the search terms are correct and to obtain preliminary answers to the research questions.

Traditionally, according to Khan et al. (2003), the SLR follows a set of standardized steps: question formulation, study location, selection and evaluation, analysis and summary, and results and reporting. The first step is to properly describe the research topic under consideration, which will help direct the search for relevant papers. The following step is to do an article database search using previously defined search terms and filters. As a result, the researcher receives a list of studies that must be selected and reviewed using inclusion and exclusion criteria, which can include language, scope of study, and other relevant criteria to the literature review. The fourth step comprises thoroughly examining the final selection of studies to summarize each contribution, limitations, and subjects relevant to answering the SLR questions. Finally, the fifth stage is concerned with results and reporting, which includes graphs depicting the insights and tables summarizing the SLR results, among other outputs.

The snowballing analysis is another method for conducting a literature review. This method involves the researcher using a reference list or citations to well-known articles on the subject (WOHLIN, 2014). So, it is essentially a thorough dive into a certain papers' references. The second and third steps of the snowballing analysis differ from those of the SLR. Instead of a structured search in databases, the snowballing approach focuses on the key articles in a given subject, and with these key papers, the researchers use inclusion and exclusion criteria to filter the reference list, and then begin the analysis, summary, results, and reports similarly to the SLR. Snowballing can be done in two ways: forward or backward. The references in the important publications are used in the former search. The citation to the key papers is used in the latter search.

The quality of the final list of studies, however, can be influenced by the researcher's experience, as they may not completely comprehend what the significant publications in a specific area are. Furthermore, according to Jalali and Wohlin (2012), the snowballing analysis is simple to understand and replicate, as opposed to the SLR, which contains more difficult procedures for a rookie researcher. However, as a disadvantage, "the lack of randomized:

representativeness" in the snowballing study can result in biased conclusions, according to Geissdoerfer et al. (2017).

It should be noted that the snowballing technique is intended to be complementing rather than a replacement for SLR (WOHLIN, 2014).

## 4.3. Methodology

In this section, an adapted systematic literature review is proposed, combining SLR and snowballing analysis, using the following sets: question formulation, study location, selection and evaluation, analysis and summary, snowballing analysis, and results and reporting. All steps and information used and extracted during the adapted SLR processes are summarized in Figure 9.

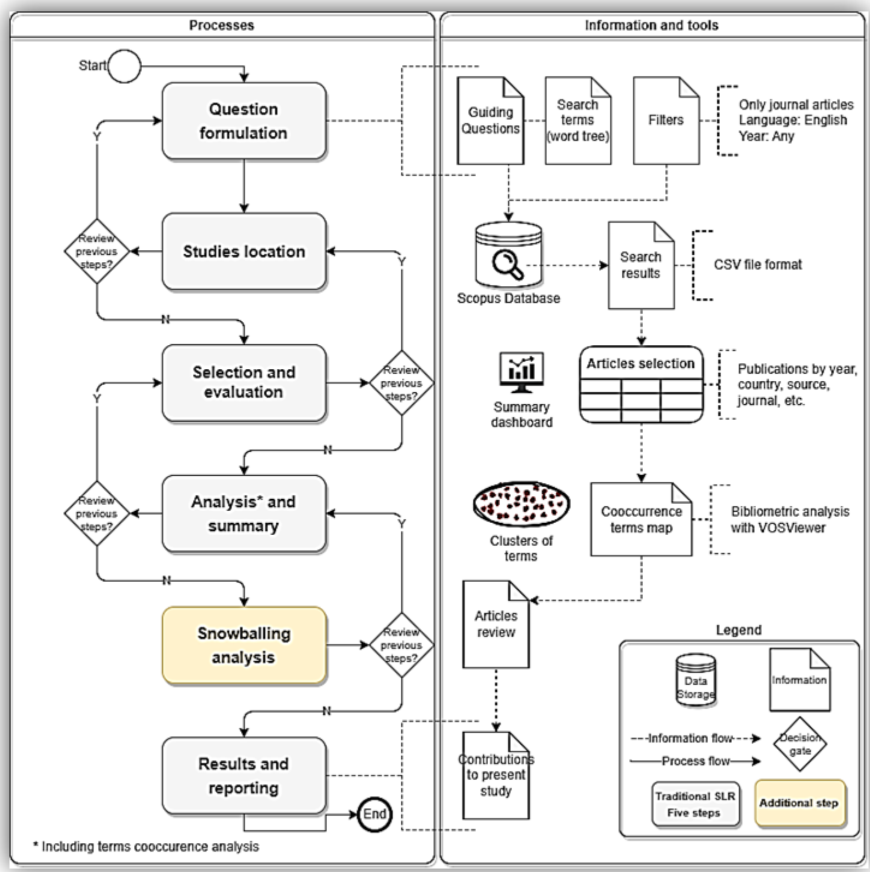



Figure 9 - Workflow that represents the adapted SLR used in the methodology.

The SLR questions in this study are:

- How do construction planning methods handle uncertainties?

- What kinds of uncertainties are taken into account in construction planning methods?

Following that, a comprehensive literature search was conducted, in which Scopus database was used to identify relevant studies by using inclusion and exclusion criteria. Basically, the search engine was configured to find only journal articles written in English without period limit and with the search terms, derived from the initial questions, occurring in the title, abstract, or keywords. The search terms involved words related to construction planning methods (e.g., “construction plan*” and “construction schedule*”) and uncertainties (e.g., “uncertain*” and “risk*”).

Thus, 444 studies were discovered using the database’s search terms and filters. It is worth noting that 67% of the found papers were taken with the search term “risk*” and 33% with “uncertain*”. This distinction indicates that the studies are more concerned with variability and controlled scenarios than with completely unknown scenarios. It is also worth noting that some authors may not have the same understanding of the distinctions between risk and uncertainty, as suggested by Feng et al. (2022). The current study concentrated on uncertainty scenarios that are considered in construction planning methods.

All article titles from the search results were read to determine whether or not they fit the scope of the current work. To avoid mistakes, this evaluation was repeated twice, and 226 articles were removed. The remaining 218 studies' abstracts were then read using the same logic and process as the title reading. The sample was reduced to 103 papers after the second evaluation, which means approximately 23% of the initial sample.

The fourth step in the proposed SLR involved extracting data, such as authors and publication year, from the chosen studies, to develop classification criteria and organize the articles in a logical manner. So, a bibliometric analysis was performed using the ‘VOSViewer’ tool (VAN ECK; WALTMAN, 2010) to better understand the study clusters, categorize similar studies, and identify potential scientific gaps. To that end, it is necessary to define some configurations. To begin, it was established that only the terms presented in the abstract and title would be examined, with structured abstracts and copyright statements being excluded. Second, the terms were counted using binary logic, which means that even if a term appears multiple times in the field, the tool will count each occurrence as one. Third, it was determined which terms should be dropped (generic terms such as contribution, example, and so on) and which should be synonymized. Fourth, the analysis was set up to include only terms that appeared at least three times.

As a result, the analysis yielded six clusters, which are denoted by different colors in Figure 10. The higher the font size, the more frequent the term appears, and the closer the terms, the more they appear together (cooccurrence).

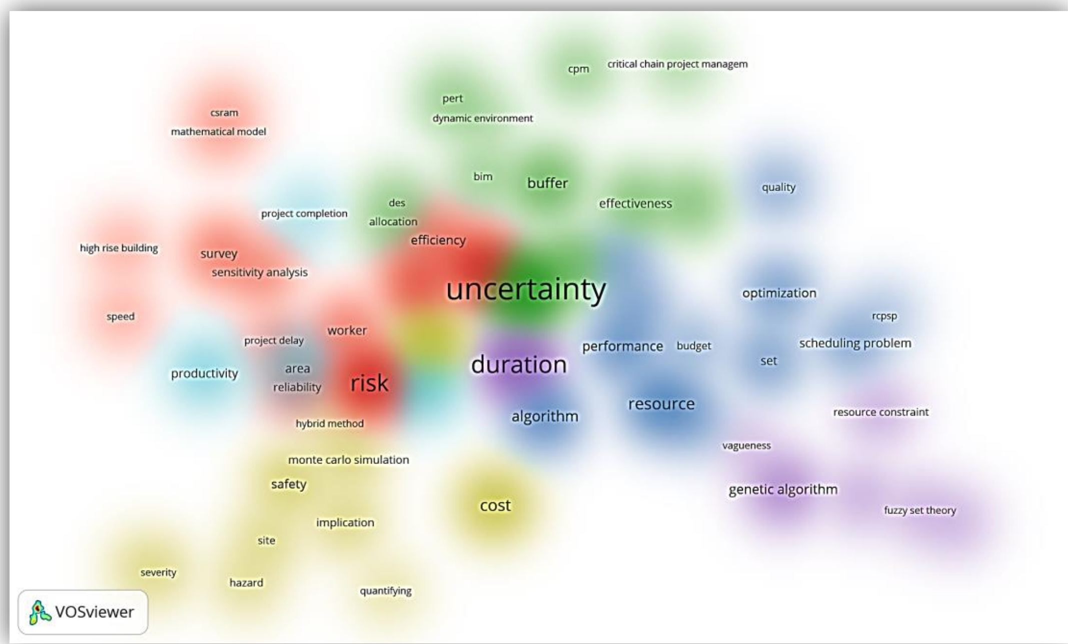



Figure 10 - Cooccurrence terms mapping and clustering resulted from the selected papers.

As expected, the term “uncertainty” has the highest occurrence and is in the center due to the SLR objectives and search terms. The terms “PERT”, “CPM”, and “buffer” are also included in the same “uncertainty” cluster. Close to “uncertainty”, “duration” is the second most cited term, as expected given that planning methods typically deal with activity and project duration. Terms such as “duration”, “genetic algorithm”, and “resource constraint” are included in the cluster identified for “duration”, and can indicate possible techniques used to consider multiple types of uncertainties in the planning method. The third most frequently used term is “risk”, which can be attributed to possible misunderstandings on the distinction between uncertainty and risk, as discussed in the introduction section. Terms like “risk”, “worker”, “efficiency”, and “survey” were grouped together in the same cluster. There is a cluster involving the terms “productivity” and “project completion”, almost mixed with the previous cluster, with no clear conclusion. On the other hand, there is a cluster with terms like “resource”, “optimization”, and “algorithm”, which again indicates possible techniques and construction planning methods purposes that take uncertainties into account. The final cluster that was discovered was associated with “cost”, “safety”, and “Monte Carlo simulation”, which is more closely related to the risk cluster than the duration and resource clusters.

The final sample was then thoroughly read to answer the SLR questions. This step included a sensitivity analysis to classify the studies based on the types of uncertainties being considered (cost, environmental impact, safety, resources, site layout, weather, quality, and others) and the techniques used (fuzzy logic, machine learning, discrete event simulation, probabilistic analysis, Monte Carlo simulation, stochastic processes, information modelling, and others) while considering, but not limited to, the clusters suggested by the bibliometric analysis.

Then, a forward snowballing analysis was carried out to collect papers that were not found by the SLR mechanism but were included in references and were related with the scope of this work. This approach was applied in the key papers observed in the SLR results.

In the final step of the SLR, the authors synthesized the findings of the selected studies to identify interesting discussions and conclusions related to the question formulation. Also, the authors also proposed the creation of a summary table resuming all main suggestions for future work described in the SLR. This step is critical for understanding the overall scenario of the topic in the context of the guide questions. It is intended to answer, even partially, the SLR questions and to provide substantial material to orient new studies that aim to reinforce the main works identified or fill scientific gaps.

## 4.4. Results and discussions

The categorization revealed that the majority of articles (56%) deal with uncertainties through time variables, which was expected given that construction planning methods typically work with deadlines and activity duration. When the papers that are not focusing on time were examined, four main areas represent 81% of the other topics studied as uncertainties in construction planning methods: resources (29%), cost (22%), environment (14%), and safety (16%). It is also worth noting that many works combine those areas, such as the construction duration estimation model proposed by Lee et al. (2009), which considers both weather conditions (environmental issue) and work cycles (time issue). These groupings follow a nearly identical division as shown in Figure 11, with a high occurrence for the four areas mentioned.

Moreover, the Table 5 shows specifically the related uncertainty, the solution method, and the use or not of Artificial Intelligence (AI) adopted in the selected papers from the last 6 years.

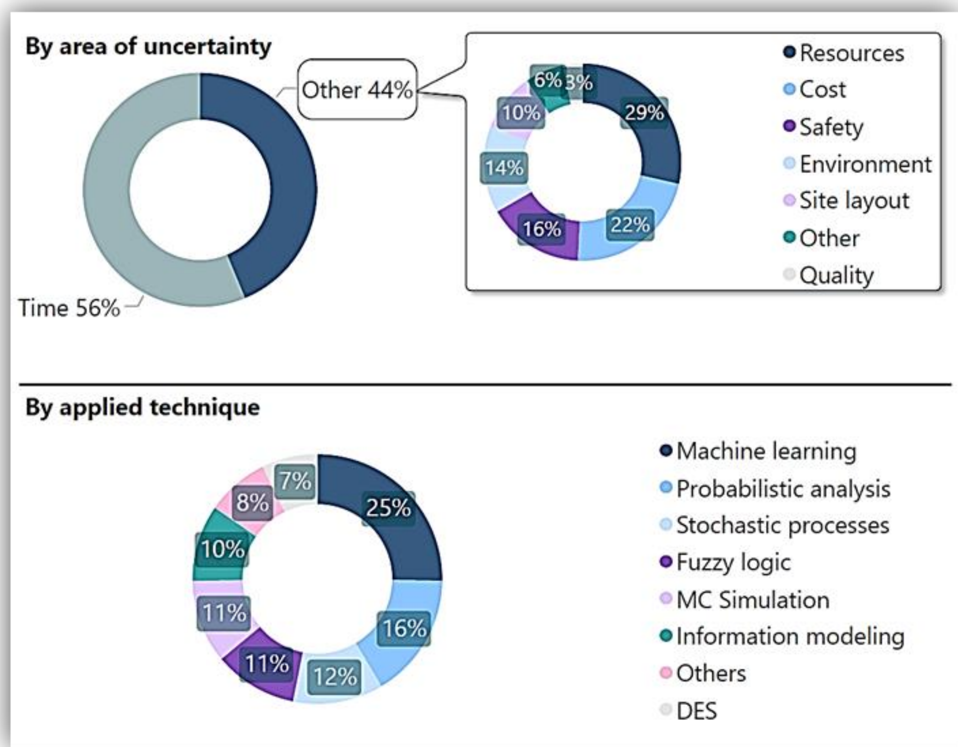



Figure 11 - Papers divided by area of uncertainty and applied technique.

A division by technique was performed besides the area grouping, as shown in Figure 11. This chart shows that some procedures, such as Machine Learning (ML), probabilistic analysis, stochastic processes, Fuzzy Logic (FL), Monte Carlo (MC) and Discrete Event Simulation (DES), and information modelling are frequently used in uncertainty analyses for construction planning. It should be noted that traditional techniques such as probabilistic analysis, stochastic processes, and MC simulation are losing ground to ML algorithms, which gained popularity in the 2000s because of technological advances, primarily in processing hardware solutions. However, some of these algorithms are internally based on traditional techniques, such as the Genetic Algorithm (GA) proposed by Leu and Hung (2002), which searches for probability distributions that best describe project duration under resource constraints.

Table 5 - Selected papers published since 2018: Related uncertainties, adopted solutions, and AI use.

| **Articles** | **Related uncertainties** | | | | | | | **Solution methods** | | | | | | | | **AI used?** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Cost** | **Envir.** | **Resour.** | **Safety** | **Time** | **Quality** | **Other** | **BIM** | **DES** | **FL** | **MC** | **ML** | **Probab.** | **Stoch.** | **Other** | |
| Hu et al. (2023) | | | | • | | | | • | | | | • | | | | • |
| Zhang and Lin (2023) | | | | • | | | | | | | | | | | • | |
| Chen et al. (2023) | | | | | • | | | | | | • | • | | | | • |
| Adedokun et al. (2023) | | | | | | | • | | | | | • | | | | • |
| Wang et al. (2023) | | | | | • | | | | • | • | | • | | | | • |
| AlJassmi et al. (2023) | | | | | • | | | | | | | • | | | | • |
| Hosny et al. (2022) | | | | | | | • | • | | | • | | | • | | |
| Kammouh et al. (2022) | | | | | • | | | | | | | | | | • | |
| Hong et al. (2022) | | | | | • | | | | | | | | | | • | |
| Sharma et al. (2022) | • | • | • | | • | | | | | | | • | | | • | |
| Kedir et al. (2022) | | | • | | • | | | | | | | • | | | • | • |
| Feng et al. (2022) | | | | | • | | | | • | | | | | | • | • |
| Milat et al. (2022) | | | | | • | | | | | | | • | | | | • |
| Canca and Laporte (2022) | | | | | • | | • | | | | | | | • | • | • |
| Ramani and Kumar (2022) | | | | • | • | | | | | | | | | | • | |
| Fitzsimmons et al. (2022) | | | | | • | | | | | | • | • | | | • | • |
| Alhussein et al. (2022) | | | | | | | • | | | | | | | | • | |
| Chen et al. (2021) | | | | | | | • | | | | • | • | • | | | |
| Sarkar et al. (2021) | | | | | • | | | | | | | | | | • | |
| Abadi et al. (2021) | | | | | | | | • | | | | | | | • | |
| Cheng and Zhang (2021) | | | | | • | | | | | | | • | | | • | • |
| Kulejewski et al. (2021) | | | | | • | | | | | | | | | | • | |
| Plebankiewicz et al. (2021) | • | | | | • | | | | | • | | | | | | • |
| Isah and Kim (2021) | | | • | | • | | | | | | | | | • | • | |
| Taghaddos et al. (2021) | | | | | | | • | | • | | | | | | • | |
| Mohamed et al. (2021) | | • | | | • | | | | | | | | | | • | |
| Liu et al. (2021) | | | | | • | | | | | | | | | | • | |
| Biruk and Rzepecki (2021) | | | | | • | | | | | | | | | | • | |
| Ansari et al. (2021) | | | | | • | • | | | | | | • | | | • | • |
| Kaveh et al. (2021) | • | • | • | • | • | • | | | | | | • | | | • | • |
| Zhang and Wang (2021) | | | | | • | | | | | | | • | • | | | |
| Hassan et al. (2021) | • | | | | • | | | | | | • | • | | • | | • |
| Chakraborty et al. (2020) | • | | | | | | | | | | | • | • | | | |
| Abbasi et al. (2020) | | | | | • | | | • | • | | | | | | • | |
| Hosny et al. (2020) | | | • | | | | • | • | | | | | | | | |
| Jaśkowski et al. (2020) | | | | | • | | | | • | | | | • | | | |
| Zohrehvandi and Khalilzadeh (2019) | | | | | • | | | | | | | | | | • | |
| Maronati and Petrovic (2019) | • | | | | • | | | | | | • | | • | • | | |
| Wang et al. (2019) | | | • | | | | | | | | | | | | • | |
| Husin (2019) | | | | | • | | | • | | | | | | | | |
| Tran and Long (2018) | • | | | | • | | | | | | | • | | | • | • |
| Rahman et al. (2018) | • | | | | • | | | | • | | | | | | • | |
| Li et al. (2018) | | | • | | | | | | | | | • | | | • | • |

A more detailed discussion about the areas and techniques identified in the selected papers is presented in the following sections.

### 4.4.1. Uncertainties related to time, cost, and resources

A common approach used to address time issues as uncertainties in construction planning methods is to draw on previous experiences and user subjectivism. For example, Mulholland and Christian (1999) developed a study that involved quantifying uncertainties in a construction chronogram using expert knowledge and experience, lessons learned, and project information. Similarly, understanding that there is subjectivism in the information provided to estimate construction duration, AbouRizk and Sawhney (1993) proposed a system to assess uncertainties caused by planner subjectivism using FL.

Another common subset of time uncertainties is the time buffer, which is essential for simulating both optimistic and pessimistic scenarios in activity duration. In fact, Nasir et al. (2003) claim that the definition of upper and lower duration values can have an impact on risk management. For that reason, Sarkar et al. (2021) proposed a Critical Chain Project Management by improving the buffer sizing through the integration of multiple uncertainties that affects the construction schedule, such as environmental disasters and resources restrictions.

Note that uncertainties are typically treated as risks, especially when they are time related. Consider the study of Chen et al. (2023), who investigated the interdependence of risks in building construction schedule using Bayesian networks and MC simulation. They suggested a planning strategy that, when compared to standard methods (CPM and PERT), resulted in more accurate construction time due to its ability to foresee the sequence of risks. An interesting aspect of the study performed by Chen et al. (2023) is that they suggested that the literature is limited in approaches that include risk interdependence, which could be interpreted as a lack of understanding on deeper uncertainties.

Returning to the discussion on time buffers, Ma et al. (2014) proposed a framework to size buffers and allocate resources based on the critical chain concept, similarly with the study of Sarkar et al. (2021). Furthermore, they emphasized that improving information flow can reduce uncertainties in construction activities, which is accordance with Abbasi et al. (2020). The authors proposed a construction planning method that uses information extracted from a Building Information

Modeling (BIM) model and DES searches for optimal activity durations that represent realistic scenarios.

The complete reading of the selected papers revealed that information modelling, primarily related to BIM and Virtual Design and Construction (VDC), is widely used in construction planning method proposals, such as the work developed by Li et al. (2009). The authors investigated virtual construction prototypes created with BIM and VDC to analyze and optimize the schedule through the visualization of “what-if” scenarios. It is worth noting that the solution in this case is influenced in part by the planner's subjectivism regarding the virtual model, bringing back the importance of studies like Mulholland and Christian (1999) and AbouRizk and Sawhney (1993).

To summarize, some light can be shed on current planning methods that account for uncertainties for time issues. First, the duration of the activity is considered in different scenarios, ranging from pessimistic to optimistic. It is already a result of the PERT implementation and its use as a model for new approaches. Second, the time buffer is essential not only for covering uncertainties during construction, but also for carrying out meaningful risk management. Third, model visualization is a feature that has been investigated to help with the search for better solutions and the impact of subjectivism on the schedule.

Another major point of discussion is resource constraints. Construction planning methods have traditionally assumed that resources are limitless and, thus, always available during the construction phase, but this is not the case. To solve this issue, many studies are using AI algorithms. Li et al. (2018), for example, used multi-objective optimization algorithms and metaheuristics; Kim and Ellis (2009) presented a hybrid and adaptative GA; and Leu et al. (1999b) and Leu and Hung (2002) proposed a GA to find optimal solutions in resource allocation problems.

However, due to their multidimensionality, resources can be abstract and difficult to plan. The term “resources” refers to the equipment, workers, materials, and other auxiliary products required to carry out construction activities. As a result, Hosny et al. (2020, 2022) proposed a tool to model workspaces and detect interferences between them to gain a better understanding of some uncertainties that can be assessed before construction begins. Nonetheless, other studies deal with resources during the construction process, suggesting “live” planning methods, such as the work done by AlJassmi et al. (2023). Their work comprises of a neural network-based planning system that self-recovers the construction schedule by collecting and analyzing worker productivity rate on a regular (daily or weekly) basis.

When looking at studies that discuss cost uncertainties, two sub-areas were identified: cost estimate and cash flow. Cheng et al. (2013) proposed an inference model based on Support Vector Machine (SVM) and time series, in which FL is used to work with cash flow problems and construction estimates. In relation to cost estimation, Chakraborty et al. (2020) highlighted, after comparing multiple ML algorithms, that the use of a hybrid ML model to deal with uncertainties in the cost issue, associated with a probabilistic approach, is recommended.

It is important to note that many studies combine cost with other categories of uncertainty, such as time (HASSAN; EL-RAYES; ATTALLA, 2023), environmental impacts (SHARMA et al., 2021), and so on. One of these categories, safety, is regarded by the authors as the one for which uncertainties are most difficult to estimate, since it is associated with human factors such as emotions, health, and a plethora of random variables that extend beyond a number or a historical data set.

### 4.4.2.
### Uncertainties related to other variables

In relation to the safety issue, due to the high level of uncertainty that is involved, from a heart attack to an explosion that can result in a construction accident, many authors carry out questionnaire surveys to assess the uncertainties. For instance, Zolfagharian et al. (2014) proposed an automatic tool for safety planning based on a risk matrix calibrated with a survey applied to safety and construction managers.

Information modeling and simulations are also commonly used in studies related to safety uncertainties. Benjaoran and Bhokha (2010) proposed a rule-based integrated system that allows the user assessing and reviewing construction planning through model visualization and, thus, viewing potential work-related accidents. Goh and Askar Ali (2016) presented a hybrid simulation framework to facilitate the integration of safety uncertainties and construction activity sequence. They used DES, system dynamics, and agent-based simulation to achieve this goal.

Sometimes, however, uncertainties related to safety require spatiotemporal analysis. Hu et al. (2023), for example, presented a strategy for dealing with safety accidents induced by crane operations, based on a spatiotemporal analysis that was based on a BIM model. They collected data using the BIM methodology, applied AI algorithms to perform path analysis (connected to the crane's position), and visualized the results using a hazard exposure heatmap.

Uncertainties related to cognitive aspects were also identified in the SLR results. For example, Alhussein et al. (2022) utilized agent-based methods to investigate improvisational behavior in construction planning that emerges from unanticipated uncertainty. They discovered that improvised solutions are produced more frequently by managers than by laborers. It should be noted that agent-based techniques to deal with uncertainties are common in recent literature, with numerous research works such as those by Zhang and Lin (2023), Goh and Askar (2016), Abadi et al. (2021), Kedir et al. (2022).

Weather is other issue that is commonly studied in planning methods. For instance, Pan (2005) addresses rainfall uncertainty by proposing a construction planning approach that assesses the impact of rainfall on construction duration by using historical rainfall data and expert knowledge.

Other studies go beyond the specific category of uncertainties, but rather about risk inference in infrastructure project construction planning. Chen et al. (2021) created a method that does not require observed data and can be useful in cases where historical data is unavailable. The authors employed MC simulation with Bayesian networks to infer risks in infrastructure building scheduling.

### 4.4.3. The growth of AI techniques

Aside from the types of uncertainties covered by the present SLR, certain discussions are necessary about the strategies used in these planning methods. Historically, the SLR indicated that uncertainties were frequently evaluated using statistical analysis (TOURAN, 1986), mathematical formulations (WOOLERY; CRANDALL, 1983), and simulation - typically DES (SAWHNEY; ABOURIZK; HALPIN, 1998) and MC (SUKUMARAN et al., 2006). In fact, Figure 12 shows that almost 50% of the planning methods found in the SLR are related to these traditional techniques.

However, more recent studies found in the SLR revealed that AI methods are commonly utilized. Some construction planning methods use evolutionary algorithms to deal with multi-objective optimization (HASSAN; EL-RAYES; ATTALLA, 2023; MILAT; KNEZIĆ; SEDLAR, 2022; TRAN; LONG, 2018), FL to deal with uncertain data (PAWAN; LORTERAPONG, 2016; PLEBANKIEWICZ; ZIMA; WIECZOREK, 2021; SZCZESNY; KÖNIG, 2015), neural networks to forecast schedules in real-time (ALJASSMI; ABDULJALIL; PHILIP, 2023), rules induction methods to interpret and assess construction scenarios (FENG et al., 2022), among others. Figure 12 depicts a summary of the AI techniques revealed in the SLR. The widespread usage of FL and GA (53%) may indicate that they are the most promising methodologies to account for uncertainty in building planning procedures. However, the other part (47%) of the AI methods uses other ways that could also be seen as a promise for portraying random events in construction plan.

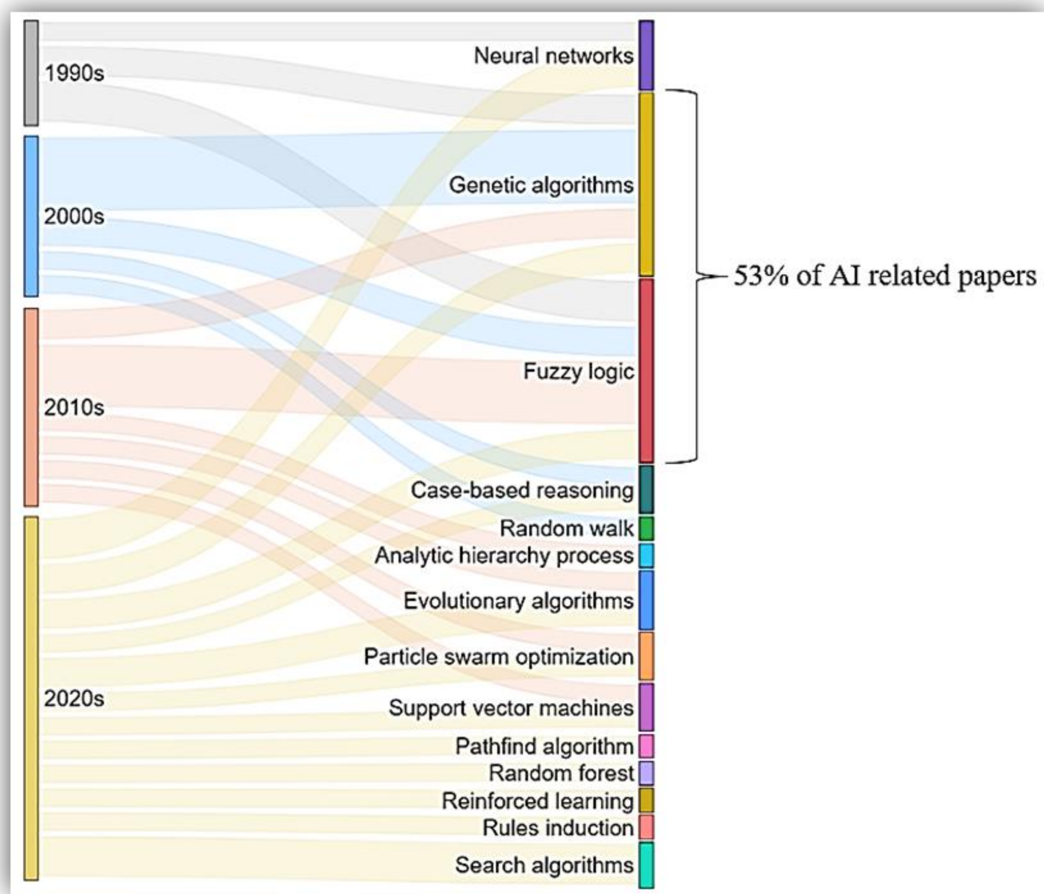



Figure 12 - AI techniques utilized over the time in the construction planning methods.

Also, it is important to note that many studies propose hybrid planning methods that combine traditional and contemporary techniques, such as the work done by Fitzsimmons et al. (2022), who combined MC simulations with support vector machines, a well-known AI technique, to predict project delays while accounting for uncertainty. Furthermore, the integration of FL with GA has been reported by Moon et al. (2015) and Cheng et al. (2013). The former was used to consider the random variables, while the latter was utilized to optimize the outcomes. Indeed, some studies have already assessed AI approaches applied in certain themes of planning with uncertainties, such as Chakraborty et al. (2020), who analyzed six AI algorithms to anticipate cost and discovered that a hybrid solution produces better estimates.

### 4.4.4. Perspectives for future works

The present work suggests that new studies may consider unusual uncertainties (not just time, for example) and hybrid solutions, that combine traditional methods with advanced algorithms, in the development of new construction planning methods. But it is important to highlight the recommendations for future work that are provided in the selected studies. Researchers can use these suggestions to direct their studies to fill scientific gaps completely or partially and, as a result, improve this research field.

In the current work, twenty suggestions from the reviewed articles were identified to guide future studies. Many authors suggested as future works the expansion of their methods to other countries, industries, and more detailed data, aiming to validate their approaches with different test environments. Moreover, it was identified that some studies recommend dynamic approaches to deal with actual and field data, to create a kind of "live planning". Certainly, to make it viable, a user-friendly interface and advanced algorithms are needed, which are other two recommendation given by some authors. Finally, some authors proposed cognitive studies to better understand human behavior and planner attitudes toward construction planning (Table 6).

Table 6 - Suggestions for future works identified in the SLR.

| # | Suggestions for future studies | Number of related studies |
|---|---|---|
| 1 | Context expansion (application in other industries or contexts) | 10 |
| 2 | Increasing sample data (application in more details) | 6 |
| 3 | Algorithms improvements | 6 |
| 4 | Development of dynamic approaches to deal with real-time data | 6 |
| 5 | Capturing field data with monitoring technologies and use of actual data | 6 |
| 6 | Geographical expansion (methods implementation in other countries) | 4 |
| 7 | Consideration of multiple variables | 4 |
| 8 | User friendly interface | 4 |
| 9 | Better understanding of variables and parameters | 3 |
| 10 | Development of microsimulation and micro modelling to deal with more details | 3 |
| 11 | Processes automatization | 3 |
| 12 | More case studies applying existing methods | 3 |
| 13 | Better computational performance | 2 |
| 14 | Cognitive studies to better understand people behavior and resilience | 2 |
| 15 | Creation of knowledge database based on field data and/or past experiences | 2 |
| 16 | Creation of specific databases to support similar studies | 2 |
| 17 | Investigation of hybrid simulation | 1 |
| 18 | Improvements in math formulations | 1 |
| 19 | Consideration of constraint conditions. | 1 |
| 20 | Deeper understanding of uncertainties | 1 |

### 4.4.5. The answers to the SLR questions

• How do construction planning methods handle uncertainties?

There are several ways that construction planning methods deal with uncertainties, and some commonly used strategies have been identified. First, as observed in Hossen et al. (2015); Mulholland and Christian (1999); Nasir et al. (2003); Rozenfeld et al. (2009), planners use risk assessment techniques to identify potential sources of uncertainty and to evaluate the likelihood and impact of these risks during construction project. They investigated methods for mitigating or managing these risks, such as developing contingency plans and acquiring additional resources. Second, it was discovered in some studies (ANSARI, 2021; CHEN et al., 2021; FORD; LANDER; VOYER, 2002; TRAN; LONG, 2018) that planning methods handle uncertainties by relying on flexibility and adaptability.

Flexibility allows for adjustments to be made in response to changing circumstances caused by uncertain factors. Building in contingencies allows for scope changes, and being open to alternative approaches are all examples of this. Third, many of the proposed methods make use of computational resources and simulations to better understand various scenarios and critical sequences that may occur during the construction phase. Following the studies developed by Woolery and Crandall (1983), Zhong et al. (2016), MC is one of the most used techniques for simulation purposes. However, methods that use DES, such as Abbasi et al. (2020); Goh and Askar Ali (2016); Szczesny and König (2015), should be mentioned as an approach to understanding construction scenarios with uncertainties. Finally, a brief discussion on construction monitoring and lessons learned is provided next. Some planning methods (BI et al., 2015; KAMMOUH et al., 2022; SZCZESNY; KÖNIG, 2015) establish systems for ongoing monitoring of the project to identify potential issues early on and make necessary adjustments. There are also studies focused on the creation of knowledge bases, such as Pan (2005), which assesses the impact of rain on project completion based on historical data and expert experiences. Overall, the key to dealing with uncertainties in construction planning methods involves identifying uncertainty sources, strategies for collecting and modelling construction data, flexibility to work with many possible scenarios, and computational solutions, such as simulations, to capture multiple solutions and outcomes caused by uncertainties.

• What kinds of uncertainties are taken into account in construction planning methods?

According to the SLR results, uncertainties related to time (construction duration, project delay, productivity rates, and so on) are the most used in construction planning methods, followed by resource and cost issues. The reason for this is most likely because construction management has traditionally been based on three pillars: time, cost, and resources. However, after thoroughly reading the selected articles, it was discovered that many methods deal with multiple variables at the same time, such as the study carried out by Leu et al. (2001) to optimize time and cost in construction trade-off subject to uncertainties. Furthermore, the findings point to other issues that have been investigated by numerous studies, such as quality, environmental impacts, and safety. In terms of safety issues, simulation techniques are typically used, such as the work done by

Wang et al. (2016), and historic databases support planning methods by providing previous knowledge to draw future possible scenarios, such as the CHASTE approach (ROZENFELD; SACKS; ROSENFELD, 2009). Similarly, in terms of environmental impacts, historical data is frequently used in planning methods associated with ML algorithms and fuzzy analysis, such as the work developed by Pan (2005). Furthermore, the most used techniques are ML algorithms and probabilistic analysis, with stochastic analysis coming in third with some studies relating to time, cost, resources, environmental impacts, and/or site layout. It is also worth mentioning the use of MC simulation in conjunction with stochastic processes, which is present in at least seven studies.

## 4.5. Conclusions

The authors used an adapted SLR to examine the state of the art in construction planning methods with uncertainties. The results revealed that most approaches consider time issues as a variable to consider uncertainty, but they also exposed that there are many methods that consider multiple variables at the same time. Furthermore, the findings showed that there is currently no common system in use, and that a combination of traditional techniques and advanced algorithms is being used to estimate uncertainties during the construction phase.

These findings have important implications for the understanding of this research area, which requires more studies not only consolidating existing methods, but also validating new methods with benchmarking data. As indicated by the summary of suggestions for future works, there is a need for universalization and dynamization of current methods, extending the application to other project contexts, countries, etc., and bringing actual data to create a “live planning”.

This paper encourages construction planners to use methods that account for uncertainties and do not overlook out-of-the-ordinary planning variables that may have a significant impact on the project because of crisis events. Likewise, the review suggests that new research on construction planning methods should take those uncertainties into account while exploring different approaches that have already been discussed in the literature. Moreover, the authors suggest that new studies can be oriented to construction planning methods that deals with higher level of uncertainties, being able to capture unusual random events commonly observed in infrastructure project.

Finally, it is critical to emphasize that there is no close answer for considering uncertainties in construction planning approaches. On the contrary, some studies suggest that hybrid methods may be the best option for dealing with some sorts of uncertainties. Rather than repeating failing planning methods, that do not consider uncertainties during crisis events, academics and industry practitioners could dive into AI growth and strive to uncover atypical uncertainties with atypical strategies that can help anticipate the future of construction planning.

# 5
# A stochastic pure birth model for predicting and help prevent accidents in energy infrastructure construction projects

Paper submitted by Cristiano S. T. do Carmo and Elisa D. Sotelino in a peer-reviewed international journal.

This study proposes a novel construction planning approach to predict safety events amid un-certainties in the construction industry. Recognizing historical challenges and recent crises, the research aims to help enhancing worker safety in construction projects. Current planning methods often lack the ability to handle random variables related to accidents, leading to overly optimistic projections. In contrast, the proposed approach leverages stochastic processes, an underutilized classical theory in the literature related to construction planning, coupled with modern computational power. Using the Knowledge Discovery in Databases (KDD) framework, real construction data is refined for training predictive models. The study presents two solutions: probabilistic distribution and stochastic processes. Results from 39 projects reveal that the probabilistic solution is optimist, and the stochastic solution provides a cautious outlook. While both methods fit some projects well, the probabilistic solution excels in minimizing false positives, while the stochastic approach offers superior precision. By balancing precision and recall, the stochastic approach outperforms in F1-score and the area under the precision-recall curve (PR-AUC score). Further analysis supports its advantage in matrix similarity scores. Notably, the potential integration of advanced Artificial Intelligence (AI) methods is highlighted within the robust stochastic framework.

## 5.1. Introduction

According to the Brazilian Statistics Institute – IBGE (2023), in the last 22 years the construction industry accounted for 5.1% of gross domestic product added in Brazil on average, and behind this figure are numerous repercussions on the local economy, such as job creation and housing deficit reduction. This industry employed 7,7 million of people in 2019, according to the same sources. As a result, studies related to construction projects difficulties and potential solutions is a constant demand, as it will affect not only the country's economy, but also the lives of workers.

Its significance in the local and global contexts is, thus, undeniable, particularly during economic crises due to the COVID-19 pandemic and the Ukraine War, for example. The effects of these important events on this industry might linger for years, affecting economic, political, and social fields and capable of generating new difficulties or rescuing existing problems for technical discourse. In the case of the pandemic, the subject of Occupational Health and Safety (OSH), which was a contested issue in the 1970s, has returned to be one of the most cited study issues in the previous two years. A quick search in the Scopus database with the terms "construction industry" and "safety management" yielded 207 journal articles since 2020.

In this sense, understanding the issue of OSH in the construction industry as a field that requires further research, the current study aims to develop a predictive model of construction accidents. More specifically, the concept of stochastic processes and knowledge discovery in databases are applied in a dataset related to safety occurrences in infrastructure projects, to grasp the study´s main inquiry: how to better predict and prevent accidents in construction projects using not only the past but also future scenarios?

The paper is organized into the sections listed next. The first section, introduction, addresses the historical and local context of the construction challenge that is addressed in the current study. The second section, theoretical framework, outlines the key concepts used in this research. The third section presents the literature review, scientific searches, and summarizes the articles that contributed to the present effort. The fourth section, related to the proposed methodology, outlines which approaches, tools, and processes were used to solve the topic under consideration. In the fifth section, results and discussions, the proposed methodology applied to a case study with real project data and the results are detailed analyzed to understand the discovered knowledge. Finally, in the conclusion section, a summary of the contributions and limitations of the present work, and suggestions for future work are presented.

## 5.2. Theoretical framework

The two major concepts employed in the proposed methodology are presented in this section. The first, on safety management, focuses on presenting current research that deals with this topic, highlighting the utilized approaches. The second concept is related to stochastic processes. This concept´s necessary theoretical background required to comprehend the suggested methodology is provided. Because the aim of this work is dedicated to apply the existing knowledge of stochastic processes to accident prevention in the construction sector, no deep conceptual development related to that technique is presented.

### 5.2.1. Safety management in the construction industry

To understand how accidents are managed in construction projects, a literature review was carried out. It focused on locating studies that propose solutions on occupational safety in the construction industry. The review methodology used the following search terms with their synonyms: “safety management”, “construction industry”, and "data modeling"; and filters were applied to limit the search to only journal articles published in journals in English. After an analysis of titles and abstracts, some studies were discarded, and others were selected for full reading. The contributions of the selected papers to the development of the current study are explained next.

The first observation after reading the selected articles was that many of them propose the adoption of Building Information Modeling (BIM) methodology to support OSH analysis in construction planning. For example, Zhang et al. (2015) propose a BIM tool that automatically identifies possible points of worker fall accidents during the construction phase. The innovation of their work is that the tool automatically models the safety components that prevent the fall in the BIM model and, thus, without the modeler´s intervention, thus, reducing the modeling effort. It is worth mentioning, however, that the tool requires a BIM model with detailed geometric and temporal data before the actual construction begins, which is not yet the reality in the construction industry in many countries, including Brazil.

Similarly, Sacks et al. (2009) created a system named CHASTE, that takes as input data traditional construction planning and the BIM model. Based on the layout of employees and equipment in the building site, it uses probabilistic algorithms to analyze possible victims exposed to loss-of-control scenarios (such as falling ceramic tiles from the facade) using spatial and time variables. As a result, the technique generates data matrices of worker´s exposure in all planned tasks and sub-activities, categorized by accident type and severity. However, the authors discovered during the implementation in a case study that the algorithm does not take into consideration distinct construction approaches.

The use of the spaciotemporal BIM model to identify conflicts and risk exposures is found in several articles, as is the case with the work developed by Choi et al. (2014). However, most methods have as a limitation the non-consideration of randomness inherent in the construction process, which can lead to deterministic results that do not accurately describe the construction site's reality. Like Choi et al. (2014). Su et al. (2018) also studied a spaciotemporal mathematical model using MATLAB. They used singularity functions and expressions to take uncertainties into account and used the BIM model as input. Although not strictly focused on safety management, it may be an indicator of how to consider uncertainties in conjunction with a BIM model.

Another line of research found in the selected articles is related to ergonomics and ergometric research that may be causal factors in construction accidents or harm to employees' health. For example, Golabchi et al. (2015) employ BIM model´s data to simulate work-related musculoskeletal illnesses. However, their approach requires a high level of model detail for ergonomic posture analysis and biomechanical analysis, which may not match the reality of many building projects. A similar limitation was identified in the work by Zhang et al. (2015). Without the need for such a complex modeling, Zaalouk and Han (2021) investigate work environments and offer a project parameterization with modular constructs to reduce the difficulties and hazards of worker´s poor posture. This is done using genetic algorithms with multicriteria optimization. Finally, the work of Wang et al. (2016), which provides an occupational safety planning that is updated in real time. This research focuses on the development of underground caverns and the use of a geometric and temporal BIM model to update simulations of the construction processes with real-time geological data that may signal possible structural instability of the cave. Although a Structured Query Language (SQL) database is used for the dynamic model, the authors claim that the method is less difficult than prior proposed methods for the same purpose.

As observed, the topic of safety management in the construction industry is being studied in many research fields, specially related to BIM and data mining using AI techniques. Also, the studies revealed that the uncertainties are rarely examined, which can be one of the causes of accidents – unplanned events. Studies point in the direction of a dynamic model and a more simplified geometry to make viable methods of construction planning, as is pursued in the current study.

It is worth pointing out that, aside from the recent studies, traditional and well-known articles are still references for handling uncertainty in safety management. The study developed by Chua and Goh (2005) is an example of that. In their work, the authors examined the ideal statistical distribution for construction events. Due to a lack of scientific work aimed at modeling construction safety occurrences, the authors proposed the homogeneous Poisson distribution to account for the intrinsic random nature of accidents. Using a dataset of 14 projects, they demonstrated that the Poisson parameter may be utilized as a quantitative indication of safety in railroad construction projects using the chi-square goodness-of-fit and scatter tests. Furthermore, construction incidents, according to them, can be separated into n Poisson subprocesses that represent diverse categories, such as the type of occurrence and the severity of the accident. The authors suggest that big databases be used in future studies to better understand the systemic elements that cause an accident and the applicability of this distribution (Poisson) in various types of construction projects.

It is worth mentioning that the use of Poisson process to describe the accident probability is not specific to construction projects. Nicholson and Wong (1993) had previously investigated this issue in the context of traffic accidents and concluded that a Poisson distribution can be assumed for annual events. This type of distribution was also employed by Sari et al. (2009) to explain the number of failures, which is equivalent to accidents, associated to coal mines. Also, Janardan (1998) investigated the number of failures related to computer chips and concluded that the Poisson distribution is suitable to understand the average failure rate.

However, according to Zhang et al. (2021), there are few research works that employ stochastic processes as a theoretical foundation for understanding safety risk in construction projects. The combination of stochastic theories with construction risk, they claim, is "meaningful and reasonable," and these random processes are used in a variety of applications such as failure detection and asset integrity. Furthermore, Chua and Goh (2005) have shown that the Poisson process may be utilized to represent the randomness of construction accident occurrences. The current work further contributes to the scientific field by utilizing stochastic process theory.

The research area of Machine Learning (ML) is also observed in the articles selected in the literature review. It should be noted that most of the studies (GERASSIS et al., 2017; SARKAR, Sobhan et al., 2019; SHIN et al., 2018; TRILLO CABELLO et al., 2021; XU; ZOU, 2021) are directed to the evaluation of the causes of accidents. Consequently, the datasets chosen in those research field include extensive information on the accident, such as the victim's age and gender, which is outside the scope of the current study. However, two works (LI, Xin et al., 2021; ZHANG, Fan et al., 2019) employed a dataset like the one available for the current study, with accident records collected throughout time at different construction sites and with varying severity levels (fatal, first aid, etc.). Nevertheless, it should be emphasized that the databases used by both works contain more information and are available to the public from the OSHA. The current study, on the other hand, makes use of a private company's database that has limited records of safety incidences. Yet related to ML, no specific algorithm seems to predominate, but decision trees and association rules techniques are the most frequently used in these works (AMIRI; ARDESHIR; FAZEL ZARANDI, 2017; DHALMAHAPATRA et al., 2019; SARKAR, Sobhan et al., 2019; SHIN et al., 2018; TRILLO CABELLO et al., 2021; XU; ZOU, 2021).

As conclusion from the literature review it was observed that most work planning approaches do not account for random variables associated with work accidents. So, the ensuing construction timelines may be unduly optimistic, since they mainly consider contractual, financial, and other similar problems. Besides that, it reveals that strategies, tools, and methodologies (e.g., decision tree algorithms, association rules, and BIM) have been used in specific scientific articles on safety management in construction projects.

### 5.2.2. Stochastic processes

A stochastic process is a family of functions that use a parameter to translate each instance of a sample space to a new domain. This function map, also known as a random process, is frequently tied to a time parameter, which can be discrete or continuous and relies on how the experiment's observations are carried out. Because the observed instances might be discrete or continuous, continuous, or discrete stochastic processes with continuous or discrete parameters are defined.

As described by Beltrami (2013), the number of calls that arrives in a fire department over the time is an example of a stochastic process, specifically known as a Poisson process. In this process, the observation (call) is discrete and counted in unit increments, because it is not possible that more than one call will arrive at the same moment. However, the observed time is continuous, the time difference between accidents does not follow a standard increment, and an instance can occur at any time instant.

Another example of a stochastic process is the birth and death process, defined by Kendall (1948) as a stochastic process $n(t)$ with non-negative integer values ($S$ states) and birth and death rates as a function of time, i.e., $\lambda_s(t)$ and $\mu_s(t)$, respectively. Furthermore, Karlin and McGregor (1955) defined the birth and death process as a random walk process with a continuous or discrete time parameter.

In this paper, the Poisson and pure birth models are utilized to understand and predict the transitions of safety occurrences in a construction project. The pure birth process is a simplification of the birth and death process, in which the death rate is zero for any state and the birth rate $\lambda_s(t)$ represents the rate of new occurrences (an accident, a near miss incident, or a safety observation) in time $t$ and state $s$. In other words, an accident cannot be deleted once it has been recorded, and the higher the birth rate, the more safety events can occur.

Note that previous studies considered safety states in different aspects. Li et al. (2016) assumed that the kinds of safety occurrence describes the safety state, namely: normal incidences, near-misses, accidents, and nonoccurrence. Zhang et al. (2021) assumed that the risk factor contributes to an evolution from risk to accident, which can be understood as two stochastic states: risk and accident. The present study, however, considers the stochastic state as the number of safety events, similarly with the stochastic times-series analysis proposed by Marhavilas et al. (2013).

Figure 13 shows the diagram that describes the evolution of safety events over the time for a typical construction project. Each construction project is specified by its own safety events rates, but together they describe the stochastic process, where the circles represent the possible safety states, varying from 0 events to $N$ possible safety events, and the arrows represents the possible transitions between states ruled by the birth rates.

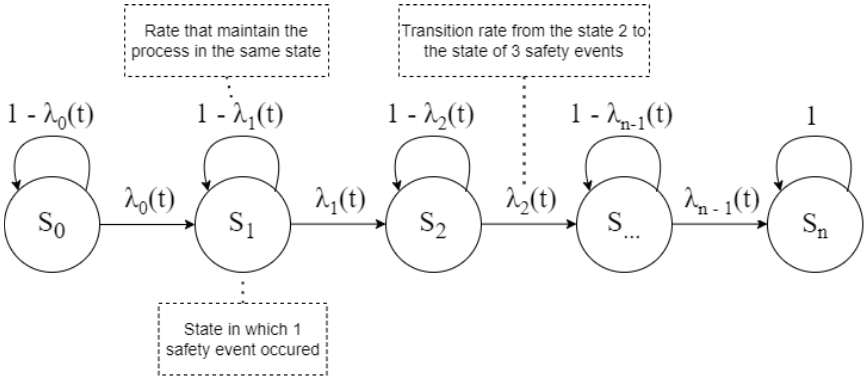



Figure 13 - States diagram that represents the safety events in a construction project.

Thus, for each project, the probability that a safety event occurs in a non-negative time interval $dt$ ($p_{ij}(dt)$), from state $S_i$ to state $S_j$ where $i$ and $j$ can assume any value between 0 and $n$, can be formally described as:

$$p_{ij}(dt) = \begin{cases} \lambda_i(t)dt + o(dt), if\ j = i+1 \\ 1 - \lambda_i(t)dt + o(dt), if\ j = i \\ 0, otherwise \end{cases} \tag{16}$$

Where,

- $o(dt)$ is any function that satisfy the condition $\lim_{dt \to 0} \frac{0(dt)}{dt} = 0$;

Furthermore, assuming two hypothetical states: $S_a$ in time $t$ and $S_b$ in time $t + dt$, it yields:

$$p_{ab}(t+dt) = \lambda_{b-1}(t)p_{a,b-1}(t)dt + [1 - \lambda_b(t)]p_{ab}(t)dt + o(dt)$$

Then, subtracting both equation sides from $p_{ab}(t)$ and dividing by $dt$, the result is the Kolmogorov forward equation, as demonstrated by Miller and Childers (2012):

$$\frac{\partial p_{ab}(t)}{dt} = \lambda_{b-1}(t)p_{a,b-1}(t) - \lambda_b(t)p_{ab}(t) \tag{17}$$

Organizing as a matrix, following Karlin and McGregor (1955), the previous equation can be rewritten as:

$$\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{A}$$

Where, $\boldsymbol{P}(t)$ represents the transition probabilities:

$$\boldsymbol{P}(t) = \begin{bmatrix} p_{00}(t) & p_{01}(t) & \cdots & 0 & 0 \\ p_{10}(t) & p_{11}(t) & & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \cdots & p_{n-1,n-1}(t) & p_{n-1,n}(t) \\ 0 & 0 & & p_{n,n-1}(0{,}5) & p_{n,n}(t) \end{bmatrix}$$

And **A** represents the infinity generator matrix:

$$\mathbf{A} = \begin{bmatrix} -\lambda_0(t) & \lambda_0(t) & \cdots & 0 & 0 \\ 0 & -\lambda_1(t) & & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \cdots & -\lambda_{n-1}(t) & \lambda_{n-1}(t) \\ 0 & 0 & & 0 & -\lambda_n(t) \end{bmatrix}$$

The right diagonal of A represents the vector of transition rates, which is related to the diagrams shown in Figure 13. In the current work, this vector $\boldsymbol{v}$, named the transition vector, is one of the parameters used to evaluate the trained models as shown in Equation 18.

$$\boldsymbol{v} = [\lambda_0(t) \quad \lambda_1(t) \quad \cdots \quad \lambda_{n-2}(t) \quad \lambda_{n-1}(t)] \tag{18}$$

A Markov chain is formed when the likelihood of an event occurring depends only on the recent past and not on the distant past (BOLCH et al., 2006). In other words, a Markov process holds the attribute of memoryless, and its property can be deduced from Equation 17, which can be expressed mathematically as follow:

$$p\left(X_{t_{n+1}} \leq s_{n+1} \,\middle|\, X_{t_n} = s_n, X_{t_{n-1}} = s_{n-1}, \dots, X_{t_0} = s_0\right) = p\left(X_{t_{n+1}} \leq s_{n+1} \,\middle|\, X_{t_n} = s_n\right)$$

Also, the Markov process holds the semigroup property (KARLIN; MCGREGOR, 1955), which means that knowing the transition probabilities from states $i$ to $k$ and $k$ to $j$ at times $t$ and $s$, one learns the transition probability from $i$ to $j$ at time $t + s$. Thus, using the semigroup property, it is conceivable to estimate transition probabilities in the future based on prior records, or in other words, it is possible to develop a predictive model of accidents if the accident history of previous jobs is known.

Note that, in Equation 17, the analytical solution for the transition rates is not trivial, but some authors (CHUA; GOH, 2005) indicate that the accidents distribution (states versus time) can be associated with the Poisson function, which simplifies the Equation 16, since for any time t, $\lambda(t) = \lambda$. As a result, the pure birth model can be interpreted as a Poisson model with the following probability distribution:

$$p_{s(t)}(S) = \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \delta(S-n) \;\mid\; \delta(S-n) = \begin{cases} 1, se\ n = 0,1,2,\dots \\ 0,\ otherwise \end{cases}$$

Where,

- $s(t)$ is the state of safety events in time $t$;
- $\delta(S-n)$ is the Dirac delta.

So, the cumulative distribution function is described as:

$$P(\{t < t_1\}) = P(n(t) = 0) = \left.\frac{e^{-\lambda t}(\lambda t)^m}{m!}\right|_{m=0} = e^{-\lambda t}$$

$$P(\{t_1 \leq t\}) = 1 - e^{-\lambda t}$$

And the probability density function is given by:

$$F_{t_1}(t) = P(\{t_1 \leq t\}) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

Therefore, applying these concepts into the scope of the present study, the time of the first incidence of occupational safety on site follows an exponential distribution. As expected, projects with a history of higher accident rates (higher $\lambda$) have a larger risk of new accidents occurring at the start of the work (lower $t$).

## 5.3. Methodology

To better understand how to predict and help prevent construction accidents using past and future scenarios, the present study proposes a construction planning method oriented to worker safety based on random processes. To that end, this work's methodology adopts the notion of Knowledge Discovery in Databases (KDD), which is a process of discovering new knowledge and patterns in databases (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). In this paper, the new knowledge refers to the behavior of transitions between safety states during the construction phase of an infrastructure project and the database comprehends of past safety records of several construction projects of this type.

The proposed KDD procedure used in this work is illustrated in the flowchart shown in Figure 14. Note that there are 3 steps of data treatment, involving the tasks related to preprocessing, such as data transformation, and organization. After the preprocessing and transformation tasks, a decision gate is inserted to allow for the review of previous steps. In doing so, the proposed methodology occurs in cyclical workflows, going back when it is necessary. Another important point is the two moments to explore the data – the first one occurs with the raw data and the second one, after all preprocessing and transformation operations. This is important to avoid that the data mining operations produce inconsistent or unreal data.

After the dataset organization step, the next steps involve model training during data mining, and model evaluation during postprocessing. By evaluating the trained models, the user, with the support of the algorithm, can interpret the meaning of each model. For instance, after the evaluation step a trained model can assume a pessimistic or optimistic behavior related to safety events. Finally, if there is no need to review previous step, the proposed methodology ends a cycle with a new knowledge, which is the best models to predict a new project based on past projects. However, if a new project is concluded, then the history should be updated with new data collection and the process restarted. Therefore, the proposed methodology works in repetitive cycles with real-time data, which is a challenge for construction planning methods that deal with uncertainties (BI et al., 2015; FENG et al., 2022; GOH; ASKAR ALI, 2016; KIM; ELLIS, 2009; LI, Qian et al., 2018; MA et al., 2014).

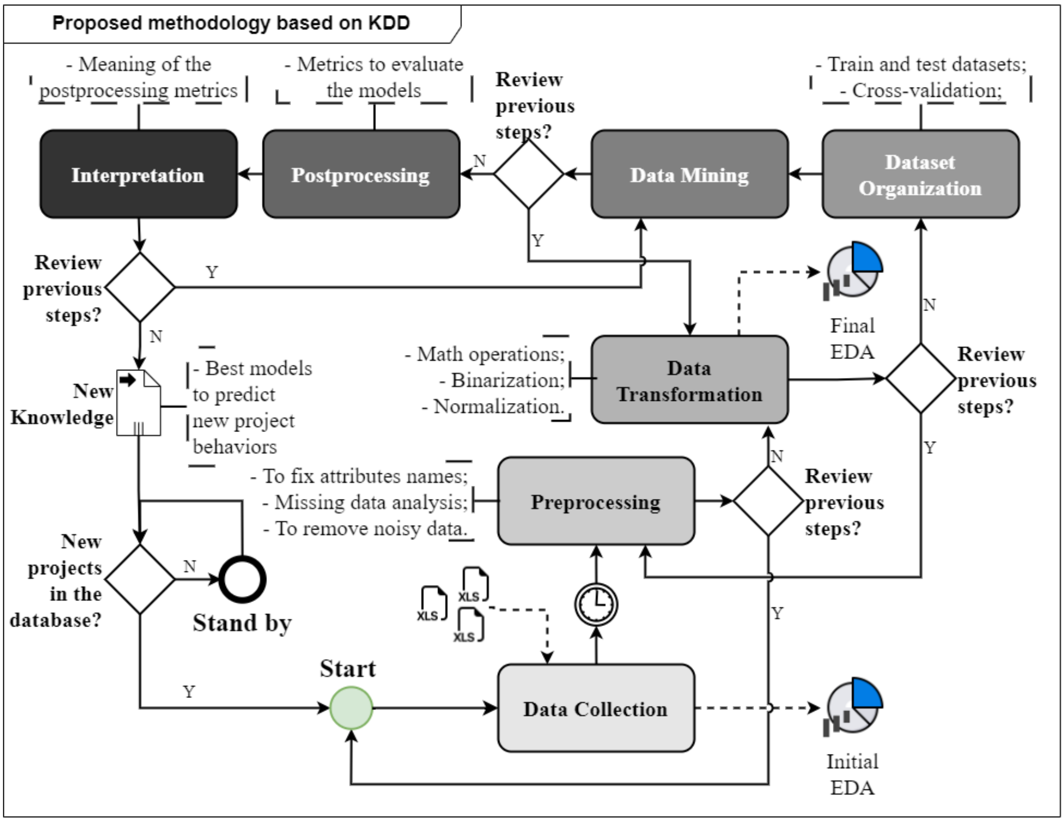



Figure 14 - Adopted methodology (Knowledge Discovery in Database) for the data mining process. (Adapted from Fayyad et al., 1996)

### 5.3.1. Data collection

The data collection step includes functions to extract the files from the database, to identify the attributes and instances in each file, and to organize the database resulting in the Exploratory Data Analysis (EDA). Performing an EDA is important to understand weak and strong points of a dataset, which can include irregularities and possible biased data (HILL, 2006). Also, in this paper, the EDA was used to gather data insights that helped to understand the database and the techniques that can be applied in the next steps. The functions were written in Python programming language using the Google Colab platform. This programming environment is useful to collaborate with other colleagues and code in distinct devices, but some attention should be paid in relation to data protection and privacy.

The dataset used in this work is composed of 39 renewable-energy power plant construction projects, the majority of which use wind and photovoltaic technologies. Since the available safety data is sensitive and confidential, the company's name is not mentioned. It is also worth noting that the database includes power plants ranging in size from 50 MW to 500 MW nominal capacity and located on four continents.

Also, the projects have distinct number of safety records, safety occurrences, and locations, which can be better understood in the exploratory analysis presented in the next section. Note that the number of projects included in this study is larger than that employed in earlier research with comparable goals (CHUA; GOH, 2005). Moreover, the safety data are provided in spreadsheets, typically as weekly reports.

To explore the dataset, ontologies theory was used. As discussed by Amaral et al. (2021), foundational ontologies are important to enhance the KDD methodology, specifically due to the insights that it can provide beyond the domain ontologies. However, the authors concluded that more practical studies are necessary to integrate the foundational ontologies with KDD techniques. Through this problem deep understanding that identifies relations and conditions inside the dataset, the researcher can find a better foundation for classification purposes and interpretation of results. In fact, with abstract topics, like risk and value, Sales et al. (2018) concluded that the ontological investigation help define the "deep connections" between them. In this paper, to better understand the uncertainties intrinsic in the dataset, a data taxonomy is proposed as shown in Figure 15.

The current study adopted domain ontologies, but the results indicate that the foundational ontologies may be useful to understand better the uncertainties and the risks events related to them.

The data conceptual model revealed that the scope of this study is limited to construction projects related to renewable energy, mainly related to wind and solar technologies. Also, it is possible to observe that there are three entities that describe a safety occurrence. First, the safety record that is registered by the user and contains the safety events counted during a time period. Through the record date and the counted events, the entity related to the safety event transitions is obtained. This represents the specie that will be predicted during the data mining process and be useful to adopt the stochastic modeling. With such a prediction, the construction planner would be able to understand the most likely moment where an accident could occur.

Also, note that safety records involve data that identifies the worker, company, among other sensitive information, and usually this kind of information is not available to the planning team due to data protection laws. Therefore, the proposed approach does not use any kind of confidential data that can interfere in the practical application of this method.

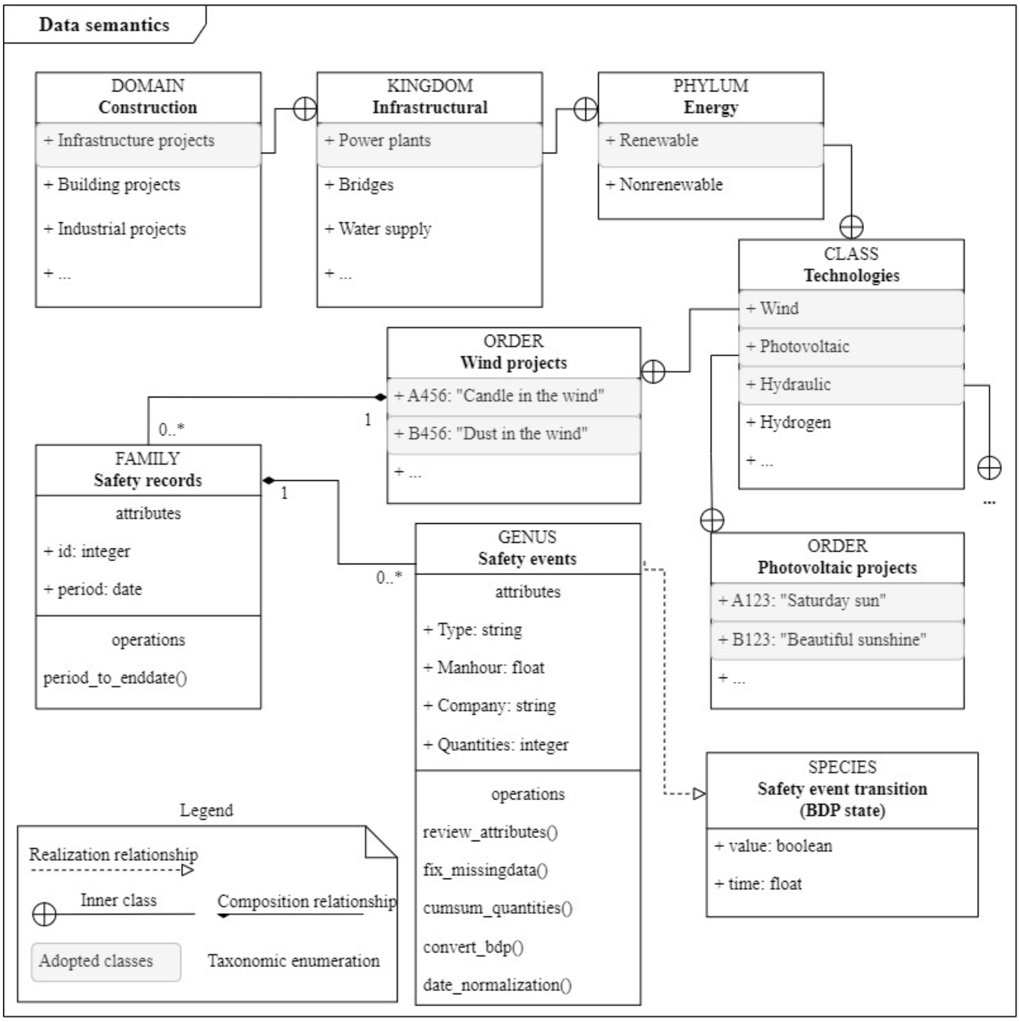



Figure 15 - Data conceptual model describing the relationships between data and attributes.

### 5.3.2. Data preprocessing

After data collection, the data from all projects were merged into one data table, named in Python as "*dataframe*", to facilitate the preprocessing and the other procedures. The preprocessing techniques used in this work were attributes renaming, attributes addition, data merging, missing data analysis, data replacement, data removal, and data type analysis. The first phase is the feature selection to identify the total number of attributes, unnecessary ones, incorrect names, among others.

During the merging function, attributes that were incorrectly named or misspelled names, such as “Accindents” instead of “Accidents”, were renamed and new attributes, such as “Period start date” and “Period end date”, were added to improve the unified data table. The unified dataset was, then, checked to identify noisy data, which can be related to missing values, unacceptable user inputs, among others. The missing data analysis is also important to avoid wrong data interpretations in the KDD process (SILVA; ZÁRATE, 2014). This analysis also involves the treatment of the missing data. Basically, the missing data can be removed, maintained, or replaced in the dataset. Considering the scope of the present study, the missing data can occur mainly due to human error, which means that the user can wrongly insert some value or forget to insert it. Therefore, in the proposed methodology, the missing data was managed in two ways: elimination and replacement with zero value, resulting in two unified datasets to comprehend if the sample size affects the results.

The last pre-processing function used in this study consisted of data type analysis to check if the values were consistent with the attribute data type described in Figure 15. In the negative case, the value data type was changed.

Table 7 shows the resulted data attributes after pre-processing techniques. It was assumed that the missing data represents the nonoccurrence of safety events, since it is likely that the user forgot to insert a value in the weeks with no safety occurrences. Therefore, the replacement with zero value was used in the proposed methodology.

Table 7 - Dataset attributes related to the prediction model of safety events.

| Data attributes | Taxonomy hierarchy | Description | Type of input | Type of variable |
|---|---|---|---|---|
| **c.project** | ORDER | Energy infrastructure construction project represented by an abbreviation | Automatic | String |
| **c.safid** | FAMILY | Number that identifies the safety record in the project | Automatic | Integer |
| **c.period** | FAMILY | Period in which the safety occurrences were counted in the safety record | Automatic | String |
| **c.enddate** | FAMILY | End date of the record period | Automatic | Date |
| **c.startdate** | FAMILY | Start date of the record period | Automatic | Date |
| **c.safobs** | GENUS | Occurrences presented in the safety record related to safety observations | Manual | Integer |
| **c.nearmiss** | GENUS | Occurrences presented in the safety record related to near miss events | Manual | Integer |
| **c.fstaid** | GENUS | Occurrences presented in the safety record related to first aid events | Manual | Integer |
| **c.accid** | GENUS | Occurrences presented in the safety record related to accidents | Manual | Integer |
| **c.fataccid** | GENUS | Occurrences presented in the safety record related to fatal accidents | Manual | Integer |

### 5.3.3. Data transformation

The third step in the KDD process refers to data transformation. Here, math operations between attributes, data conversion (e.g., date to integer value), data grouping, normalization, and binarization are the main subprocesses to prepare the database for data mining.

The first stage of data transformation is data grouping, in which the instances related to the same project in the same period were summed. Then, to understand the evolution of safety events over the time, a cumulative sum was carried out. After that, a binarization process was developed to convert the safety occurrences into two possibilities: safety event occurrence (1) or no safety event (0) and, so, the BDP states were created. This is important to prepare the dataset to the pure birth process modeling, in which the allowed instances are one single birth (1) or no birth (0). The binarization, thus, produced the actual species, which are related in the data taxonomy with the transition between safety states.

Note that this procedure removes the magnitude of safety events that occur in a period, but since this study is based on weekly safety reports, it can be assumed that after the first event it is unlikely that another would occur in such a short period of time. To summarize these operations, an example of a sequence of data preprocessing is presented in Table 8 and Table 9.

Table 8 - Example of dataset after the preprocessing operations.

| c.project | c.enddate | c.accid |
|---|---|---|
| **ABC1** | 02/08/2022 | 0 |
| **ABC1** | 09/08/2022 | 1 |
| **ABC1** | 09/08/2022 | 1 |
| **ABC1** | 16/08/2022 | 0 |
| **ABC1** | 23/08/2022 | 1 |
| **ABC1** | 23/08/2022 | 0 |
| **DEF2** | 26/08/2022 | 0 |

Table 9 - Example of dataset after the transformation operations.

| c.project | c.enddate | c.accid (sum) | c.accid (cumulative sum) | c.accid (transitions) | c.accid (BDP states) |
|---|---|---|---|---|---|
| **ABC1** | 02/08/2022 | 0 | 0 | 0 | 0 |
| **ABC1** | 09/08/2022 | 2 | 2 | 1 | 1 |
| **ABC1** | 16/08/2022 | 2 | 2 | 0 | 1 |
| **ABC1** | 23/08/2022 | 3 | 3 | 1 | 2 |
| **DEF2** | 26/08/2022 | 0 | 0 | 0 | 0 |

In the next phase of data transformation process, this work´s methodology adopts data conversion and normalization. Data conversion consists of converting date values (e.g., “1900-01-01”) into integer values (e.g., “16720200”). The integer values represent the number of days from a reference date (“1900-01-01”) until the converted date. In doing so, the integer values can be normalized through a MinMax technique, the maximum being 1 (one) and the minimum being 0 (zero). It is worth noting that the entire date was saved because it will be useful in future studies, such as the impact of weather conditions.

Therefore, the final dataset that was used for data mining purposes is composed of four main attributes: project name, normalized period end date, safety transitions (births), and birth process states. Table 10 shows an example of 5 instances of this dataset.

Table 10 - Example of the preprocessed and transformed dataset used in the data mining process.

| c.project | c.enddate_norm | c.accid (transitions) | c.accid (BDP states) |
|---|---|---|---|
| **ABC1** | 0.20 | 0 | 0 |
| **ABC1** | 0.27 | 1 | 1 |
| **ABC1** | 0.35 | 0 | 1 |
| **ABC1** | 0.43 | 1 | 2 |
| **DEF2** | 0.45 | 0 | 0 |

With this final dataset, a new exploratory study was carried out to visualize the available data for utilization in the processing techniques. The step-curve and the histogram of pure-birth states were plotted for each project. Through the graphs, projects with noisy data and data anomalies could be visually identified and discarded from the dataset. Also, to focus on construction phases related to civil engineering activities, such as earth movements, foundation pouring, among others, the final datasets were reduced to instances reported up to 60% of construction completion ($t \leq 0.6$). This time criterion is defined after the analysis of the projects´ timelines included in the dataset.

### 5.3.4. Data processing

Then, the KDD stage related to data processing was initiated by dividing the dataset into training and test datasets. This train-test split followed a project-oriented division with cross-validation. In other words, for the dataset with 39 projects, $X$ % of them was applied to separate the training dataset and $(1 - X)$ % to the testing set, as shown in Figure 16. Then, from the training projects, $N$ folds were organized to be used during the data mining for training and test purposes. After the initial training and testing, a final evaluation was done with the $(1 - X)$ % of projects separated for testing. The trial values of $X$ were 67% and 89%, and the trial values of $N$ were 3 and 6. This variation is important to understand its effects in the data training. Note that the cross-validation is useful to select models that predict safety events in new construction projects.

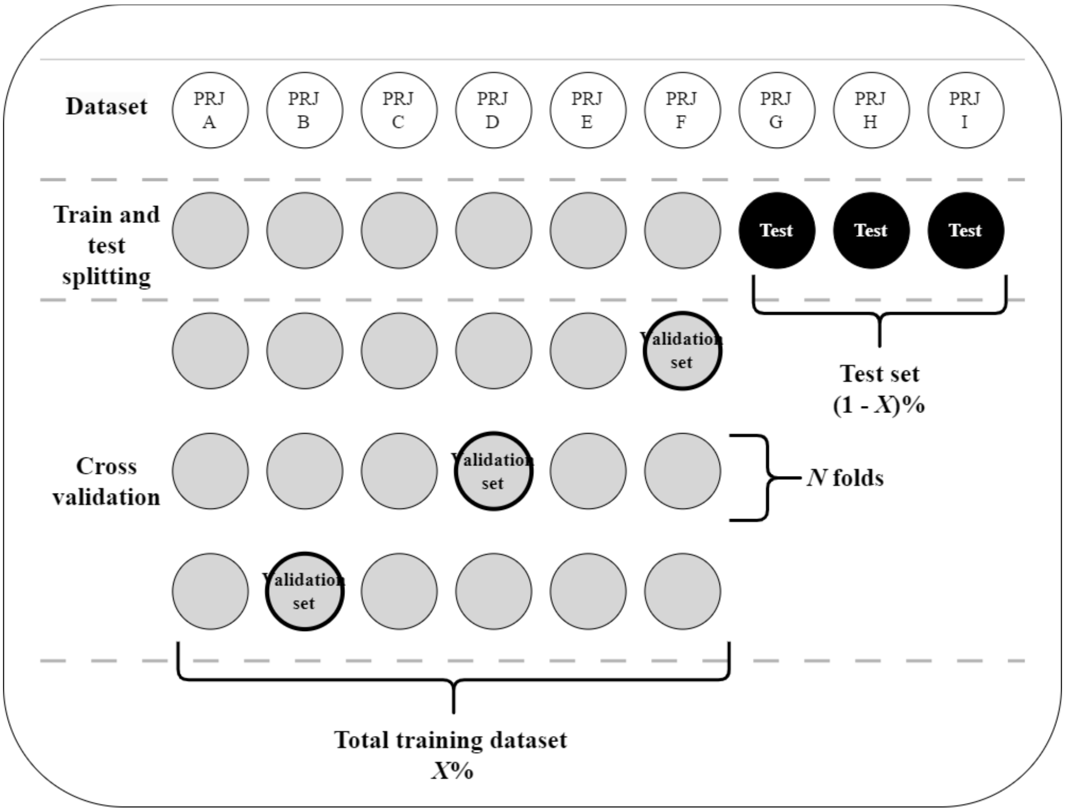



Figure 16 - Dataset division and organization including the cross-validation technique.

However, it may be not useful for ongoing projects, because the dataset can be understood as time series and traditional cross-validation techniques would not be adequate since it can eliminate the meaning of time sequence. Therefore, for that specific case, the walk forward validation or adapted cross-validation techniques (time series cross-validation), would be more appropriate to deal with time series (PARDO, 1992).

Following the proposed KDD process, the next step consists of data processing, which is divided into two parts: probabilistic distribution and stochastic solutions. The first part consists of fit the training dataset to the best probabilistic distribution performing the Kolmogorov-Smirnov test for goodness of fit. This specific metric was chosen because the best fit cannot result in a normal distribution and thus, other tests, such as the Student's T-Test, would not be applicable (MASSEY, 1951). During this part of data processing, two parameters were defined to control the learning process in KDD: the number of samples for the Kolmogorov-Smirnov test (ksN) and the threshold p-value for the Kolmogorov-Smirnov test (KS_criteria) that rejects the null hypothesis. For the former it was assumed the value of 1000 and the latter was fixed at 5%.

Therefore, using the training datasets that came from cross validations, the probabilistic solution results in a trained model that contains the best distribution for each construction time and for each step in the cross-validation process. Note that the attribute used in this part is the histogram of transitions between safety states and so, the model returns the transition probabilities as described in the right-hand side of Equation 17.

The second part in the data processing step is the stochastic solution. Using the Python package BirDePy (HAUTPHENNE; PATCH, 2021), this solution included the parameter estimation for each project training dataset, assuming that the stochastic process is pure birth or Poisson, and time continuous. So, the main parameter is the estimation method, defined as the direct numerical maximization (dnm), also known as Maximum Likelihood Estimation (MLE). The initial guess for the parameter value, and the parameter bounds are also possible hyperparameters to use, but to simplify the methodology, the values were defined as fixed. The initial guess ($p_0$) assumed the value of $1e^{-6}$ and the parameter bounds ($p_{bounds}$), $[0, 1e^{+6}]$.

So, to evaluate the training model, simulations were carried out to predict the global behavior of safety events transitions in a future construction project. Using the probabilistic model, 1000 simulations were performed, and using the stochastic model, 1 simulation was carried out because of the excessive processing time required. The simulations resulted in a predicted column related to the cumulative sum of safety events transitions, which is the attribute used to train all models. Note that each trained model passes by three distinct evaluations, two during the validation step and one with the final dataset. The evaluation with cross validation is based on the confusion matrix, comparing the actual column in the test fold (inside the initial training dataset) with the predicted one. Basically, there are only two possible values for a prediction: 1 (safety event transition occurs) and 0 (nothing occurs), and thus, the confusion matrix is 2x2, composed of true positive, false positive, false negative and true negative values. So, the following performance metrics were used to rank the trained models:

- Precision score (PRC): to understand from the positive predictions, how many are truly positive. Thus, this would suggest that the model is able to correctly predict safety events transitions.

$$PRC = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

- Recall score (REC): to understand from the positive cases, how many are predicted positive by the model. Thus, this would suggest that the model is capable of quantitatively predict safety events transitions.

$$REC = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

- Specificity score (SPC): to comprehend from the negative predictions, how well the model is predicting the nonoccurrence of safety events transitions. Thus, this would attest to the model´s ability to avoid false predictions of safety events transitions.

$$SPC = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

- F1 score (F1S): to understand in combination the precision and sensitivity scores, through the harmonic mean. It would capture the good models based on both metrics.

$$F1S = 2 \times \frac{PRC \times REC}{PRC + REC}$$

- PR-AUC score: to comprehend, in case of imbalanced classes (negative class is more frequent than positive class), the model performance in predicting positive cases. It deals with the area under the curve generated by the precision and recall scores along several executions of the training.

### 5.3.4. Data postprocessing

After the best models were ranked according to the metrics, the final evaluation was carried out to compare the two solutions. The final evaluation consisted of examining the predicted transition probability matrix, created by each model, with the actual one originated from the initial test dataset. As shown in Equation 17, the transition matrix is useful to understand the stochastic process through the birth rates between states. By creating a predicted stochastic process and comparing with the real observed (actual) process, the model is evaluated by the right diagonal that represents the transition rates for each time. Note that to understand the evolution of safety events transitions, the proposed methodology uses the probability of transition in a time interval ($\Delta t = t < t_1$), instead of a time instant ($t$). The following equation exemplify the matrix used as ground truth and was originated by the right diagonal presented in the Equation 6 applied for each time interval considered.

$$E_{n,m} = \begin{bmatrix} \lambda_{0|t<t_1}(t) & \lambda_{1|t<t_1}(t) & \cdots & \lambda_{m-1|t<t_1}(t) \\ & \vdots & & \\ \lambda_{0|t<t_n}(t) & \lambda_{1|t<t_n}(t) & \cdots & \lambda_{m-1|t<t_n}(t) \end{bmatrix} \tag{19}$$

Note that the construction manager would know from the past what was the birth rate between safety events in each period of the construction duration. The adopted time periods refer to a quarter of the total time, which means the following normalized times: 0.15, 0.30, 0.45, and 0.60. Through the data visualization in the data transformation step in the KDD process, it is observed that this period refers to approximately a quarter in a year. However, the ground truth matrix does not allow the prediction of birth rates for states that were not observed before, i.e., the exact solution can only repeat the past to predict the future. So, the solutions provided in the trained models were used to capture random events that can predict future transitions given a time period even if this type of transition did not occur before. The probabilistic model, in theory, should result in more controlled future scenarios, since the fitted distribution are semi-deterministic equations, i.e., given the same input, the equation returns the same output in terms of transition probability, but does not always return the same output in terms of occurrence of safety events transitions. However, due to the higher level of chaos associated with stochastic processes, the stochastic model should be able to predict totally unexpected events, such as natural disasters.

To evaluate the matrix similarity between the trained and exact solutions, the sum of local distances between them is calculated using the following equation:

$$\boldsymbol{MS} = \sum_{\boldsymbol{j}}^{\boldsymbol{m}} \sum_{\boldsymbol{i}}^{\boldsymbol{n}} \left| \boldsymbol{E}_{ij} - \boldsymbol{S}_{ij} \right| \tag{20}$$

Where,

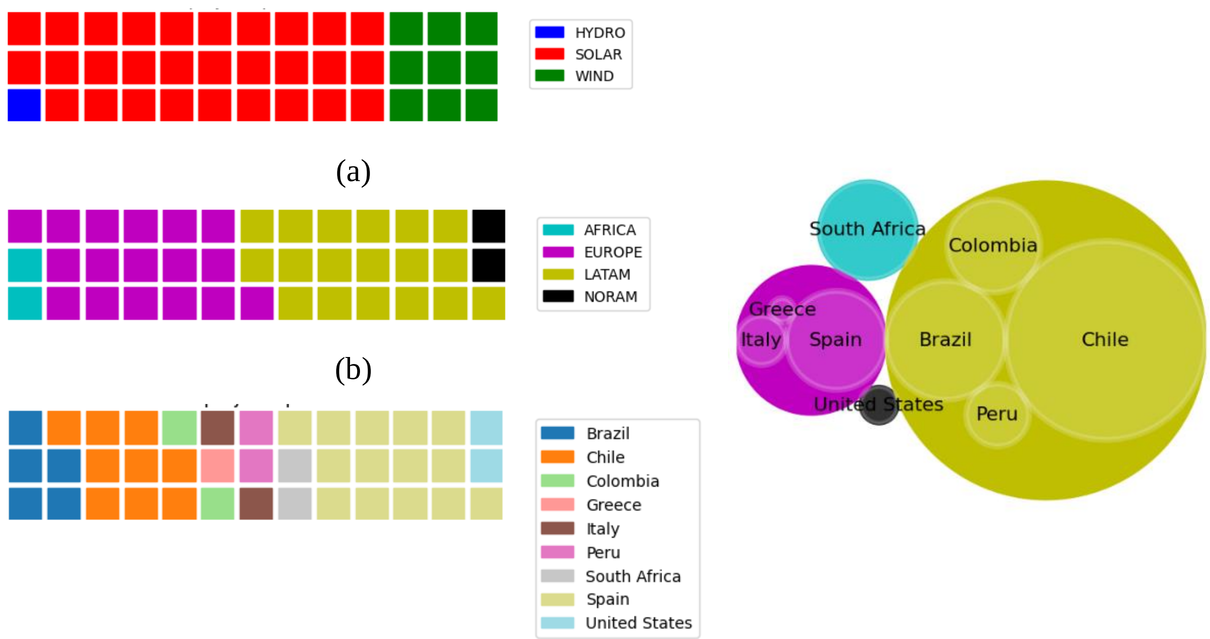
- $\boldsymbol{MS}$ is the similarity value matrix;

- $n, m$ are, respectively, the number of rows and columns of the matrices;
- $\boldsymbol{E}_{ij}$ is the value in position $ij$ of the exact solution matrix $\boldsymbol{E}$;
- $\boldsymbol{S}_{ij}$ is the value in position $ij$ of the trained solution matrix $\boldsymbol{S}$.

## 5.4. Results and discussions

The proposed methodology was applied to the available database, which included projects from various countries, from different continents and three types of renewable energy technologies. It is important to note that all databases originated from a single company that has operations in numerous nations worldwide. The results related to model training are described next.

The first Exploratory Data Analysis (EDA) is shown in Figure 17, describing the number of projects, files, and instances per technology (taxonomic class), continent, and country. With the total of 39 projects, 74% are photovoltaic projects, 23%, wind projects, and 3% hydroelectric. Most of the projects are in Europe (44%) and Latin America (46%), concentrated in Brazil (13%), Chile (23%), and Spain (36%). From the total number of projects, 6511 files containing safety records were collected, with a total of 31469 instances related to safety events. The total number of distinct attributes is 146 identified in all files, but only 10 were considered in the analysis due to General Data Protection Regulation restrictions.



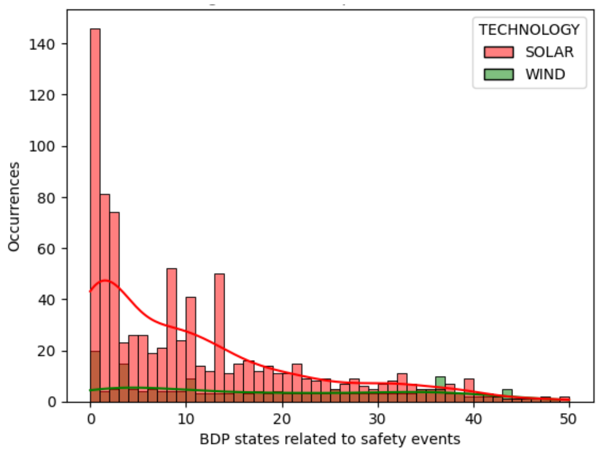
(c) (d)

Figure 17 - Exploratory data analysis to understand the technologies (a), regions (b), and countries (c) associated with the safety dataset and its instances (d).

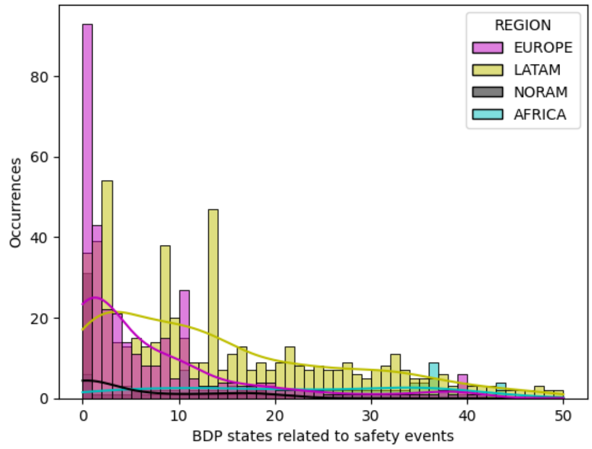
This initial data analysis reveals that developing countries are the ones with the highest number of safety records, whose cause is not in the scope of the present study. Moreover, this initial EDA reveals that the data may be biased for the Latin America (LATAM) projects, because they constitute a large parcel of the total number of instances. The same occurs with the solar projects, that are in majority. Therefore, the cross-validation technique is adopted to reduce the level of bias related to these points. Also, this analysis revealed that the database is composed of unbalanced classes, since the nonoccurrence is more frequent than the occurrence of a safety event and, thus, the performance metrics must deal with this unbalance between classes.

The second moment of the EDA is shown in Figure 18, describing the histogram of states of safety events organized by technology, continent, and country. Note that this second analysis occurs after data preprocessing and transformation steps and, thus, some technologies, continents, or countries may have been eliminated due to the cleaning or removal process. The histograms reveal, as expected, that initial states of safety events are more frequent and after 15 states, the Kernel Density Estimation (KDE) curve tends to decrease to zero. Note that some projects contain 50 safety events, which represents more than one safety event per week in a construction project that lasts 1 year. Moreover, the photovoltaic projects tend to have a smaller percentage of safety events when compared to wind projects, since the KDE curve is more frequent in the beginning of x axis.

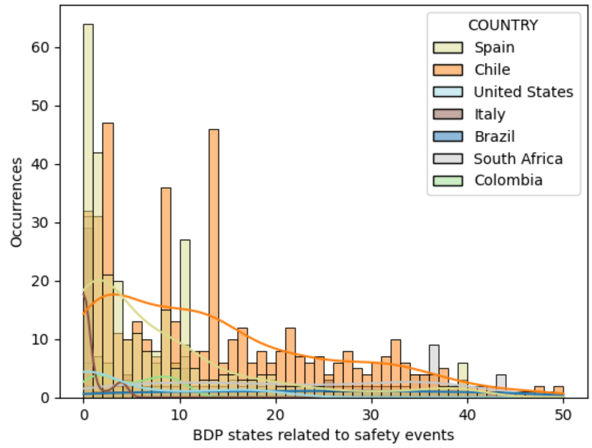
In terms of location, the KDE curves reveal that the European projects have more frequent zero safety events than the LATAM projects since the curve reaches its peak near 3 safety events. Also, the LATAM projects present distinct KDE curve behaviors. It is worth pointing out that the number of instances can affect these curves, but for an initial data analysis the histograms provide enough information to identify the possible bias that the data can assume.




(a)




(b)




©

Figure 18 - Exploratory data analysis to understand the occurrence of safety states in the dataset, according to the technology (a), region (b), and countries (c).

In sequence, the outcome from the data mining is provided. The best probabilistic and stochastic model for each training and testing combination and its results and evaluation in the cross-validation step are briefly represented in the experiment plan detailed in Table 11.

From a total of 31 trained probabilistic models, 13 (42%) were pre-evaluated and selected for the final testing, which reduced to 6 (19% of total) selected models. For each model, 1000 simulations were carried out to compare with the actual dataset in order to extract the mean score values. Note that a convergence test was used to determine the number of simulations, and the results demonstrated that after 1000 simulations, the results were essentially the same.

The best models were the ones with lower values for MS and higher values for PR-AUC and F1-Score in each group of the cross-validation datasets, as listed in Table 11. Note that even the not trained model achieves a recall score higher than 0.3, which indicates that the predictions performed quantitatively bad. However, the model's specificity resulted in scores higher than 0.8, which means that the models are not doing false predictions of safety events transitions.

In general, as observed in some predicted step curves (Figure 19), the probabilistic models tend to be optimistic –almost all curves are beneath the actual step curve. The reason for that may be related with the fact that the distribution functions tend to fit better with the safety events transitions with higher occurrence (histogram peak), ignoring random events that can create another peak in advanced construction phases. Thus, the probabilistic model would be more appropriate for construction projects with lower level of uncertainties when compared with the past projects.

Moreover, there is no clear conclusion to what the best distribution function would be. As suggested by some authors (ABOURIZK; HALPIN, 1992), the beta distribution could be an adequate function to describe safety events, but the results of this work do not confirm this assumption. On the contrary, similarly with the conclusions of Tesfaye et al. (2015), several functions can be adequate to describe the project completion time. In fact, future studies should incorporate more possibilities, beyond the sixty distributions tested in the present work.

For the stochastic models, from a total of 588 trained models (294 related to pure birth and 294 to Poisson process), 266 (45%) were pre-evaluated and selected for the final testing, which reduced to 4 selected models (0.7% of the total).

Table 11 – Experiment plan with the best models after the cross validation and model evaluation.

**Experiment plan – Best models**

| **Model** | **Training and testing datasets** | | | | | **Probabilistic model** | | | | **Model evaluation** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | Main division | | Cross-validation | | | Parameters | | Trained model | | During CV (mean scores) | | | | | | Final test | |
| | test projects | test projects | folds | test projects | training projects | ksN | KS p-value threshold | Best distribution | KS p-value | PRC | REC | SPC | F1S | PR-AUC | Status | MS | Status |
| 3 | 9 | 18 | 3 | 6 | 12 | 1000 | 0.05 | genpareto[a] | 0.8686 | 0.651 | 0.223 | 0.907 | 0.183 | 0.662 | Passed | 7.69 | OK |
| 8 | 9 | 18 | 6 | 3 | 15 | 1000 | 0.05 | genpareto[a] | 0.7914 | 0.642 | 0.291 | 0.935 | 0.271 | 0.576 | Passed | 8.84 | OK |
| 13 | 9 | 18 | 9 | 2 | 16 | 1000 | 0.05 | foldnorm[b] | 0.7194 | 0.698 | 0.233 | 0.830 | 0.233 | 0.723 | Passed | 9.54 | OK |
| 21 | 3 | 24 | 3 | 8 | 16 | 1000 | 0.05 | halfnorm[c] | 0.6076 | 0.716 | 0.109 | 0.866 | 0.183 | 0.742 | Passed | 11.45 | OK |
| 23 | 3 | 24 | 4 | 8 | 16 | 1000 | 0.05 | halflogistic[d] | 0.5634 | 0.643 | 0.124 | 0.908 | 0.180 | 0.615 | Passed | 7.75 | OK |
| 27 | 3 | 24 | 6 | 4 | 20 | 1000 | 0.05 | halflogistic[d] | 0.4428 | 0.723 | 0.140 | 0.859 | 0.228 | 0.720 | Passed | 12.11 | OK |

| **Model** | **Training and test datasets** | | | | | **Stochastic model** | | | | | **Model evaluation** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | Main division | | Cross-validation | | | Parameters | | | Trained model | | During CV (mean scores) | | | | | | Final test | |
| | test projects | training projects | folds | test projects | training projects | p0 | p0_bound | estimation method | Pure Birth parameter | Poisson parameter | PRC | REC | SPC | F1S | PR-AUC | Status | MS | Status |
| 24 | 9 | 18 | 3 | 6 | 12 | 1e-6 | [0, 1e6] | dnm | 1.84e-03 | | 0.928 | 0.551 | 0.780 | 0.685 | 0.871 | Passed | 4.4 | OK |
| 75 | 9 | 18 | 3 | 15 | 3 | 1e-6 | [0, 1e6] | dnm | 1.73e-03 | | 0.760 | 0.528 | 0.500 | 0.623 | 0.755 | Passed | 9.3 | OK |
| 173 | 3 | 24 | 6 | 4 | 20 | 1e-6 | [0, 1e6] | dnm | 1.90e-03 | | 0.694 | 0.714 | 0.358 | 0.704 | 0.688 | Passed | 3.9 | OK |
| 217 | 3 | 24 | 3 | 8 | 16 | 1e-6 | [0, 1e6] | dnm | | 6.17e-03 | 0.740 | 0.751 | 0.385 | 0.732 | 0.736 | Passed | 2.1 | OK |

[a] Distribution defined by a generalized Pareto continuous random variable.

[b] Distribution defined by a folded normal continuous random variable.

[c] Distribution defined by a half-normal continuous random variable.

[d] Distribution defined by a half-logistic continuous random variable.

0.716 This color represents the best metric scores.

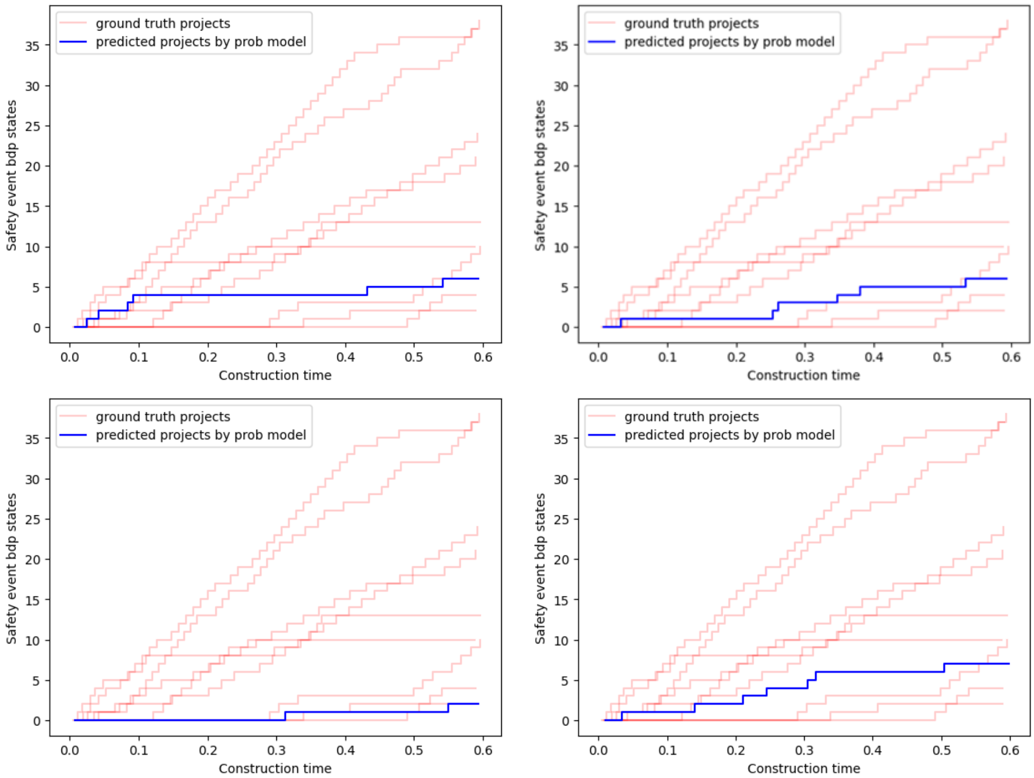



Figure 19 - Prediction results related to the probabilistic models.

Note that the higher number of models compared with the probabilistic model is due to the fact that for each training project, one stochastic model is estimated, which is not true for the probabilistic model. The probabilistic solution requires a large number of projects to have a representative histogram. To compensate this discrepancy and to reduce the processing time, only one simulation is carried out for each stochastic model.

Using the same selection criteria of the best probabilistic models, it is observed, in Table 11, that the recall scores are better than the probabilistic solution and, consequently, the F1-score is also better. In this sense, the stochastic solution tends to produce better predictions for positive classes (safety events transitions) both qualitatively and quantitatively. However, the specificity score shows lower values compared with the probabilistic solution, which can indicate that the stochastic solution produces more false positives, or, in practical terms, it is more pessimistic. Depending on the application, construction planning could take the more optimistic or more pessimistic approach. In the context of infrastructure projects, which are typically uncontrolled and have a higher level of uncertainty, the pessimistic solution should be better. However, if the values of specificity are too low, this would indicate that the model is being too pessimistic and, thus, not realistic.

When compared to the probabilistic solution, the stochastic models present better values of matrix similarity, which can suggest that the birth rates were better predicted. In fact, the predicted step curves seem to be more similar with the actual ones (see Figure 20) for the stochastic solution and, even when there are few safety events transitions, some stochastic models predicted similar step curves. This adaptability indicates that maybe the stochastic models can also represent optimistic scenarios in the construction projects, it will depend on the trained project. Using the stochastic models, the planner will have a family of possible predictions, but using the probabilistic ones, she/he will have just one standard family member (i.e., the taxonomical level that represents the safety record) that can better represent the others.

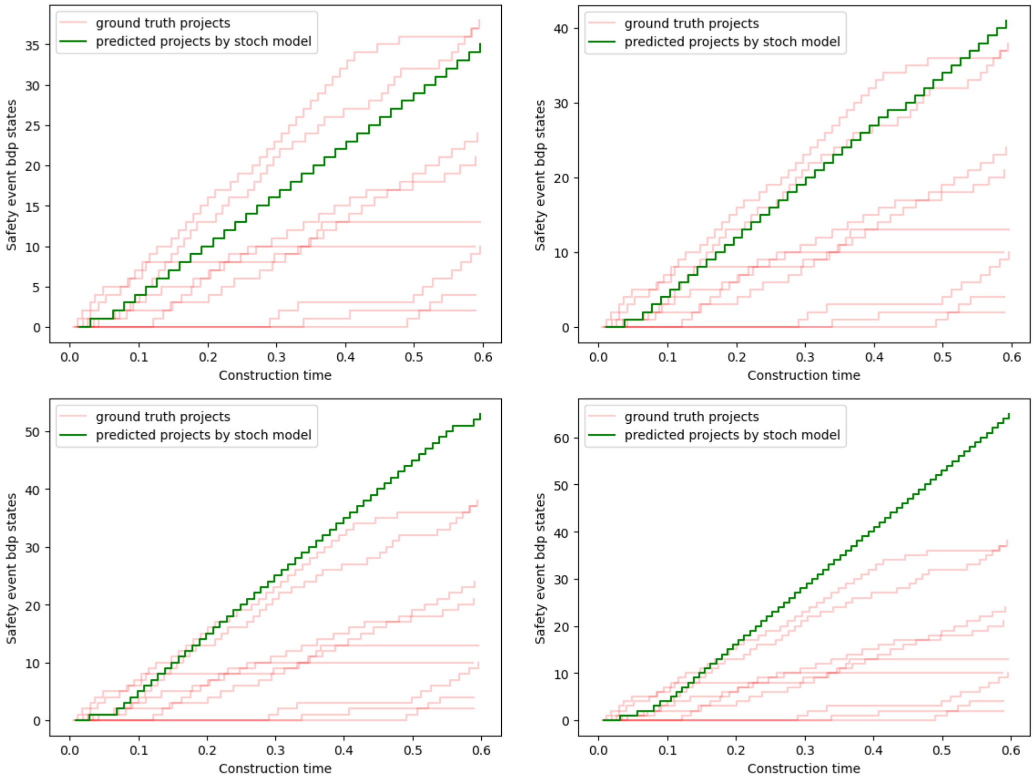



Figure 20 - Prediction results related to the stochastic models.

In terms of uncertainty, as expected, the results suggest that the stochastic model captures a higher level of uncertainty when compared to the probabilistic solution. Therefore, as the conclusions given by Zhang et al. (2021), the stochastic models seem to be more useful for construction projects that have more chances of having random events. Comparing both solutions, as shown in Figure 21, it can be clearly seen that the stochastic models capture more safety events than the probabilistic solutions.

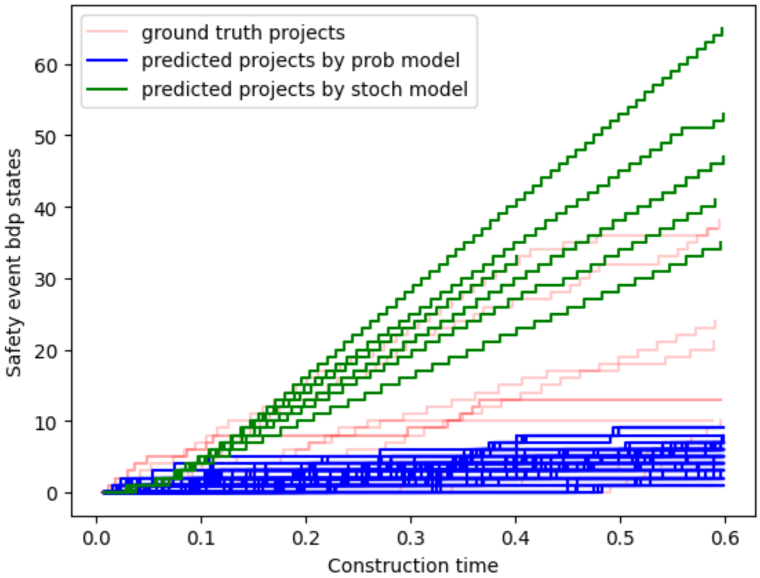



Figure 21 - Comparison betwen the prediction results from the probabilistic and stochastic models.

Next, a discussion related to the stochastic process adopted in each solution provided. From the total number of 266 selected models, 97 (36%) are pure-birth process estimation, and only 1 of the best 4 models is estimated by a Poisson process. Differently from other studies (CHUA; GOH, 2005; ZHANG, Zhenhao; LI; YANG, 2021), it seems that the Poisson is not the best process to estimate safety occurrences in construction projects. From the results, it can be concluded that the Poisson process is more optimistic than pure birth. Again, the level of uncertainty may explain the reason for that: the pure-birth process works like a random walk, which is more chaotic, and the Poisson process, even in the stochastic part, captures less uncertainties related to safety events transitions.

## 5.5. Conclusions

The construction industry has been tested many times during the recent crisis events, and that worker safety is a historical issue to be dealt with. Therefore, the current study proposes a construction planning approach to predict the transitions between safety states, capturing random events with high level of uncertainties. To that end, the authors reviewed the literature and concluded that most planning methods do not handle uncertainties with random variables related to work accidents, which generally results in optimistic scenarios. The ones that deal with random events usually adopt information models and machine learning techniques to manage the safety plan in construction projects. Stochastic processes, which are part of a classical theory that have the purpose to capture the random nature of the construction industry, have not yet being reported in the literature. The present study´s hypothesis is that combined with the currently available computational power, stochastic processes can be a good alternative to prepare infrastructure projects to deal with totally unplanned events, like a work accident that had never occurred before, focusing on the safety worker issue.

The proposed methodology based on the KDD concept involves data treatment steps to optimize and improve real construction datasets, preparing them for training models that will predict the species of safety events transitions. Based on the previous methods adopted in the literature, the current work presents two solutions: one based on probabilistic distribution and the other on stochastic processes. By applying these solutions in a real dataset with 39 projects, the results revealed that the probabilistic approach tends to assume optimistic scenarios for real projects, while the stochastic approach, more pessimistic ones. As observed in the step curves, both solutions fit relatively well in some real projects, but with different best scores. The probabilistic solution presents the best specificity scores – highlighting its ability to not predict a false positive class – and the stochastic solution presents better precision scores, even with only one simulation – highlighting its ability to predict a true positive class. Also, combining precision and recall scores, the stochastic solution results in better F1 and PR-AUC scores, and analyzing the generator matrix of both solutions, the stochastic solution results in better values of MS score as well. Therefore, the results indicate that quantitatively the stochastic solution provides better predictions than the probabilistic approach.

It is also worth pointing out that the pure-birth estimation seems to return better scores when compared to the Poisson model estimation (both estimations in the stochastic solutions), which leads to the conclusion that when the level of uncertainty is high, the Poisson model is able to capture this level of randomicity. Thus, as suggestion for future works, it is recommended the use of walk-forward validation to better understand the differences between Poisson and pure-birth models and to make the approach useful for ongoing projects, since the training and testing datasets will be done with incremental time. Also, to better categorize the qualitative results in terms of pessimistic, realistic, and optimistic model, future studies can investigate the application of clustering techniques in the stochastic models, expanding the approach to more projects and countries, too. Finally, as highlighted in the literature, the stochastic solution should also be tested with recently developed AI methods.

# 6
# A hybrid solution to consider the stochastic nature of safety incidents on project delays in construction planning methods

Paper ready to be submitted by Cristiano S. T. do Carmo and Elisa D. Sotelino to a peer-reviewed international journal.


## Abstract

The construction sector is inherently full of uncertainties, such as political and social worries, and infrastructure projects are considerably more chaotic due to their scale. This context brings to light construction safety hazards, which are typically overlooked in construction planning methods. Using the stochastic quasi birth and death process and neural network models, the current study proposes a new construction planning method that considers safety events and their effects on project delay. A literature study revealed that there is a scientific requirement for practical research related to infrastructure that connect technical applications with industry practices. As a result, the goal of this work is to cover some of the research gaps mentioned in the literature by employing a real-life database and applying it to energy infrastructure projects. The results show that ignoring safety variables leads to the assumption that safety occurrences will occur; however, by applying the stochastic solution, construction planners can better understand the implications of safety events on delay events and vice versa. In addition, for live planning, this study recommends the use of statistical and neural network algorithms capable of forecasting bivariate time series. The validation scores show that the neural networks model performed nearly twice as well as the statistical method during the initial period of the construction phase.


## 6.1.
## Introduction

Unlike other industries, which have a controlled work environment, civil construction is well-known for its dangerous workplace (TAM; ZENG; DENG, 2004). In 2021 in Brazil, according to the Brazilian Statistical Yearbook of Works Accidents, it was responsible for the sixth highest number of accidents when compared to all economic activities in the country. This is often due to the terrible condition for workers the high cost for constructors due to the insurance premiums (HINZE; DEVENPORT; GIANG, 2006) and employee absence, that reduces team productivity. In fact, Waehrer et al. (2007) concluded that an extra cost of $1.36 billion (2002 dollars) is due to construction work-related accidents.

In times of crisis, however, the construction industry is under a lot of pressure to increase productivity, since it accounts for a large percentage of countries' gross domestic value – around 13% at global level (McKinsey, 2020). This scenario of cost and time pressure, following Pinto et al. (2011), can result in worst safety performance due to relaxing rules and processes. The current study focuses on safety management issues to avoid the increase of work accidents and, consequently, losses in terms of workers, and extra-costs for the company.

Related to the safety impacts in project management, Soltani and Fernando (2004) have already found that a safest path for the worker can reduce in almost 30% the construction activity cycle and process times. Also, according to Koehn and Musser (1983) apud Zhou et al. (2015), safety regulations can result in construction cost reduction, from 2.8% to 1.4%. However, following Carmo and Sotelino (2023), most planning methods currently adopted in construction projects do not consider the safety variable as critical for the schedule validation.

Thus, the current study proposes a new construction planning method to evaluate the effects of safety incidents on construction delays, using the fundamental stochastic theory of Quasi Birth and Death Processes (QBDP) and machine learning techniques. In fact, Carmo and Sotelino (2023) verified that the use of a stochastic model for safety occurrences predictions can perform better than the forecasting models adopted so far, such as statistical methods and Poisson model.

The paper is structured as follows. The first section discusses the context in which this study fits. The second section provides the theoretical foundation needed to comprehend the stochastic processes used in the proposed method, as well as the application of such theory in construction project management. The third section describes the methods used in the current work, such as the Markov transition diagram and the computational techniques used to process real-project datasets. The fourth section presents the findings and discussions related to the incorporation of such methods into a real-world database of energy infrastructure construction projects. Finally, the fifth section provides a summary of conclusions, limitations, and future work suggestions.

## 6.2. Theoretical framework

### 6.2.1. Stochastic processes in project management

A Systematic Literature Review (SLR) was conducted using the Scopus database to identify the important research works, their contributions to the body of knowledge, and the research gaps concerning the use of stochastic process theory in construction project management. The adopted literature review approach consists of five key steps: research questions, study location, study selection and assessment, analysis and summary, and reporting and interpretation (KHAN et al., 2003), which are discussed in Carmo and Sotelino (2023).

The SLR questions were:

- What are the primary contributions linked to stochastic processes in construction management in the literature?
- How were the theories implemented in the studies?

The search terms were subsequently formulated for use in the database engine, with the goal of achieving the greatest number of works relevant to the issue that could answer, at least partially, the inquiries. The search query included the following terms: TAK("STOCHASTIC* PROCESS*" OR "STOCHASTIC* MODEL*" OR "STOCHASTIC* ANALY*") AND TAK("MANAGEMENT") AND TAK("CONSTRUCTION PROJECT" OR "CONSTRUCTION INDUSTR*" OR "CONSTRUCTION ACTIVIT*" OR "CONSTRUCTION WORK*")). The acronym "TAK" indicates that the term must occur at least once in the title, abstract, or keywords.

The initial results returned 75 studies from the database without any filtering. However, two filters were required to eliminate publications that were irrelevant or in a language that was incomprehensible to the authors. As a result, the language filters – only English papers (results in 74 studies) – and the publishing type – only journal articles (58 studies) – were used.

Then, using inclusion/exclusion criteria, all 58 titles were examined, and studies that were not directly related to construction project management were excluded (e.g., "*Governmental Investment Impacts on the Construction Sector Considering the Liquidity Trap*" – Alshboul et al, 2022), and the sample was reduced to 43 studies. Finally, all 43 abstracts were evaluated, and 24 articles were chosen for the SLR following the same inclusion/exclusion criteria.

Note that many studies mentioned stochastic variables (e.g., "*Physical Distancing Analytics for Construction Planning Using 4D BIM*" – Hosny et al., 2022) or coefficients (e.g., "*Research on extremely short construction period of engineering project based on labor balance under resource tolerance*" – Peng et al., 2022) but did not use stochastic processes themselves and were, thus, excluded. This occurred because some studies interpret the term "stochastic" as an alternative expression for "risk and uncertainty" (e.g., "Stochastic analysis for managing risk of delay in Duri oil construction projects, Indonesia" – Sandhyavitri, 2022).

Following the SLR processes the 24 journal papers were thoroughly read, and the key studies and their contributions are detailed next.

Yang (2005) conducted one of the pioneering works highlighted in the SLR, proposing a time-cost tradeoff analysis using stochastic formulations, primarily with uncertainties due to funding unpredictability. A probability function was defined to a chance-constrained programming, a subdivision of stochastic programming. The study then translated the stochastic formulation into a deterministic expression that was evaluated using a small building project, after analyzing the possible random variables and the probabilistic distribution that they could adopt. According to the author, the proposed methodology could assist planners in quantifying the impact of the examined uncertainty. The current study has a similar goal, which is to determine the impact of safety uncertainties on project time.

Tseng et al. (2009) provided a technique to estimate contingency considering uncertainties that affect project duration and cost utilizing stochastic programming as well. They accounted for random events caused by incorrect or inadequate data, resulting in change orders, and permitted delays during the construction phase. Also, they created the stochastic process using mathematical expressions that assumed specified probabilistic distributions for the random variables, like Yang (2005). To deal with stochastic processes, they used the real options technique rather than chance-constrained programming. In fact, they claimed that using predetermined distributions could be a restriction of their model, but other distributions could be employed in future works.

It is interesting to note that the applicability of the Markov chain's memoryless property into the project management process was noted by Tseng et al. (2009). As a result of a numerical application, they discovered that increasing project riskiness led in a decrease in construction length, because the contingency rises project cost in order to reduce project duration, as a result of dynamic crashing (risks mitigation based on optimal decisions). This conclusion is intriguing because it suggests that, when adequately defined and combined with cost contingency, uncertainty may assist reduce construction length. Following their findings, the present study used Markov processes to characterize stochastic processes and hypothesized the following: when adequately established and supplemented with proactive safety management (i.e., acts prior to an accident occurring) the uncertainties associated with safety accidents can reduce the duration of a construction project.

In terms of safety, an accident prediction model can be thought of as a stochastic process (MARHAVILAS; KOULOURIOTIS; SPARTALIS, 2013). The authors created a risk assessment tool based on a stochastic harmonic analysis of time series. They discovered that this method was time-consuming and necessitated a large enough accident database spanning a few years. Furthermore, they concluded that Markov chain and neural network models can improve accident forecasting in the medium term, and they pointed out that, due to the non-linearity of random data, exponential smoothing and auto-regression models are inapplicable for times-series forecasting that follows a stationary condition. It is worth noting that, like the other research, the study developed by Marhavilas et al. (2013) presented the stochastic character primarily through the use of stochastic variables.

Li et al. (2016), on the other hand, made a significant advance in this research area by studying the stochastic states that describe the process and developing a mathematical model to comprehend them. More specifically, they developed a live construction planning method that makes use of real-time location technologies and a Markovian stochastic process. The safety states referred to the type of safety occurrence (e.g., near-miss, accident, or fatal accident), and the research focused on the relationship between them rather than the causes. In fact, many research works proposed safety prediction models based on accident causes (DHALMAHAPATRA et al., 2019; XU; ZOU, 2021; ZHANG, Fan et al., 2019). However, due to the randomness of the construction environment, the current study followed a similar approach to that adopted by Li et al. (2016), focusing solely on safety events that could be produced by a variety of factors. This assumption is particularly necessary to avoid using sensitive data that could disrupt the implementation of any method, since the General Data Protection Regulation restricts the use of worker information or related data and is linked to the consideration of legal issues indicated by Zhou et al. (2013).

Zhang et al. (2021) recently added to this area by investigating the evolution of safety risk as a random process and by converting risk to accident as a random process. They discovered that the use of stochastic processes theory to construct safety risk management is still 'very rare' in the literature. As a result, they created a conceptual framework to integrate stochastic processes with safety accidents utilizing Markov, normal, and Poisson processes. The goal of this study was to anticipate the risk-accident process, and they determined that, on a macro level, the Poisson process is an adequate model to predict quantitatively the number of safety accidents, which is consistent with previous research (CHUA; GOH, 2005). Chapter 5 further examined the adoption of Poisson process to represent safety incidents in energy infrastructure construction projects, but they concluded that the pure birth process can also be used and leading to better estimation. As a result, the current study used the quasi birth and death process to integrate the random variables of safety and duration.

As a conclusion from this literature review, the initial questions can be answered, as follows.

- What are the primary contributions linked to stochastic processes in construction management in the literature?

  A.: As observed in the literature, the adoption of stochastic processes is commonly associated with logistics and procurement activities (CARON; MARCHET; PEREGO, 1998; HSU; ANGELOUDIS; AURISICCHIO, 2018; NG; FANG; UGWU, 2008), risks analysis (MARHAVILAS; KOULOURIOTIS; SPARTALIS, 2013; ZHANG, Zhenhao; LI; YANG, 2021); and project time and cost evaluation (ESPINOZA, 2011; FARSHCHIAN; HERAVI, 2018; YANG, 2005).

Most studies proposed mathematical models and focused only on the theorical implications and did not use any kind of data mining process. Three works (LI, Heng et al., 2016; MARHAVILAS; KOULOURIOTIS; SPARTALIS, 2013; ZHANG, Zhenhao; LI; YANG, 2021), however, showed good examples of application of stochastic processes into the construction safety management, which is interesting for the current study.

- How were the theories implemented in the studies?

  A.: The Markov chain theory is commonly adopted to consider the transitions between stochastic states, as observed in Li et al. (2016). Also, the Poisson process is used to forecast the number of accidents during the construction phase, as observed in Zhang et al. (2021), which is a well-known approach in this research field since the work of Chua and Goh (2005). However, this literature did not identify the use of the birth and death process, which can be described from a Markov process perspective and simplified with the Poisson model. This indicates that there is a scientific gap that, if filled, can contribute to the literature with theorical advances.

### 6.2.2. Stochastic processes

A stochastic process can be defined as a set of functions $f(x_i, t)$ that maps each instance $x_i$ of the sample space $\Omega$ and a parameter $t$ into a new domain $\Upsilon$. This function map, also known as a random process, is usually coupled to a time parameter, which might be discrete or continuous. Furthermore, stochastic analysis is typically used to explain the general behavior of an experiment (ALBUQUERQUE, 2017). For example, the observation of safety events during the construction phase of an infrastructure project is an experiment where the safety occurrences are the instances $x_i$ in the domain $\Omega$, and the random process $f(x_i, t)$ (or stochastic process) results in the occurrences over time in a new domain $\Upsilon$, similarly with a time series.

Formally, it can be described as:

$$f(x_i, t) : \Omega \to \Upsilon : x_i \to f(x_i, t) \mid x_i \in \Omega , t \in \Upsilon , f(x_i, t) \in \Upsilon \qquad (21)$$

The Birth and Death Process (BDP) is a type of stochastic process in which the instance is discrete and can rise (birth) or decrease (death) in unit increments along a continuous period, i.e., it exhibits a discrete stochastic process with a continuous parameter. This type of random process is commonly used in biological research to better understand species evolution, such as in the whooping crane population (MANDJES; SOLLIE, 2022). It is also being used in research on communication channels, such as the signal 6G with integrated sensing and communication (ZHANG, Zhengyu et al., 2023). Despite the fact that Chapter 5 demonstrated the potential benefits of using it to understand the evolution of accidents, studies based on BDP theory in construction management are extremely rare. As a result, the current study is a pioneer in this field.

A popular technique to define a BDP is to create diagram states that correspond to all conceivable states of the instance $x_i$ and the transitions between them. It is worth noting that the graphic depicts the entire history of $x_i$ over time $t$ and so provides an alternate approach to comprehend the map function that represents the stochastic process.

A popular technique to defining a BDP is to create diagram states that correspond to all conceivable states of the instance $x_i$ and the transitions between them. It is worth noting that the diagram depicts the entire history of $x_i$ over time t and so it provides an alternative approach to comprehend the map function $n(t)$ that represents the stochastic process. The states diagram is shown in Figure 5, where the circles indicate the states $S_i$ and the arrows reflect the transition rates $\lambda$ (birth) and $\mu$ (death). In general, the transition rates vary with time $t$, i.e., $\lambda(t)$ and $\mu(t)$.

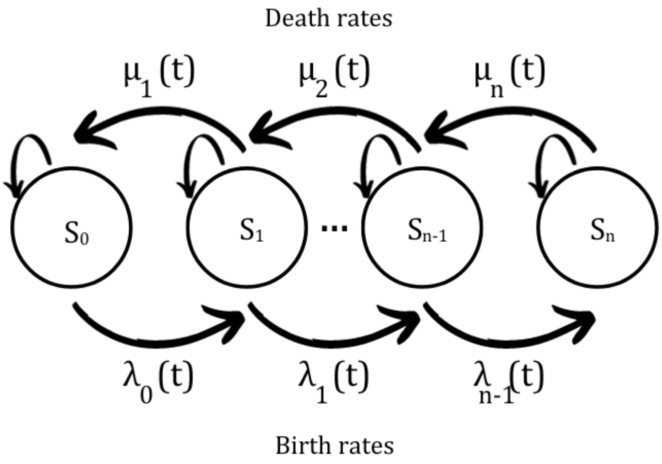



Figure 5 - States diagram that represents a typical birth and death process.

Formally, the possible transitions are as follows:

$$n(t+\delta) = \begin{cases} n(t)+1 \\ n(t)-1 \\ n(t+\delta) = n(t) \end{cases}$$

Where,

- $n(t)$ is the map function that yields the number of occurrences in a particular experiment;
- $\delta$ is a time increment.

Thus, by linking the transition rates, one obtains the transition probabilities $p_{ij}(dt)$ in the infinitesimal time $dt$:

$$p_{ij}(dt) = \begin{cases} \lambda_i(t)dt + o(dt), if\ j = i+1 \\ \mu_i(t)dt + o(dt), if\ j = i-1 \\ 1 - \lambda_i(t)dt - \mu_i(t)dt + o(dt), if\ j = i \\ 0, otherwise \end{cases}$$

Where,

$i$ and $j$ indicate the past and subsequent states, respectively, and belong to the set of integers, i.e., $i, j \in \mathbb{Z}$;

$o(dt)$ is any function that satisfy the condition $\lim_{dt\to 0} \frac{o(dt)}{dt} = 0$.

In addition, when two generic states $a$ and $b$ are observed, the transition probability at time $t$ given an increment $dt$ follows the following equation:

$$p_{ab}(t+dt) = \lambda_{b-1}(t)p_{a,b-1}(t)dt + [1-\lambda_b(t)-\mu_b(t)]p_{ab}(t)dt + \mu_{b+1}(t)p_{a,b+1}(t)dt + o(dt)$$

The first parcel in the right side of the equation represents the transition from state $b-1$ to state $b$, the second parcel represents no transition – past and subsequent states are $b$, and the third parcel represents the transition from state $b$ to state $b+1$.

Furthermore, as demonstrated by Miller and Childers (2012), the forward Kolmogorov equations can be obtained by deriving both sides of the equation by $dt$:

$$\frac{dp_{ab}(t)}{dt} = \lambda_{b-1}(t)p_{a,b-1}(t) - [\lambda_b(t)+\mu_b(t)]p_{ab}(t) + \mu_{b+1}(t)p_{a,b+1}(t)$$

As shown by Karlin and McGregor (1955), it can be organized in a matrix format as:

$$\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{A}$$

Where, $\boldsymbol{P}(t)$ is an $n \times n$ matrix and represents the transition probabilities, in which $n$ represents the maximum possible state in the process:

$$\boldsymbol{P}(t) = \begin{bmatrix} p_{00}(t) & p_{01}(t) & \cdots & 0 \quad 0 & \\ p_{10}(t) & p_{11}(t) & & 0 \quad 0 & \\ \vdots & & \ddots & \vdots & \\ 0 \quad 0 & & \cdots & p_{n-1,n-1}(t) & p_{n-1,n}(t) \\ 0 \quad 0 & & & p_{n,n-1}(0{,}5) & p_{n,n}(t) \end{bmatrix}$$

And **A** represents the generator matrix with $n \times n$ dimension:

$$\mathbf{A} = \begin{bmatrix} -\lambda_0(t) & \lambda_0(t) & \cdots & 0 \quad 0 & \\ \mu_1(t) & -(\lambda_1(t)+\mu_1(t)) & & 0 \quad 0 & \\ & \vdots & \ddots & \vdots & \\ & 0 \quad 0 & \cdots & -(\lambda_{n-1}(t)-\mu_{n-1}(t)) & \lambda_{n-1}(t) \\ & 0 \quad 0 & & \mu_n(t) & -\mu_n(t) \end{bmatrix}$$

The pure birth process is a specific example of the BDP in which the transition $\mu_i(t)$ is zero for any state $i$ and time $t$, resulting in the following generator matrix:

$$\mathbf{A} = \begin{bmatrix} -\lambda_0(t) & \lambda_0(t) & \cdots & 0 & 0 \\ 0 & -\lambda_1(t) & & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ 0 \quad 0 & & \cdots & -\lambda_{n-1}(t) & \lambda_{n-1}(t) \\ 0 \quad 0 & & & 0 & 0 \end{bmatrix}$$

This work adopts the Quasi Birth and Death Process (QBDP), which combines both BDP and pure birth process. QBDP, as defined by Wallace (1969), is a continuous-time Markov process in which the transition probability matrix is composed of transition probability submatrices, both with tridiagonal structure. Indeed, according to Fadiloglu and Yeralan (2002), QBDP is "a generalization of the birth-death process," and following van Leeuwaarden and Winands (2004), homogeneous QBDP can be understood as a two-dimensional Markov chain. As a result, the possible states are represented by the Kronecker product between the BDP and the pure birth states, resulting in a vector of two scalars.

## 6.3. Proposed methodology

### 6.3.1 Theorical formulation

In this study, the QBDP approach is based on states that are vectors $(d, s)$ related to delay $(d)$ and safety $(s)$ states, as opposed to the BDP approach, which states are scalars. Also, to use a similar notation used in many studies (FADILOGLU; YERALAN, 2002; MANDJES; SOLLIE, 2022; OSOGAMI, 2005), the safety states are referred to as levels, and the delay states are referred to as phases, as shown in Figure 22.

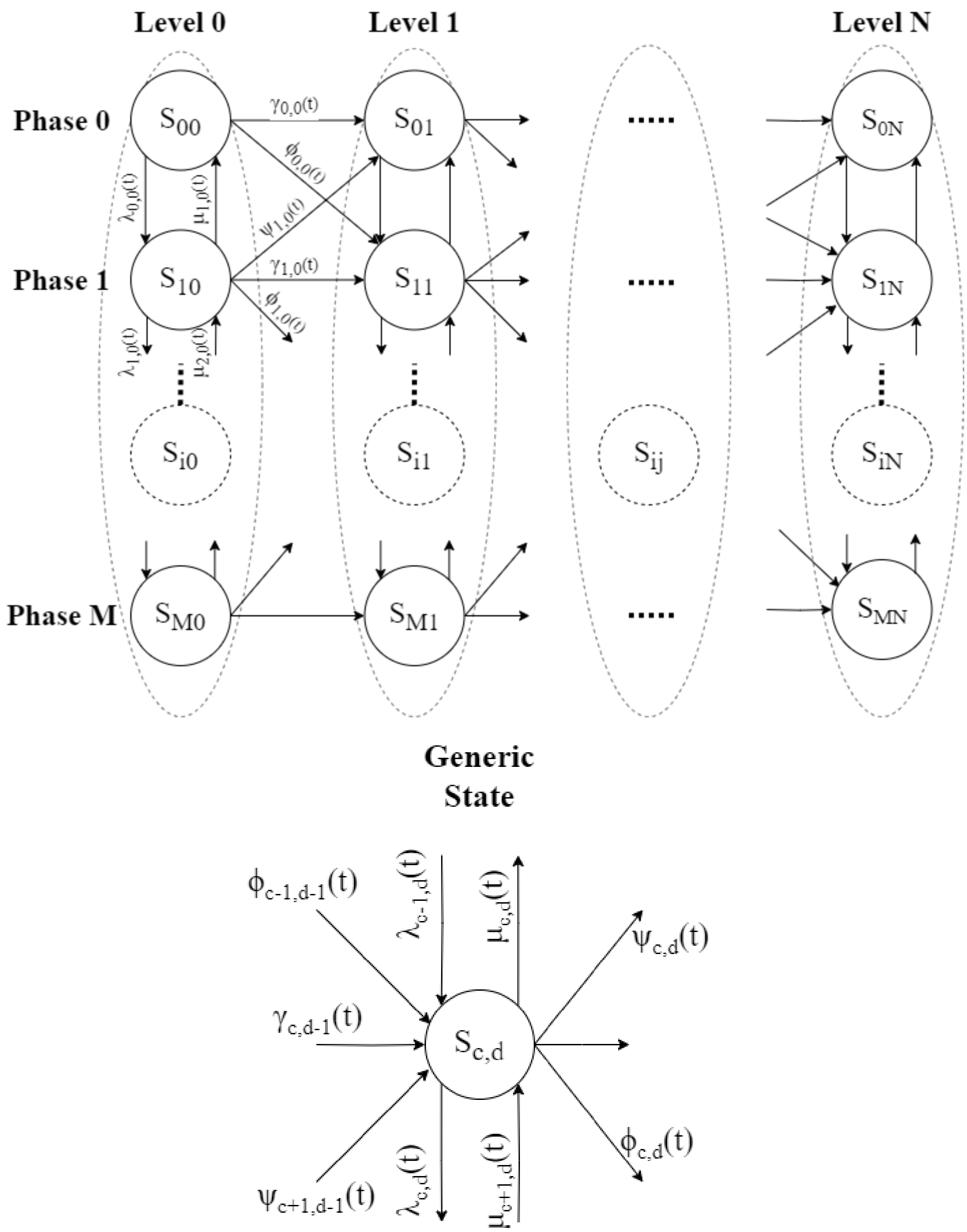



Figure 22 - States diagram that represents a quasi birth and death process considering safety states as levels.

Formally, the function map that defines the stochastic process is $f(\,d_i,\, s_i,\, t)$ and the possible states, $\boldsymbol{s}$(t), are:

$$\Omega = \boldsymbol{s}(t) = \{(d, s) \mid d = 0,1,2,3, \ldots; s = 0,1,2,3, \ldots\} \qquad (22)$$

It should be noted that the adopted QBDP assumes only states from the set of non-negative integer numbers, i.e., $\{d, s \in \mathbb{Z}^{+}\}$, as well the birth and death process.

Therefore, at time $t$ and given a time increment δ, the following transitions are possible:

$$\boldsymbol{s}(t+\delta)=\begin{cases}\boldsymbol{s}(t)+(0,0)\\ \boldsymbol{s}(t)+(0,1)\\ \boldsymbol{s}(t)+(1,0)\\ \boldsymbol{s}(t)+(1,1)\\ \boldsymbol{s}(t)+(-1,0)\\ \boldsymbol{s}(t)+(-1,1)\end{cases}$$

Similarly, as previously described, the transition probabilities $\boldsymbol{p}_{(i,j),(m,n)}(dt)$, which denotes a transition from a state $(i,j)$ to a generic state $(m,n)$ in the infinitesimal time $dt$, are given by:

$$\boldsymbol{p}_{(i,j),(m,n)}(dt)=$$
$$\begin{cases}\lambda_{ij}(t)dt+o(dt),\ if\ (m,n)=(i,j)+(1,0)\\ \mu_{ij}(t)dt+o(dt),\ if\ (m,n)=(i,j)+(-1,0)\\ \gamma_{ij}(t)dt+o(dt),\ if\ (m,n)=(i,j)+(0,1)\\ \phi_{ij}(t)dt+o(dt),\ if\ (m,n)=(i,j)+(1,1)\\ \psi_{ij}(t)dt+o(dt),\ if\ (m,n)=(i,j)+(-1,1)\\ \left[1-\left(\lambda_{ij}(t)+\mu_{ij}(t)+\gamma_{ij}(t)+\phi_{ij}(t)+\psi_{ij}(t)\right)\right]dt+o(dt),\ if\ (m,n)=(i,j)\\ 0, otherwise\end{cases} \tag{23}$$

And the transition probabilities at time $t$ given an increment $dt$ are:

$$\begin{aligned}\boldsymbol{p}_{(i,j),(m,n)}(t+dt)&\\ &=\gamma_{m,n-1}(t)\boldsymbol{p}_{(i,j),(m,n-1)}(t)dt\\ &+\lambda_{m-1,n}(t)\boldsymbol{p}_{(i,j),(m-1,n)}(t)dt\\ &+\mu_{m+1,n}(t)\boldsymbol{p}_{(i,j),(m+1,n)}(t)dt\\ &+\phi_{m-1,n-1}(t)\boldsymbol{p}_{(i,j),(m-1,n-1)}(t)dt\\ &+\psi_{m+1,n-1}(t)\boldsymbol{p}_{(i,j),(m+1,n-1)}(t)dt\\ &+\left[1-\lambda_{m,n}(t)-\mu_{m,n}(t)-\gamma_{m,n}(t)-\phi_{m,n}(t)\right.\\ &\left.-\psi_{m,n}(t)\right]\boldsymbol{p}_{(i,j),(m,n)}(t)dt+o(dt)\end{aligned} \tag{24}$$

As a result, the Kolmogorov forward equations can be obtained:

$$\frac{d\boldsymbol{p}_{(i,j),(m,n)}(t)}{dt} = \gamma_{m,n-1}(t)\boldsymbol{p}_{(i,j),(m,n-1)}(t) + \lambda_{m-1,n}(t)\boldsymbol{p}_{(i,j),(m-1,n)}(t) \\ + \mu_{m+1,n}(t)\boldsymbol{p}_{(i,j),(m+1,n)}(t) \\ + \phi_{m-1,n-1}(t)\boldsymbol{p}_{(i,j),(m-1,n-1)}(t) \\ + \psi_{m+1,n-1}(t)\boldsymbol{p}_{(i,j),(m+1,n-1)}(t) \\ - \left[\lambda_{m,n}(t) + \mu_{m,n}(t) + \gamma_{m,n}(t) + \phi_{m,n}(t) \right. \\ \left. + \psi_{m,n}(t)\right]\boldsymbol{p}_{(i,j),(m,n)}(t) \tag{25}$$

Rewriting the equations in matrix format, one obtains:

$$\frac{d\boldsymbol{p}_{(i,j),(m,n)}(t)}{dt} \\ = \begin{bmatrix} \boldsymbol{p}_{(i,j),(m-1,n-1)}(t) & \boldsymbol{p}_{(i,j),(m-1,n)}(t) & \boldsymbol{p}_{(i,j),(m-1,n+1)}(t) \\ \boldsymbol{p}_{(i,j),(m,n-1)}(t) & \boldsymbol{p}_{(i,j),(m,n)}(t) & \boldsymbol{p}_{(i,j),(m,n+1)}(t) \\ \boldsymbol{p}_{(i,j),(m+1,n-1)}(t) & \boldsymbol{p}_{(i,j),(m+1,n)}(t) & \boldsymbol{p}_{(i,j),(m+1,n+1)}(t) \end{bmatrix} \\ \times \begin{bmatrix} \phi_{m-1,n-1}(t) & \gamma_{m,n-1}(t) & \psi_{m+1,n-1}(t) \\ \lambda_{m-1,n}(t) & -[\lambda_{m,n}(t) + \mu_{m,n}(t) + \gamma_{m,n}(t) + \phi_{m,n}(t) + \psi_{m,n}(t)] & \mu_{m+1,n}(t) \\ 0 & 0 & 0 \end{bmatrix} \tag{26}$$

Due to the number of variables, some notations are simplified by dropping subscripts and birth and death rates are treated as constants, as follows: $\boldsymbol{p}_{(i,j),(m,n)}(t) = \boldsymbol{p}_{(m,n)}$ ; $\lambda_{m,n}(t) = \lambda$ ; $\mu_{m,n}(t) = \mu$ ; $\gamma_{m,n}(t) = \gamma$ ; $\phi_{m,n}(t) = \phi$ ; $\psi_{m,n}(t) = \psi$, for any $m, n \in \Omega$.

The steady-state equations are calculated to obtain the generator matrix that describes the QBDP process, as described in Fadiloglu and Yeralan (2002). These equations define the balance conditions and represent the border states in the states diagram. For example, it is known that there are only two possible transitions to the first state $(0,0)$, coming from the state $(1,0)$ or from the same state $(0,0)$. As a result, the sum of both possible transitions must be zero.

$$-(\lambda+\gamma+\phi)\,\boldsymbol{p}_{(0,0)}+\mu\,\boldsymbol{p}_{(1,0)}=0 \tag{27}$$

$$-(\lambda+\mu+\gamma+\phi+\psi)\,\boldsymbol{p}_{(i,0)}+\lambda\,\boldsymbol{p}_{(i-1,0)}+\mu\,\boldsymbol{p}_{(i+1,0)}\mid M>i>0 \tag{28}$$

$$-(\gamma+\mu+\psi)\,\boldsymbol{p}_{(M,0)}+\lambda\,\boldsymbol{p}_{(M-1,0)}=0 \tag{29}$$

$$-(\lambda+\gamma+\phi)\,\boldsymbol{p}_{(0,j)}+\gamma\,\boldsymbol{p}_{(0,j-1)}+\mu\,\boldsymbol{p}_{(1,j)}+\psi\,\boldsymbol{p}_{(1,j-1)}=0\mid N>j>0 \tag{30}$$

$$-\lambda\,\boldsymbol{p}_{(0,N)}+\gamma\,\boldsymbol{p}_{(0,N-1)}+\mu\,\boldsymbol{p}_{(1,N)}+\psi\,\boldsymbol{p}_{(1,N-1)}=0 \tag{31}$$

$$\begin{aligned}-(\lambda+\mu+\gamma+\phi+\psi)\,\boldsymbol{p}_{(i,j)}+\gamma\,\boldsymbol{p}_{(i,j-1)}+\phi\,\boldsymbol{p}_{(i-1,j-1)}+\psi\,\boldsymbol{p}_{(i+1,j-1)}\\+\mu\,\boldsymbol{p}_{(i+1,j)}+\lambda\,\boldsymbol{p}_{(i-1,j)}=0\;\mid M>i>0,N>j>0,\end{aligned} \tag{32}$$

$$\begin{aligned}-(\mu+\lambda)\,\boldsymbol{p}_{(i,N)}+\gamma\,\boldsymbol{p}_{(i,N-1)}+\phi\,\boldsymbol{p}_{(i-1,N-1)}+\psi\,\boldsymbol{p}_{(i+1,N-1)}+\lambda\,\boldsymbol{p}_{(i-1,N)}\\+\mu\,\boldsymbol{p}_{(i+1,N)}=0\mid N>j>0\end{aligned} \tag{33}$$

$$\begin{aligned}-(\gamma+\mu+\phi)\,\boldsymbol{p}_{(M,j)}+\gamma\,\boldsymbol{p}_{(M,j-1)}+\phi\,\boldsymbol{p}_{(M-1,j-1)}+\lambda\,\boldsymbol{p}_{(M-1,j)}=0\mid M\\>i>0\end{aligned} \tag{34}$$

$$-\mu\,\boldsymbol{p}_{(M,N)}+\gamma\,\boldsymbol{p}_{(M,N-1)}+\phi\,\boldsymbol{p}_{(M-1,N-1)}+\lambda\,\boldsymbol{p}_{(M-1,N)}=0 \tag{35}$$

According to Osogami (2005), the generator matrix $\boldsymbol{Q}$ has the following format, a block tridiagonal matrix of submatrices:

$$\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{L}_{00} & \boldsymbol{F}_{01} & 0 & \\ \boldsymbol{B}_{10} & \boldsymbol{L}_{11} & \boldsymbol{F}_{12} & \cdots \\ 0 & \boldsymbol{B}_{21} & \boldsymbol{L}_{22} & \\ & \vdots & & \ddots\end{bmatrix} \tag{36}$$

Where,

- $\boldsymbol{L}_{ij}$ refers to the transition rate submatrix related to from state $(i,j)$ to state $(k,j)$ for $i\neq k$;
- $\boldsymbol{F}_{ij}$ refers to the transition rate submatrix related to from state $(i,j)$ to state $(k,j+1)$ for any value of $i$ and $k$;
- $\boldsymbol{B}_{ij}$ refers to the transition rate submatrix related to from state $(i,j)$ to state $(k,j-1)$ for any value of $i$ and $k$.

Note that the submatrices attend to the following rule, which is important to obtain the submatrices:

$$\boldsymbol{L}_{00}\boldsymbol{1}^T + \boldsymbol{F}_{01}\boldsymbol{1}^T = \boldsymbol{B}_{10}\boldsymbol{1}^T + \boldsymbol{L}_{11}\boldsymbol{1}^T + \boldsymbol{F}_{12}\boldsymbol{1}^T = 0 \tag{37}$$

Thus, the submatrices can be extracted, following the procedures explained in Fadiloglu and Yeralan (2002):

$$L_{00} = \begin{bmatrix} -(\gamma+\lambda+\phi) & \lambda & \cdots & 0 \quad 0 & \\ \mu & -(\gamma+\lambda+\mu+\phi+\psi) & & 0 \quad 0 & \\ & \vdots & \ddots & \vdots & \\ & 0 \quad 0 & \cdots & -(\gamma+\lambda+\mu+\phi+\psi) & \lambda \\ & 0 \quad 0 & & 0 & -(\gamma+\lambda+\phi) \end{bmatrix}_{N\times N} \tag{38}$$

$$F_{01} = \begin{bmatrix} \gamma & \phi & \cdots & 0 & 0 \\ \psi & \gamma & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \gamma & \phi \\ 0 & 0 & & \psi & \gamma \end{bmatrix}_{N\times N} \quad F_{12} = \begin{bmatrix} \gamma & \phi & \cdots & 0 & 0 \\ \psi & \gamma & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \gamma & \phi \\ 0 & 0 & & \psi & \gamma \end{bmatrix}_{N\times N} \tag{39}$$

$$B_{10} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & & 0 & 0 \end{bmatrix}_{N\times N} \quad B_{21} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & & 0 & 0 \end{bmatrix}_{N\times N} \tag{40}$$

$$L_{11} = \begin{bmatrix} -(\gamma+\lambda+\phi) & \lambda & \cdots & 0 \quad 0 & \\ \mu & -(\gamma+\lambda+\mu+\phi+\psi) & & 0 \quad 0 & \\ & \vdots & \ddots & \vdots & \\ & 0 \quad 0 & \cdots & -(\gamma+\lambda+\mu+\phi+\psi) & \lambda \\ & 0 \quad 0 & & 0 & -(\gamma+\lambda+\phi) \end{bmatrix}_{N\times N} \tag{41}$$

$$L_{NN} = \begin{bmatrix} -(\lambda) & \lambda & \cdots & 0 \quad 0 & \\ \mu & -(\lambda+\mu) & & 0 \quad 0 & \\ & \vdots & \ddots & \vdots & \\ & 0 \quad 0 & \cdots & -(\gamma+\lambda) & \lambda \\ & 0 \quad 0 & & \mu & -\mu \end{bmatrix}_{N\times N} \tag{42}$$

It is also important to understand that, due to the simplification of constant transition rates, all submatrices from level 1 will repeat, except for $\boldsymbol{L}_{NN}$, i.e.:

$$\boldsymbol{F}_{01} = \boldsymbol{F}_{12} = \boldsymbol{F}_{23} = \cdots = \boldsymbol{F}$$
$$\boldsymbol{B}_{10} = \boldsymbol{B}_{21} = \boldsymbol{B}_{32} = \cdots = \boldsymbol{B}$$
$$\boldsymbol{L}_{00} = \boldsymbol{L}_{11} = \boldsymbol{L}_{22} = \boldsymbol{L}_{33} = \cdots = \boldsymbol{L}$$

With the generator matrix and using the states where the probabilities are known, Equation 37 applied in the case of QBDP can be solved through the demonstration given by Kharoufeh (2011), utilizing the fundamental matrices $\boldsymbol{R}$, $\boldsymbol{G}$, and $\boldsymbol{U}$. The analytical solution to such a problem is exhaustive, necessitating the use of numerical algorithms and computational approaches, such as the Erlangization approach (MANDJES; SOLLIE, 2022). The results are the stationary probabilities vectors, which are the current study's goal to understand the probability QBDP states during the construction phase and then predict the influence of safety states into delay ones.

The fundamental matrices must satisfy the following equations by definition:

$$\begin{gathered} \boldsymbol{B}+\boldsymbol{L}\,\boldsymbol{G}+\boldsymbol{F}\boldsymbol{G}^2=\mathbf{0} \\ \boldsymbol{F}+\boldsymbol{R}\,\boldsymbol{L}+\boldsymbol{R}^2\boldsymbol{B}=\mathbf{0} \\ \boldsymbol{L}+\boldsymbol{F}(-\boldsymbol{U})\,\boldsymbol{I}\,\boldsymbol{B}-\boldsymbol{U}=\mathbf{0} \end{gathered} \tag{43}$$

It should be noted that in case that the aim of the analysis is the effects of delay states on safety events, the states diagram is different and, consequently, the formulations too. Figure 23 shows the states diagram for this variation and the mathematical expressions can be seen in Appendix A.

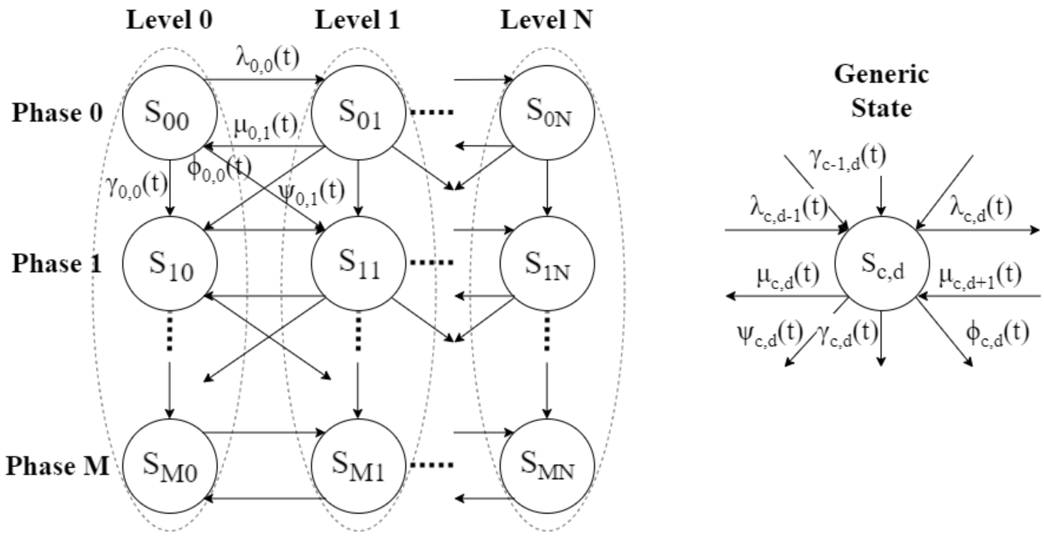



Figure 23 - States diagram that represents a quasi birth and death process considering delay states as levels.

### 6.3.2. Computational techniques

To solve the transition probabilities in QBDP, the Python module BuTools (HORVATH; TELEK, 2017) with the package Matrix Analytic Methods, that provides Markovian solution algorithms, was employed in this work. In short, given the generator matrix, this tool computes the fundamental matrices and the stationary probabilities for the QBDP using matrix analytical solution methods. See the documentation[4] for more details.

The dataset employed in this work is made up of 23 energy infrastructure construction projects, the majority of which uses wind and photovoltaic technologies. As mentioned in Chapter 5, it should be noted that the number of projects covered in this study is greater than those used in previous studies with equivalent goals (CHUA; GOH, 2005). The projects include distinct numbers of safety and delay records, as well as locations, which can be better understood in the exploratory study discussed later in this section. In addition, data on safety and delays are provided as weekly reports.

As a result, the QBDP solution represents the global random process that incorporates all projects and may be used in a new project to capture the unpredictability connected with the consequences of safety accidents on delay events. However, in an ongoing project, the QBDP may be so general that such predictions may yield unreal results ignoring the project`s particularities. Therefore, the current methodology proposes the use of neural networks models for each time series created in each project during the construction phase. Since there are two variables – safety and delay events – that are not independent and identically distributed (this condition was verified by data processing techniques), the adopted algorithms were more oriented to multivariate time series forecasting.

This hybrid approach is presented in Figure 24. It should be noted that traditional safety management, such as fall prevention barriers and the distribution of individual and collective protective equipment, will remain important and necessary in this workflow. Also highlighted in yellow is the construction monitoring and reporting process, which is critical for producing adequate and consistent data for both stochastic and AI models.

[4] Retrieved November 14, 2023, from https://webspn.hit.bme.hu/~telek/tools/butools/doc/index.html

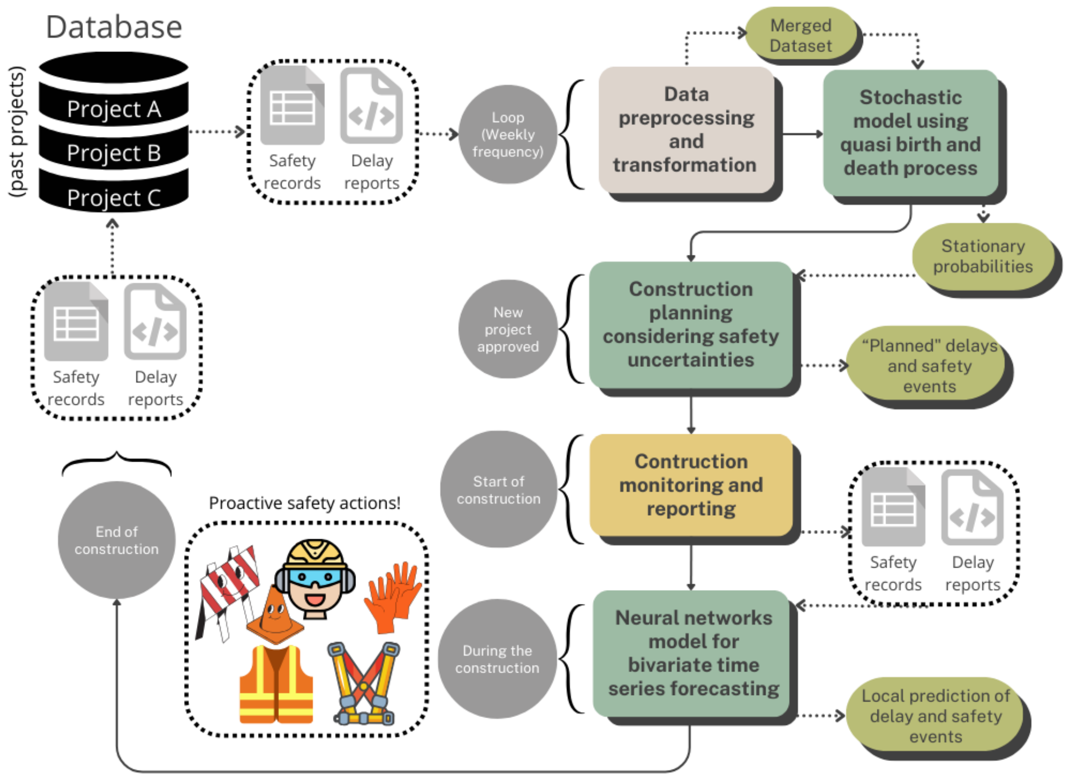



Figure 24 - Proposed hybrid solution including QBDP and neural networks model to deal with safety and delay events before and during the construction phase.

Following the methodology described in Chapter 5, a structured procedure was developed to preprocess, transform, process, and interpret the outcomes of the computational methods adopted.

The major difference between the safety and delay raw data refers to the file format and how the user interacts with it during the data gathering process. While safety reports are completed manually by the user in spreadsheets, delay reports are generated automatically in XML (Extensible Markup Language) format by a construction management platform that receives only the actual start and end dates of the construction activities from the user.

To produce the delay states, certain data preprocessing and transformation techniques were used based on the planned duration (the difference between the scheduled start and finish dates) and the actual duration of the activity. As a result, in accordance with the Knowledge Discovery in Databases (KDD) outlined in Chapter 5, the data preprocessing and transformation stages were performed individually for each raw data collection that represents the random variables. The safety and delay treated datasets were then combined into a single *dataframe* to begin the training and testing set division (data organization step) for data mining purposes, as shown in Table 12.

Table 12 - Merged dataset with delay and safety events.

| # | PROJECT | TECHNOLOGY | REGION | COUNTRY | NORMALIZED DATA | SAFETY STATES | DELAY STATES | SAFETY TRANSITIONS | DELAY TRANSITIONS |
|---|---|---|---|---|---|---|---|---|---|
| 1 | K400 | WIND | AFRICA | South Africa | 0.008584 | 0 | 85 | 0 | 0 |
| 2 | K400 | WIND | AFRICA | South Africa | 0.009009 | 0 | 84 | 0 | -1 |
| … | … | … | … | … | … | … | … | … | … |

Considering the total number of projects, an initial training and testing set division was carried out, defining 80% as training projects and 20% as testing projects. Then, in the training dataset (80% or 16 projects) the walk forward validation technique was adopted, because the random variables are time dependent. As shown in Figure 25, the following time periods were applied: 30%, 50%, and 70% of the dataset have been added since $t_0$.

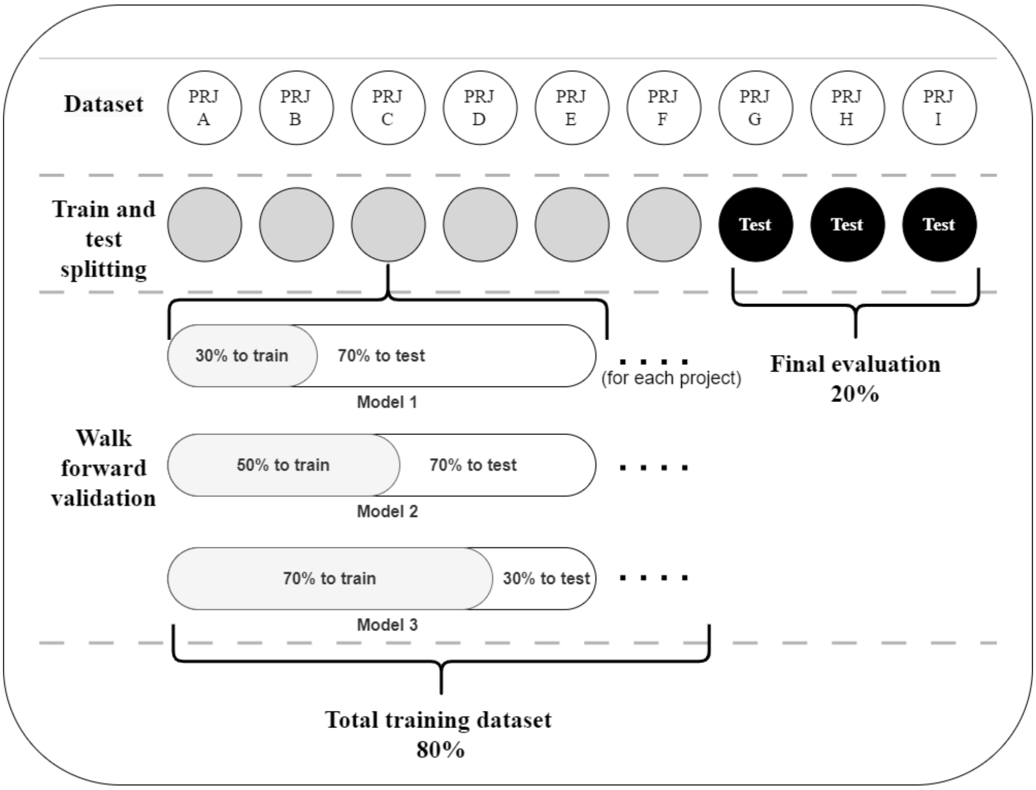



Figure 25 - Dataset division and organization including the walk forward validation technique.

It should be emphasized that the time series representing a single ongoing project, with 30% of the dataset, represents a small quantity of data to train, but this is the reality in a construction project. Each infrastructure construction project has a distinct environment, with diverse work teams, site conditions, and other aspects that add to the construction industry's uniqueness.

This small dataset in time series forecasting was common in the early months of the COVID-19 pandemics, as documented by Fong et al. (2020). In fact, the authors used only 14 instances to predict the next 6 instances. They discovered that classical algorithms, like the Auto Regression Integrated Moving Average (ARIMA), did not perform well in forecasting the time series. Also, they concluded that the polynomial neural network, that according to Fong et al. (2020) is a “prototype” of Convolutional Neural Network (CNN), performed the best time series forecasting with the lowest Root Mean Square Error (RMSE). However, Cruz-Nájera et al. (2022) found that the ARIMA technique performed better in forecasting than the applied artificial neural network technique in a study involving short time series related to crimes in cities. It should also be noted that, as Ospina et al. (2023) pointed out, the ARIMA method fails in long-term prediction. As a result, it appears that the literature is unclear about which technique should be used in short-sized time series. For that reason, the current study tested both the statistical and neural networks approaches.

Moreover, in the current study, the time series is bivariate with the safety and delay states and so ARIMA and other classical statistical techniques may be not applicable since the training data is a vector. In contrast, there are well-known statistical methods to deal with vectorial data, the current study adopted the Vector Auto Regression (VAR) and the Vector Autoregressive Integrating Moving Average (VARIMA).

The VARIMA technique was used by Karim and Ahmed (2023) to forecast and evaluate the pandemics consequences on oil prices, concluding that is applicable for short-term forecasting only, which is the case of a construction project with week reports.

A distinguishing feature of statistical methods is that, unlike neural network models, they require that the condition that variables are stationary, with no trend or seasonal variations. Each time series' stationary was assessed using the Augmented Dickey Fuller (ADF) test. When the test statistic exceeds the critical value, the null hypothesis cannot be rejected, indicating that the series is not stationary. If the stationary tests were not successful, the data was logarithmically transformed and differentiated in a single time step. In other words, these two operations were required to convert the time series into one that was stationary enough for a statistical analysis.

The Granger's and Johansen's causality test was also applied to determine the relationship between the random variables (safety and delay). The tests were required to determine whether the safety states influence the delay states and, thus, whether the time series is truly bivariate. This is significant because if variables have no relationship, they can be excluded and modeled separately. In contrast, if a relationship exists, the variables must be taken into account during the modeling phase. When the Granger's test is less than the critical value, the null hypothesis must be rejected, indicating that there are potential causal relationships between variables. For the Johansen's test, the variables have dependency relations when the p-value is greater than the critical value (0.05).

In addition to statistical methods, AI techniques were used to compare forecasting models. According to the literature, the techniques most commonly used are CNN and Long Short-Term Memory (LSTM), both of which are capable of capturing nonlinearities and, thus, do not require stationary conditions. It should be noted that, for the purposes of this work, the use of neural networks is recommended, as they deal with noise and chaotic components better than other methods, according to Stepchenko et al. (2017). Indeed, Stepchenko et al. (2017) concluded that it could outperform time series predictions using a Markov chain, which is interesting to investigate in the current study. There are also recent studies combining CNN and LSTM in multivariate time series forecasting in smart city topics (PAPASTEFANOPOULOS et al., 2023).

### 6.3.3. Data mining assumptions

The adopted QBDP assumes that transitions are made in unit increments; thus, to be consistent with this methodology, a time lag ($p$) of one was considered to fit the VAR model, i.e., $p = 1$. The variables ($K$) used in the VAR modeling are safety and delay states, i.e., $K = 2$. It should be noted that the VAR parameters did not vary due to the study's context explained previously. To fit the VARIMA model, however, two additional parameters ($p$ and $q$) are required, which define the model's order. According to Karim and Ahmed (2023), the parameters are critical for producing consistent and accurate VARIMA models. As a result, the methodology used tested four variations: $p = 1,2$ and $q = 1,2$.

In relation to the CNN and LSTM techniques, the model fitting required a different format of variable in partial time-series (block sequences). The blocks are made up of steps in (expected input) and steps out (expected output), which act as new attributes that supplement the two existing ones and improve the neural network model's performance. For example, in a project with nine instances – $[1,1,2,3,4,4,4,5,6]$; six input blocks with three steps in and six output blocks with one step out can be created:

- Input block sequences: $\left[[1,1,2], [1,2,3], [2,3,4], [3,4,4], [4,4,4], [4,4,5]\right]$
- Output block sequences: $\left[[3], [4], [4], [4], [5], [6]\right]$

Furthermore, two groups of neural network models were created: one with only the training project dataset (local) and one with all of the training project datasets (global). This is critical for understanding the proposed methodology's applicability in construction companies that do not have a database of past projects.

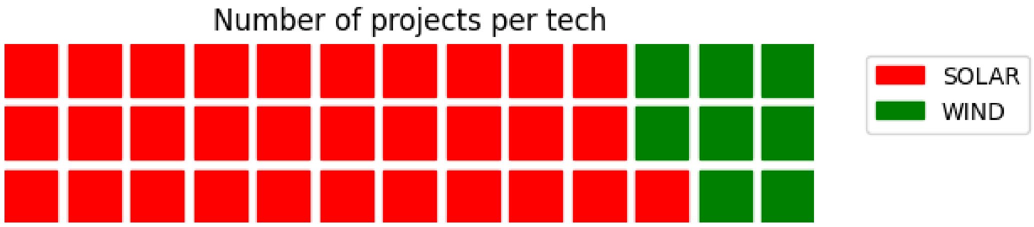
In the proposed methodology, the number of steps in were tested with 4 and 7 instances, and the steps out were fixed in 1. In addition, for both ML techniques, the activation layer was configured with a rectified linear unit (ReLU) function, the Adam algorithm was used as the optimizer, the Mean Squared Error (MSE) was used as the loss function, the number of epochs was tested with 100, and the batch size was tested with 32. The CNN and LSTM architectures are shown in Table 3.

The Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and RMSE metrics were used to evaluate the VAR, VARIMA, CNN, and LSTM models during the walk-forward validation. Because the dataset contains many actual values close to 0, MAPE score may fail, resulting in an error (divide by zero) or extremely high values. As a result, the MAPE score is only used to understand the MAE score, and the RMSE will be the final score used to test and validate the models, following Fong et al. (2020).

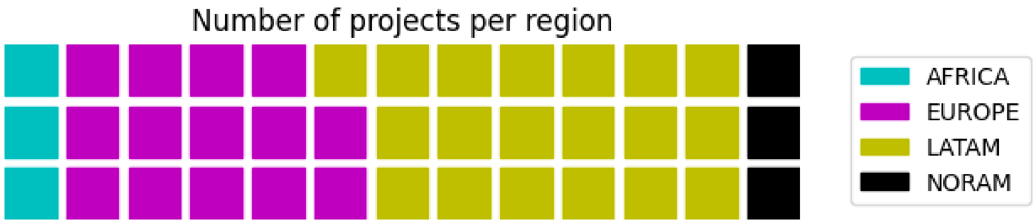
## 6.4.
## Results and discussion

### 6.4.1.
### Exploratory data analysis

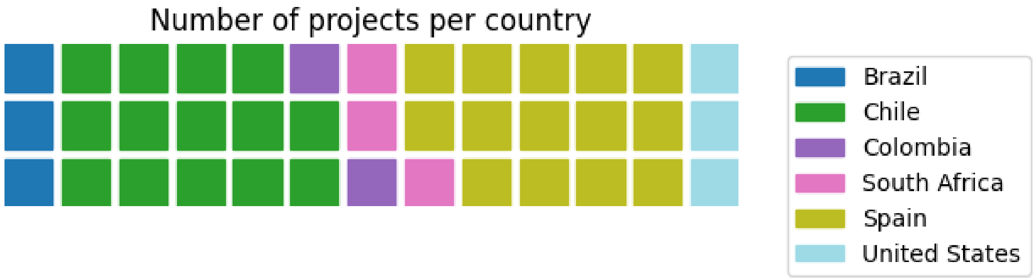
Figure 26 show exploratory data analysis using the same approach as presented in Chapter 5. The total number of projects that contain consistent data related to safety and delay events during construction is 23, with 79 percent being photovoltaic (solar) power plants and 21 percent being wind farms. Furthermore, 49 percent of them are in Latin American (LATAM) countries, the majority of which are in Chile (36 percent of the total number of projects).




(a)




(b)




(c)

Figure 26 - Exploratory data analysis to understand the technologies (a), regions (b), and countries (c) associated with the safety and delay dataset.

Figure 27 depicts histograms of safety and delay states organized by region, country, and technology in relation to the observed safety and delay events in these projects. The $y$ axis represents state occurrences, while the $x$ axis represents state number. For example, the first histogram shows that for European projects, more or less three safety events (state equal to 3) occurred nearly 120 times.

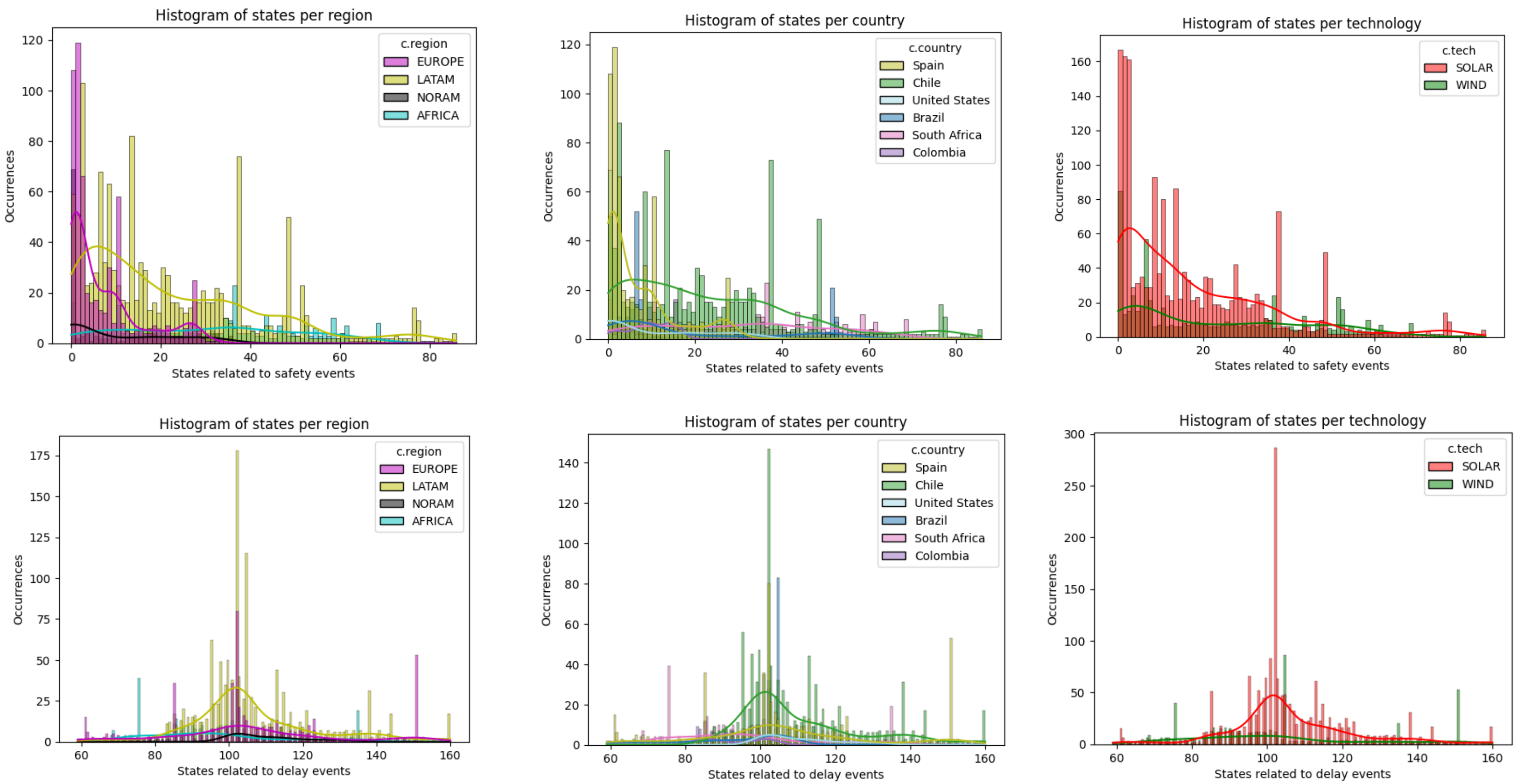
In addition, to normalize the occurrences, a density estimation is used, which is represented by the Kernel Density Estimation (KDE) curves. By examining these curves, it is possible to conclude that the LATAM region has more occurrences of safety states because the peak of the curve is located at a higher value of the axis $x$.

Concerning the delay states, to eliminate the possibility of negative delay states, like a construction advance brought about by higher worker productivity, all projects began with 102 days of construction as the default. As a result, the histogram's center is on 102, which represents zero delay events. This number was altered to the original in the final analysis of each project and was defined using the minimum and maximum observed delay states to avoid non positive states. In contrast to the safety histogram, the KDE curve for delay states shows a higher frequency of no delays in LATAM projects than in other regions.

Furthermore, the project technologies present similar KDE curves for safety states with different number of occurrences, since most projects are photovoltaics. However, they presented different delay state curves – the wind projects appear to have a higher frequency of delay events than the solar ones.

The histograms of states shown in Figure 27 also show that there are many peaks of safety and delay events, making the use of classical probability distributions to define their behavior difficult. As demonstrated in Chapter 5, by using classical probability distributions to fit the safety occurrences in such projects, random events may be missed, necessitating the use of methods that capture high levels of uncertainty.

Figure 27, however, did not depict the evolution of the project states. The time series for each project is thus plotted in, where the $y$ axis represents the observed states, and the $x$ axis represents the normalized construction time – 0 represents the start of construction and 1 represents the end. The red curves represent the evolution of safety states and the blue ones, the evolution of delay states.




(a) (b) (c)

Figure 27 - Exploratory data analysis to understand the occurrence of safety and delay states in the available dataset, according to the region (a), countries (b), and technologies (b).

The actual time-series for all the projects did not present any kind of relation or pattern between the safety and delay random variables, as shown in Figure 28. It is a stochastic process, as observed, because there is no clear understanding of the behaviors of the safety and delay states in the projects, and even less so of the influence of these two random variables. The bivariate time-series depicted in Figure 28 was used in the current study to understand the global behavior of safety occurrences into delays events and to forecast during a construction project.

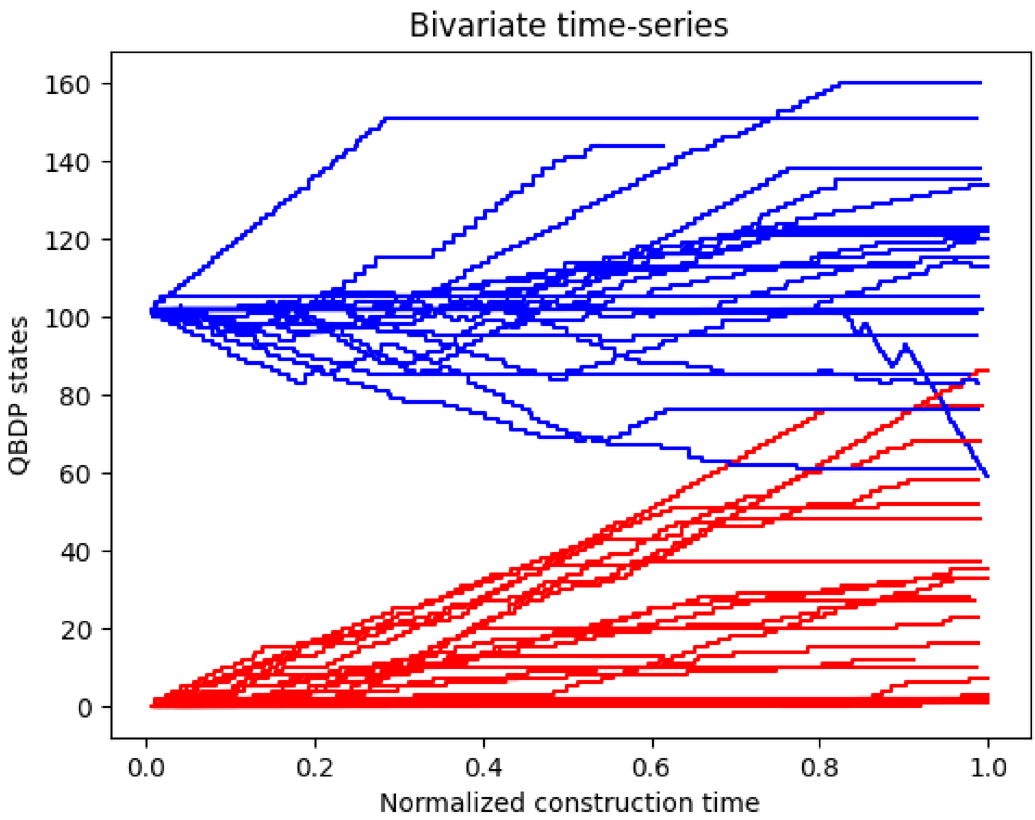



Figure 28 - Bivariate time series that represents the available dataset for forecasting.

### 6.4.2. Statistical tests

Following the EDA, the stationary tests were computed for each variable in the time series. The ADF test yielded six projects with nonstationary data or errors during the calculations. In Table 13, each project is represented by its time series and the ADF p-value. The two projects that did not return any p-values due to a calculation error were removed from the dataset. In these projects' time series, it is possible to observe a long period of time without the occurrence of safety or delay events – this is unrealistic when compared to the other projects. The authors understand that this level of noise is unacceptable. However, the four projects that failed in the ADF test were not removed but transformed to contain stationary variables.

The causality tests were also carried out to understand the adopted variables. Granger's causality test, in fact, provides a preliminary indication of the effects of safety events on project delays, which must be confirmed or denied following the data processing stage. As shown in Table 14, most projects reject the null hypothesis in both variable relationships, with the exception of two projects where there appears to be no causality relationship from safety events to delay events.

There are some causality relationships between variables in all projects, depending on the used tested. Following the Johansen's test, one project (S145) demonstrated through p-values that the variables are not dependent, with the exception of the safety variable into the delay variable. Figure 29 depicts the time series for this specific project. In fact, the occurrences of states in this project remain nearly constant during the construction phase, which is consistent with the causality test results. It was removed from the dataset as another example of an unreal project dataset.

Table 13 – Statistical test results to check stationary variables using ADF test.

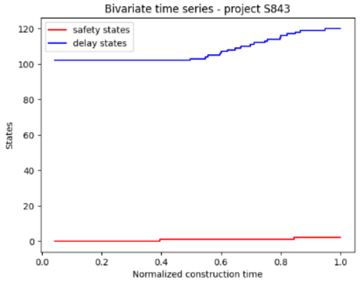
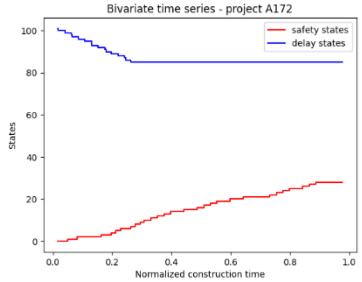
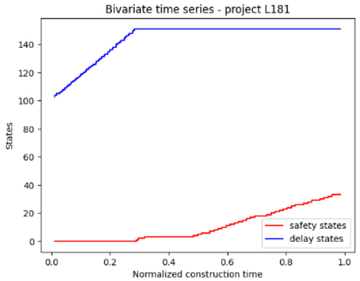
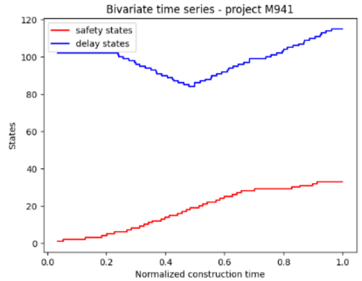
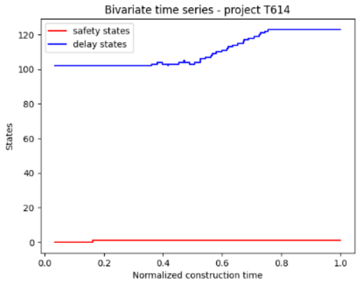
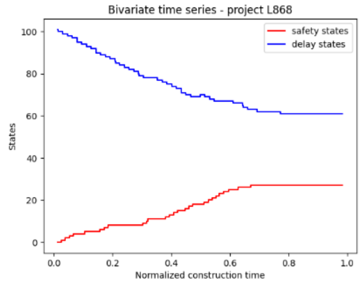
| **Failed projects** | **Bivariate time series** | **p-value** | | **Failed projects** | **Bivariate time series** | **p-value** | |
|---|---|---|---|---|---|---|---|
| | | safety variable | delay variable | | | safety variable | delay variable |
| S843 |  | 5.70e-6 (OK) | 0.85 (Failed) | A172 |  | 0.86 (Failed) | 0.82 (Failed) |
| L181 |  | Error | Error | M941 |  | 0.15 (Failed) | 0.56 (Failed) |
| T614 |  | Error | Error | L868 |  | 2.95e-5 (OK) | 0.58 (Failed) |

Table 14 - Statistical test result to check causality relationship using Granger`s test and Johansen`s test.

| Projects | p-value for the Granger's causality test | | p-value for the Johansen's causality test | |
|---|---|---|---|---|
| | Safety > Delay | Delay > Safety | Safety > Delay | Delay > Safety |
| T238 | 0.1615 | 0.0000 | 0.2766 | 0.0668 |
| S843 | 0.0000 | 0.0000 | 0.4171 | 0.0695 |
| C434 | 0.0411 | 0.0042 | 0.3652 | 0.0004 |
| V104 | 0.0000 | 0.0009 | 0.2037 | 0.0031 |
| S833 | Error | Error | 0.1757 | 0.0111 |
| A172 | 0.0000 | 0.0008 | 0.6161 | 0.2352 |
| M941 | 0.0238 | 0.0000 | 0.4659 | 0.3067 |
| L406 | 0.0006 | 0.0000 | 0.4565 | 0.2366 |
| C907 | 0.0579 | 0.0000 | 0.2524 | 0.0325 |
| M174 | Error | Error | 0.1384 | 0.0084 |
| S242 | 0.0118 | 0.4297 | 0.3618 | 0.0063 |
| S816 | 0.0182 | 0.0054 | 0.1839 | 0.1281 |
| L868 | 0.0001 | 0.0077 | 0.4215 | 0.0104 |
| A739 | 0.0003 | 0.0000 | 0.5718 | 0.1857 |
| C624 | 0.0144 | 0.0575 | 0.2648 | 0.0627 |
| A676 | Error | Error | 0.1174 | 0.0270 |
| S145 | 0.1063 | 0.3448 | 0.1626 | 0.0196 |
| G413 | 0.0002 | 0.0400 | 0.2615 | 0.0189 |
| S165 | 0.0048 | 0.0000 | 0.3378 | 0.0474 |
| K400 | 0.0467 | 0.0000 | 0.4035 | 0.1192 |
| D316 | 0.0000 | 0.0000 | 0.2562 | 0.1239 |

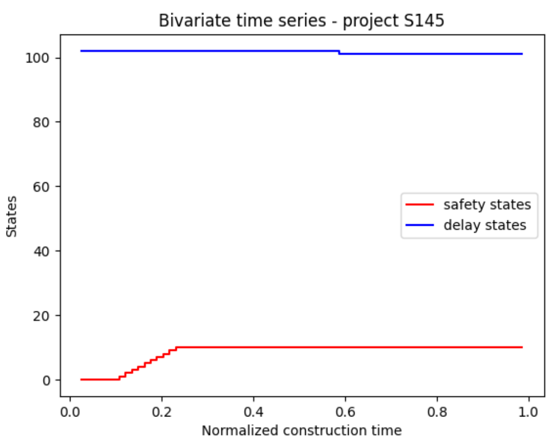



Figure 29 – Example of time series related to the failed project in the causality tests.

### 6.4.3. QBDP-based solution

To understand the global behavior between the forementioned variables, first the QDBP was modeled including all the dataset (2150 instances). Note that the size of the dataset is appropriate with the stochastic process, analogously with the study developed by Mandjes and Sollie (2022) and De Gunst et al. (2022), in which the simulated data achieved a total number of 2000 samples.

The transition rates, defined in Equation 23, were calculated and the results are presented in Table 15. Note that the double jump – simultaneous transitions of safety and delay states – are very rare, as expected; the safety and delay transitions are proportional, i.e., the frequencies are very close; and the none transition represents 36.6%, which means that only in almost one third of the construction period simulated, nothing occurred (safety events or delay events), showing how dynamic and critical is an energy infrastructure construction project.

Table 15 - Actual QBDP transition rates.

| Transition variable | Occurrence | Frequency |
|---|---|---|
| $\boldsymbol{\gamma}$ | 601 | 0.280 |
| $\boldsymbol{\lambda}$ | 440 | 0.205 |
| $\boldsymbol{\mu}$ | 295 | 0.137 |
| $\boldsymbol{\phi}$ | 22 | 0.010 |
| $\boldsymbol{\psi}$ | 6 | 0.003 |

Using the transitions rates, the fundamental matrices were calculated from the generator matrix, resulting in the following matrices $\boldsymbol{R}, \boldsymbol{G},$ and $\boldsymbol{U}$ with a residual error in the order of 1 $e^{-16}$.To facilitate results interpretation, the generator matrix was truncated to the dimension of $20 \times 20$, which means that the effects of the safety event will be examined until state 19.

$$\boldsymbol{R} = \begin{bmatrix} \mathbf{0.631} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0.755} \end{bmatrix}_{20\times20}$$

$$\boldsymbol{G} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}_{20\times20}$$

$$\boldsymbol{U} = \begin{bmatrix} \mathbf{-0.494} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{-0.420} \end{bmatrix}_{20\times20}$$

It is important to note that the fundamental matrices were checked to verify if they satisfy the Equation 43 and all the conditions were satisfied with the worse precision in the order of $1\,e^{-15}$.

Then, to obtain the stationary distribution in level 0, the method *QBDSolve* was applied and, consequently, the stationary distribution in any level was achieved by using the function *QBDStationaryDistr*. The results including all the stationary distributions are presented in Figure 30, in which the $x$ axis represents the phases (delay states) and the $y$ axis, the stationary probability given the level (safety state) which is colored and listed in the legend.

As observed, the results indicate that the higher the safety state (levels), the more likely are the higher number of safety states (phases). It is possible to conclude, thus, that if the high number of construction states coincide likely with high number of delay occurrences.

In addition, some quantitative conclusions can be drawn from the QBDP results. It is worth noting that the most likely delay state in this simulation (19) is 6. It should be noted that a most likely delay state of 1 appears only when the safety state is equal to 4. This is critical because if a construction schedule is defined as "planning" some delay days, it should be assumed that some safety events will occur. As a result, the "zero accidents" mindset must be implemented during the preconstruction phase rather than just during the construction phase, as is customary.

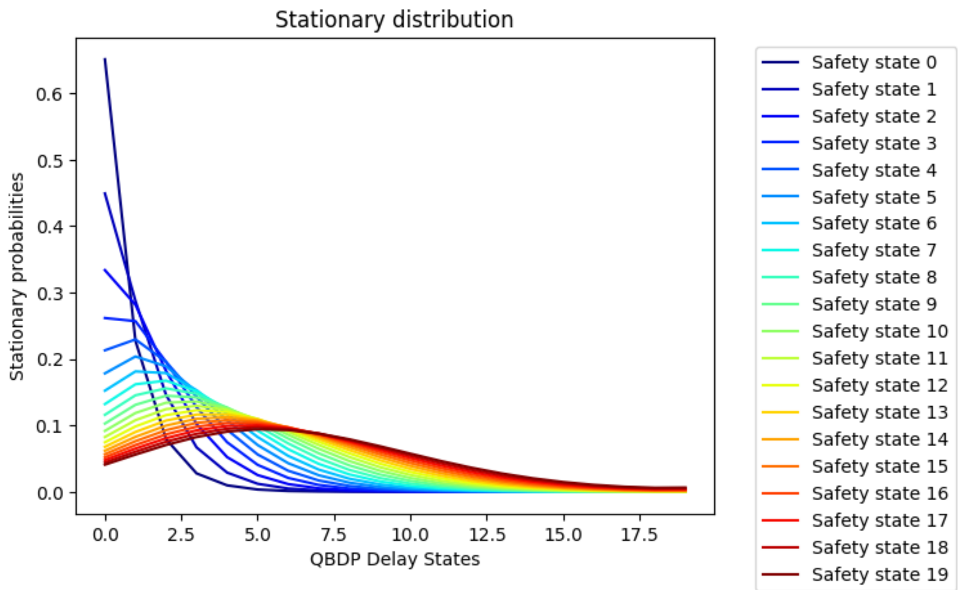



Figure 30 - Stationary probability distribution for the proposed QBDP using the actual dataset truncated to 20 safety states.

### 6.4.4. Bivariate time series forecasting

It should be noted that the QBDP solution does not support detailed forecasting in a time series format. Therefore, related to the bivariate time series forecasting applied individually in the projects, the statistical and neural networks models were compared.

Table 16 shows the experiment plan including the model parameters that were changed to evaluate the model's performance and the performance metrics. Note that the lowest RMSE scores in the overall performance is obtained using the VAR model (0.60) and the LSTM model (0.65) with 70% of training data. The VAR model used 32 instances to result in this score and the LSTM model used 60 instances. Therefore, in an initial analysis, the VAR model seems to be more adequate than the VARIMA one, and the LSTM local model resulted in better scores compared to the CNN and global models.

Table 17 shows the comparison for each percentage of training project used during the walk forward validation. The neural network models that used all of the training projects outperformed the statistical methods using the smallest train dataset for walk forward validation (30%). The best RMSE score was associated with the global LSTM model, while the worst RMSE score was associated with the VAR model – the score value was nearly double that of the best. In comparison to the largest training dataset (70 %), the results show almost the opposite. The RMSE score for the VAR model was the best, and the local LSTM model came in second. The scores were, however, very close between the models, and the sum of RMSE scores was better for the local LSTM model.

The final validation using 20% of the projects produced similar results. Locally trained models outperformed global models, with better metrics for the LSTM models using 30% of the available local dataset. Statistical and local neural network models performed similarly when using the largest dataset (70 percent). For ongoing construction projects that are nearing completion, the global CNN and LSTM models did not perform well.

Appendix B contains the entire experiment plan.

Table 16 – Experiment plan including the best models during the walk forward validation.

| | Dataset | | | Parameters | | | Resulted metrics per model | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Training Project | Walk forward validation | Trainig data size | VARIMA | | Neural networks | VAR | | | VARIMA | | | Local CNN | | | Local LSTM | | | Global CNN | | | Global LSTM | | |
| # | | % Train* | (instances, attributes) | p | q | Block sequences (steps in) | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE |
| **3** | T238 | 0,3 | (13, 2) | 2 | 1 | - | 5,73 | 0,20 | 6,44 | 5,53 | 0,05 | 6,09 | - | - | - | - | - | - | - | - | - | - | - | - |
| **12** | T238 | 0,7 | (32, 2) | 2 | 2 | - | 0,36 | 0,00 | 0,60 | 0,68 | 0,11 | 0,88 | - | - | - | - | - | - | - | - | - | - | - | - |
| **121** | L868 | 0,3 | (30, 2) | 1 | 1 | - | 4,97 | 0,17 | 5,85 | 15,96 | 0,43 | 18,02 | - | - | - | - | - | - | - | - | - | - | - | - |
| **122** | L868 | 0,3 | (30, 2) | 1 | 2 | - | 4,97 | 0,17 | 5,85 | 15,74 | 0,42 | 17,85 | - | - | - | - | - | - | - | - | - | - | - | - |
| **123** | L868 | 0,3 | (30, 2) | 2 | 1 | - | 4,97 | 0,17 | 5,85 | 17,86 | 0,48 | 19,72 | - | - | - | - | - | - | - | - | - | - | - | - |
| **124** | L868 | 0,3 | (30, 2) | 2 | 2 | - | 4,97 | 0,17 | 5,85 | 16,23 | 0,44 | 18,22 | - | - | - | - | - | - | - | - | - | - | - | - |
| **176** | S165 | 0,5 | (27, 2) | 2 | 2 | - | 1,76 | 0,09 | 2,19 | 1,89 | 0,17 | 2,42 | - | - | - | - | - | - | - | - | - | - | - | - |
| **232** | L406 | 0,5 | (39, 2) | - | - | 7 | - | - | - | - | - | - | 2,58 | 0,05 | 2,89 | 2,58 | 0,05 | 2,58 | - | - | - | - | - | - |
| **246** | M174 | 0,7 | (60, 2) | - | - | 7 | - | - | - | - | - | - | 0,65 | 0,01 | 0,77 | 0,65 | 0,01 | 0,65 | - | - | - | - | - | - |
| **272** | A676 | 0,3 | (22, 2) | - | - | 7 | - | - | - | - | - | - | 4,26 | Error | 4,38 | 4,26 | Error | 4,26 | - | - | - | - | - | - |
| **361** | C624 | 0,3 | (32, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 2,73 | 0,29 | 2,94 | 2,73 | 0,29 | 2,73 |
| **363** | C624 | 0,5 | (53, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 3,52 | 0,36 | 3,65 | 3,52 | 0,36 | 3,52 |
| **378** | S165 | 0,7 | (39, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 2,50 | 0,22 | 2,68 | 2,50 | 0,22 | 2,50 |

* From the initial time;

0,716 This color represents the best metric scores.

Table 17 - Summary of the best models and its metrics per train dataset.

| **Dataset** | **Analysis** | **Statistical techniques** | | | | | | **AI techniques** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Walk forward validation** | **Math Operation** | **VAR** | | | **VARIMA** | | | **Local CNN** | | | **Local LSTM** | | | **Global CNN** | | | **Global LSTM** | | |
| | | **Metrics** | | | **Metrics** | | | **Metrics** | | | **Metrics** | | | **Metrics** | | | **Metrics** | | |
| **% Train*** | | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE | MAE | MAPE | RMSE |
| **0,3** | Average | 16,97 | 35,9% | 21,57 | 14,74 | 34,6% | 16,81 | 15,52 | 62,0% | 16,30 | 15,52 | 62,0% | 15,52 | 11,34 | 86,8% | 12,54 | 11,34 | 86,8% | 11,34 |
| | Sum | 882,40 | | 1121,49 | 751,66 | | 857,51 | 496,62 | | 521,74 | 496,62 | | 496,62 | 362,96 | | 401,22 | 362,96 | | 362,96 |
| | Minimum | 4,97 | 9,2% | 5,85 | 5,53 | 4,7% | 6,09 | 4,26 | 14,6% | 4,38 | 4,26 | 14,6% | 4,26 | 2,73 | 15,6% | 2,94 | 2,73 | 15,6% | 2,73 |
| **0,5** | Average | 8,88 | 16,1% | 11,07 | 11,90 | 25,2% | 13,53 | 11,86 | 39,4% | 12,25 | 11,86 | 39,4% | 11,86 | 13,03 | 90,1% | 13,67 | 13,03 | 90,1% | 13,03 |
| | Sum | 461,51 | | 575,65 | 618,72 | | 703,80 | 379,46 | | 391,99 | 379,46 | | 379,46 | 386,66 | | 406,73 | 386,66 | | 386,66 |
| | Minimum | 1,76 | 2,2% | 2,19 | 1,89 | 10,8% | 2,42 | 2,58 | 4,7% | 2,89 | 2,58 | 4,7% | 2,58 | 3,52 | 15,6% | 3,65 | 3,52 | 15,6% | 3,52 |
| **0,7** | Average | 3,43 | 6,6% | 4,35 | 8,44 | 15,9% | 9,39 | 6,79 | 29,6% | 6,94 | 6,79 | 29,6% | 6,79 | 16,76 | 76,5% | 17,06 | 16,76 | 76,5% | 16,76 |
| | Sum | 178,32 | | 226,35 | 438,64 | | 488,27 | 217,13 | | 222,19 | 217,13 | | 217,13 | 501,87 | | 511,61 | 501,87 | | 501,87 |
| | Minimum | 0,36 | 0,3% | 0,60 | 0,68 | 2,8% | 0,88 | 0,65 | 0,6% | 0,77 | 0,65 | 0,6% | 0,65 | 2,50 | 14,6% | 2,68 | 2,50 | 14,6% | 2,50 |

0,716 This color represents the best metric scores.

This color represents the second best metric scores.

Figure 31 depicts the score results based on the percentage used in the walk forward validation and the models used. The $y$ axis shows the average RMSE score, the $x$ axis shows the percentage of train dataset used, and the colors show the trained and tested models. Except for the global models, all of the models improved their RMSE scores as the dataset increased. One possible explanation for this exception condition in the global model is that the global dataset is too broad and does not adequately represent local forecasting.

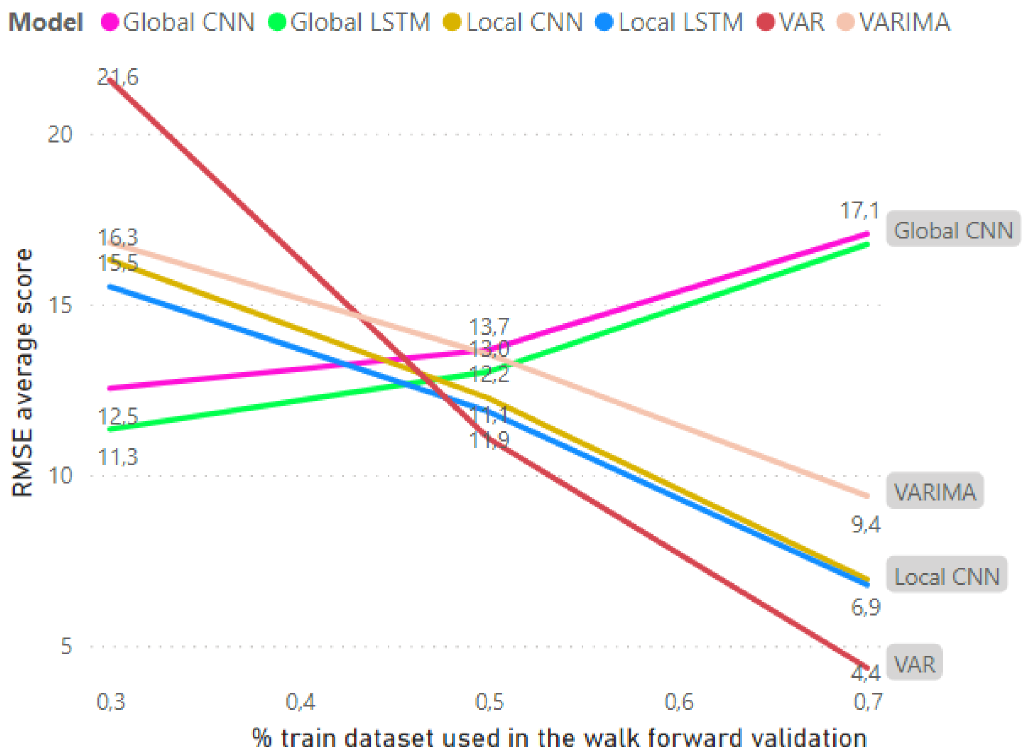



Figure 31 - Models` performance with the increasing train dataset.

The results shown in Figure 31 show that the CNN and LSTM produced very similar outputs that capture more random events than the VAR model when the training dataset is small, which is consistent with the findings of Fong et al. (2020) that verify that neural networks performed better for time series forecasting using small dataset. However, as the dataset size grows, statistical forecasting becomes more suitable for bivariate time series. Nevertheless, the neural network models continue to perform well with larger datasets, with acceptable RMSE scores.

It's also worth noting that the models have comparable RMSE metrics when the training and testing data are both 50%. It could imply that block sequences perform better when the amount of training and testing data is comparable. As a result, future research can look into walk forward validation using always block sequences with the same number of train and test samples.

As a result of analyzing the overall performance, the current study concluded that the LSTM model is the most appropriate for bivariate time series forecasting, using all available projects when the construction stage is in its early stages and only the local dataset produced during the construction phase when the construction stage is nearing its end.

It is important to note that using the neural networks model, the level of uncertainties captured is higher than the stochastic model, but the user does not obtain the global behavior of the entire set of projects to predict a new project starting without any size. Indeed, the proposed CNN and LSTM models need some input data from the project, which it is only possible to obtain after a few weeks from the start of the construction phase.

Therefore, the obtained results suggest that a hybrid solution may be necessary to deal with safety and delay uncertainties in a construction planning method. The stochastic solution, during the pre-construction phase, with all the history of previous projects, is adequate to understand the possible scenarios of safety and delays events in a new project, when there are not available data. And, the AI solution, using neural networks applied to bivariate time series forecasting, is applicable for the new data that is generated during the construction phase.

## 6.5. Conclusions

Due to importance of the safety issue in the construction industry and the lack of construction planning methods that take it into account, the current study proposed a new planning method using the quasi birth and death process and machine learning techniques. To that end, a literature review was conducted to understand how stochastic process theories are being applied in construction management. The SLR results showed that the studies are scarce but promising. Due to the high level of uncertainties inherent in a construction project, the adoption of stochastic models could be helpful in capturing totally unexpected events, as pointed out in Carmo and Sotelino (2023).

A Markovian transition diagram was, thus, proposed to describe the safety and delay events in a construction project and using computational techniques, the stationary distributions were obtained for any level and phase. The results indicate that the higher the delay state, a higher safety state is expected, suggesting that delays precede safety. In other words, it is unlikely to have many delay states without a large number of accidents. The reason for that can be further investigated in future studies by understanding how other random variables can be combined in the multivariate time-series, such as cost and productivity.

The QBDP solution can be helpful in providing an overview behavior for the company that captures better the relations and uncertainties associated with safety occurrences than the traditional methods. The proposed approach is appropriate at the global management level, since it helps to understand what are the possible number of safety states based on the “expected” delay or vice-versa when considering a new project. For instance, imposing a zero accidents policy, the likely number of delay events can be calculated to support the decision-making process combined with other analysis, such as financial risks.

It should be noted that the proposed QBDP is adequate for the preconstruction phase, and that statistical and AI techniques are suggested during the construction phase, which represents data in a bivariate time series format. Furthermore, when the construction phase began (small dataset available for time series forecasting), the LSTM model performed better than the statistical methods, as expected by the literature. As the volume of data increased, both statistical and AI techniques performed admirably.

Thus, the neural networks can play an important role during the construction monitoring, by updating the forecasting every moment that a new safety and delay events are reported. Through this real-time planning method, it is possible to capture extraordinary random events that cannot not be captured by the QBDP solution.

It is worth noting that the proposed hybrid solution was tested and validated using real-world energy infrastructure construction projects. Therefore, as suggestion for future works, it is recommended the extension of such approach for residential and commercial buildings, and other types of infrastructure projects (e.g., bridges and airports). Note that maybe for construction projects that do not have a high level of uncertainties, the proposed solution may be unnecessary, and the classical statistical methods may suffice.

# 7
# The proposed construction planning method in a real application

Case study paper to be submitted by Cristiano S. T. do Carmo and Elisa D. Sotelino to a peer-review international journal.

In this chapter, the proposed construction planning method is applied to a real scenario to illustrate its usability. The used data in this demonstration is real and corresponds to energy infrastructure construction projects from 2020 and so forth. The datasets contain distinct uncertainties and random events similar to those that occurred during COVID-19 pandemic, Ukraine war. It is also associated to each project´s local context, which could affect project duration and generate safety accidents during the construction phase. However, due to the Global Data Protection Regulation, the company and project names have been omitted.

## 7.1.
## Initial assumptions and user`s background

Once a new project is approved to be executed, the planning team must develop a construction schedule. Traditionally, the user`s experience with previous projects is used to understand the construction sequence and possible risks that may affect the project`s key performance indicators, which is usually associated with time and cost, as shown in Carmo and Sotelino (2023). In the proposed methodology, the manner in which the construction sequence is carried out does not change, but the way in which the risks are taking into account do.

Independently of the professional expertise, the planned construction duration is usually not equal to the actual construction duration and, thus, delay events must be planned in the traditional planning method. A list of possible events that can generate extra cost and project delays is, thus, generated as show in Table 19.

For each item in this list, a probability of occurrence is calculated based on previous experiences. This calculation uses the classical statistical methods showed in Chapters 4, 5, and 6. The result for one of the items is illustrated in Table 19.

Table 18 - Usual uncertainties and problems considered in the traditional construction planning method for energy infrastructure construction project.

| List of possible events that can impact the project duration and cost | | |
|---|---|---|
| Problems | | Causes |
| 1 | Incorrect time/cost estimate | Basic design or project layout not completed or not performed |
| 2 | Failure during erection/commissioning activities | Inadequate contractor/sub-contractor management |
| 3 | Execution performance rates different than expected | Lack of skilled workers/qualified professionals |
| 4 | Wrong time/cost estimate | High number of tenders to manage in a tight schedule |
| 5 | Fabrication/Delivery rates of equipment different than expected | Site logistic/accessibility external constraints |
| 6 | Change/Additional scope of work | Tightness of the time schedule for project development |
| 7 | Execution performance rates different than expected | Inadequate contractor/sub-contractor management |
| 8 | Change/Additional scope of work | High/unclear fragmentation of contracts |
| 9 | Failure during the final acceptance tests | Inadequate contractor/sub-contractor management |
| 10 | Unexpected archeological/hazardous waste findings | Environmental permitting phase not finalized |

Table 19 - Details about the uncertainties and consequences considered using the statistical approach, based on probabilistic distribution.

<table>
<tr><td>Activity:</td><td colspan="4">Civil works related to grid connection</td></tr>
<tr><td>Problem event:</td><td colspan="4">Wrong time/cost estimate due to basic design or project layout not completed or not performed</td></tr>
<tr><td>Problem type:</td><td colspan="4">Regular</td></tr>
<tr><td rowspan="2">Probability density function:</td><td>Distribution:</td><td>Triangular</td><td>Occurrence probability:</td><td>0.7</td></tr>
<tr><td>Parameters</td><td>Minimum: 6 days</td><td>Maximum: 12 days</td><td>Mode: 11 days</td></tr>
</table>

Note that this represents the default events for any new project. However, as found by the literature review carried out in Carmo and Sotelino (2023), safety events are not considered. In other words, the traditional construction planning method is based in the assumption that safety occurrences do not produce significant project extra costs or delays. However, the hypothesis of this study´s proposed method is that the possibility of safety events must be considered as a new item in the list, as follows:

Table 20 - Additional uncertainties and problems considered in the proposed construction planning method for energy infrastructure construction project.

| | Problems | Causes |
|---|---|---|
| 11 | Project duration delays | Safety occurrences observed during the construction phase |

Moreover, the possible impacts in terms of days of delay can be estimated with the proposed QBDP approach and not using the statistical approaches which resulted in optimistic scenarios and, thus, do not capture well random events.

## 7.2. Required database

A possible barrier to adopt the proposed method is the required database. Even if it is a simple record of safety occurrences with binary inputs by the user, most construction firms do not have it or, when available, they are disorganized, and many pre-processing techniques are required. The KDD process presented in Chapter 5 illustrates the required efforts to treat the dataset.

In this sense, with a real example,

Table 21 explains the first step in the proposed methodology (see Figure 24), which deals with the database organization for safety occurrences and construction duration data. It is worth pointing out that the applied algorithms are applicable for any new safety or duration report increment in the initial database, but its format must be identical to the one adopted in this study. Therefore, in a different company, it would be necessary to modify the algorithms to fit in the company`s report patterns.

It is also important to note that the adopted data preprocessing and transformation techniques optimized the database size. Approximately, 35.000 files (532 megabytes) were reduced to one file with 53 kilobytes.

Following the traditional construction planning method, the “planned” delays are independently calculated from the possible safety occurrences. This calculation is based on alternative scenarios, probabilistic distributions, and Monte Carlo simulations. These techniques appeared in the literature review carried out by Carmo and Sotelino (2023) and is also the one used by the company in the present study.

Table 21 – Available database in the start of the proposed workflow and the dataset preprocessed and transformed.

| Raw data from previous 23 projects (weekly safety and delay reports) | File format |
|---|---|
| (XML excerpt below) | .XML |


```xml
<ID>EC</ID>
<NodeId>&lt;e a0="EC" /&gt;</NodeId>
<JobPlanningType>FixedWork</JobPlanningType>
<Bid>0</Bid>
<Cost>0</Cost>
<Base>0</Base>
<EndTime>2022-11-17T17:00:00</EndTime>
<StartTime>2020-04-24T08:00:00</StartTime>
<Duration>312.16:00:00</Duration>
<Work>129.16:00:00</Work>
<ResourceUnits>119.78463</ResourceUnits>
<BaselineEnd>2021-03-24T17:00:00</BaselineEnd>
<BaselineStart>2020-02-15T08:00:00</BaselineStart>
<BaselineDuration>134.16:00:00</BaselineDuration>
<BaselineWork>653.08:00:00</BaselineWork>
<BaselineResourceUnits>0</BaselineResourceUnits>
<ActivityConstraintType>StartOn</ActivityConstraintTy
<ConstraintDate>2020-04-24T00:00:00</ConstraintDate>
<WorkComplete>0.968355</WorkComplete>
<EstimateForecastDatePerc>0.99804841295822</EstimateF
<DiffPerc>-0.02969341295822</DiffPerc>
<Weight>100</Weight>
<Milestone>false</Milestone>
<Name>Eng. &amp; Cos.</Name>
<Info />
```


.XLS

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | Subsupplier | Man Hours | Fatal Accindents | Accidents | First Aid | Near Miss | Safety Obsevation | Average People |
| 2 | [redacted] | 46 | 0 | 0 | 0 | 0 | 0 | 2.898 |
| 3 | | 6 | 0 | 0 | 0 | 0 | 0 | 192 |
| 4 | | 5 | 0 | 0 | 0 | 0 | 0 | 270 |
| 5 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | | 1 | 0 | 0 | 0 | 0 | 0 | 45 |
| 7 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 9 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | | 26 | 0 | 0 | 0 | 0 | 0 | 1.170 |
| 13 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 14 | | 12 | 0 | 0 | 0 | 0 | 0 | 495 |
| 15 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Preprocessed and transformed data** — .CSV

| | c.project | c.tech | c.region | c.country | c.norm_date | c.safety_states | c.delay_states_positive |
|---|---|---|---|---|---|---|---|
| 1 | K400 | WIND | AFRICA | South Africa | 0.008584 | 0 | 85 |
| 2 | K400 | WIND | AFRICA | South Africa | 0.009009 | 0 | 84 |
| 3 | K400 | WIND | AFRICA | South Africa | 0.018018 | 0 | 83 |
| 4 | K400 | WIND | AFRICA | South Africa | 0.018598 | 0 | 83 |
| 5 | K400 | WIND | AFRICA | South Africa | 0.027027 | 0 | 82 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 1451 | S833 | SOLAR | LATAM | Chile | 0.544444 | 2 | 77 |
| 1452 | S833 | SOLAR | LATAM | Chile | 0.551789 | 2 | 77 |
| 1453 | S833 | SOLAR | LATAM | Chile | 0.564972 | 2 | 77 |
| 1454 | S833 | SOLAR | LATAM | Chile | 0.578154 | 2 | 77 |
| 1455 | S833 | SOLAR | LATAM | Chile | 0.588889 | 2 | 78 |

1417 rows × 9 columns

## 7.3. Simulations and results

The results of the traditional construction planning method applied to a real project is presented in Figure 32. The histogram of possible days of delay according to Monte Carlo simulations is represented in the left upper corner of Figure 32, and it is similar to the histograms presented in Chapter 6. The cumulative density curve is shown below the histogram, and possible project delay scenarios are shown on the right side. The chosen scenario, as shown in the highlighted box, corresponds to an 8-day delay, which occurred in 60% of the simulations with the Full Time Project (FTP) equaling 779 days.

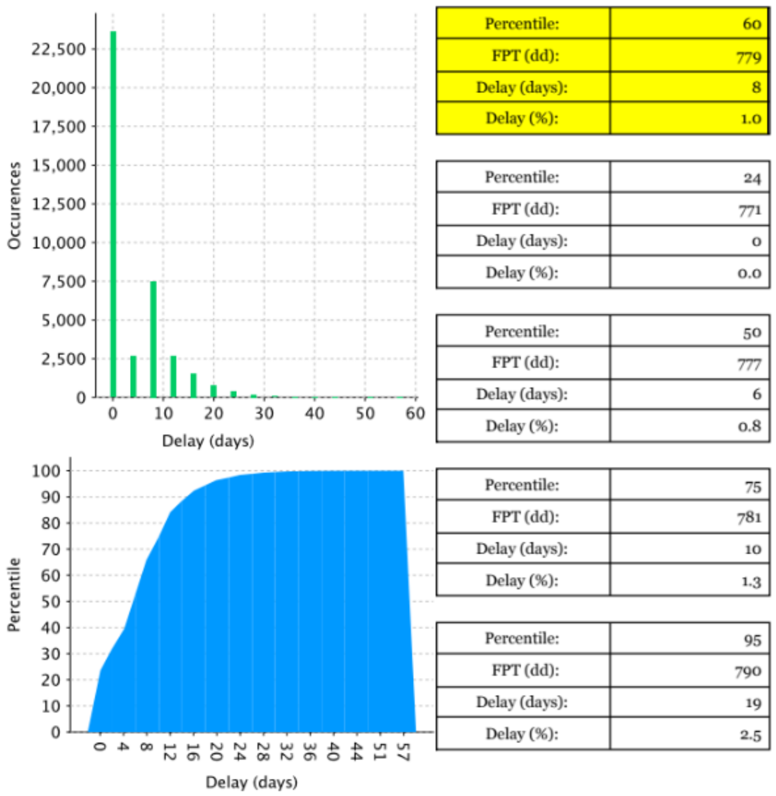



Figure 32 - Results related to the traditional construction planning approach using a real-life database.

In this case, the simulations resulted in a likely project delay equal to 8 days. Using the hybrid solution proposed in this study, the quasi birth and death model trained with the history of projects results in 38 safety events to achieve likely 8 delay events, following the stationary probabilities calculated from the generator matrix, as described in Chapter 6. Note that one delay event results in at least one delay day, and consequently at the worst case, the "planned" delay days are equal to 8. It may suggest that the "planned" delay days chosen in scenario could produce 38 safety events during the construction phase.

The proposed construction planning method, therefore, adds a new line to evaluate in the possible scenario, as shown in Table 22.

Table 22 - New output created in the proposed construction planning method.

| | |
|---|---|
| Percentile: | 60 |
| FTP (dd): | 779 |
| Delay (days): | 8 |
| Delay (%): | 1.0 |
| Prediction of safety events: | 38 |

## 7.4. Construction monitoring

In general, a company does not want to convert the planned delays into actual delays and, therefore, some techniques must be adopted during the construction phase to prevent safety events and delays. Traditionally, the S-curves are used and represent the planned and actual economical and physical progress of such construction over the time, as shown in Figure 33. The gray dashed curve represents the planned construction progress, the yellow curve represents the authorized, through financial evaluation, construction progress, the blue curve represents the actual progress, and the green curve represents the progress forecasting.

Basically, if the actual curve starts to distance from the planned curve, the construction team turns on the alert to avoid new delays. However, if this is not caught soon enough it may become so late to take action and difficult to understand the real cause of such distance between actual and planned curves.

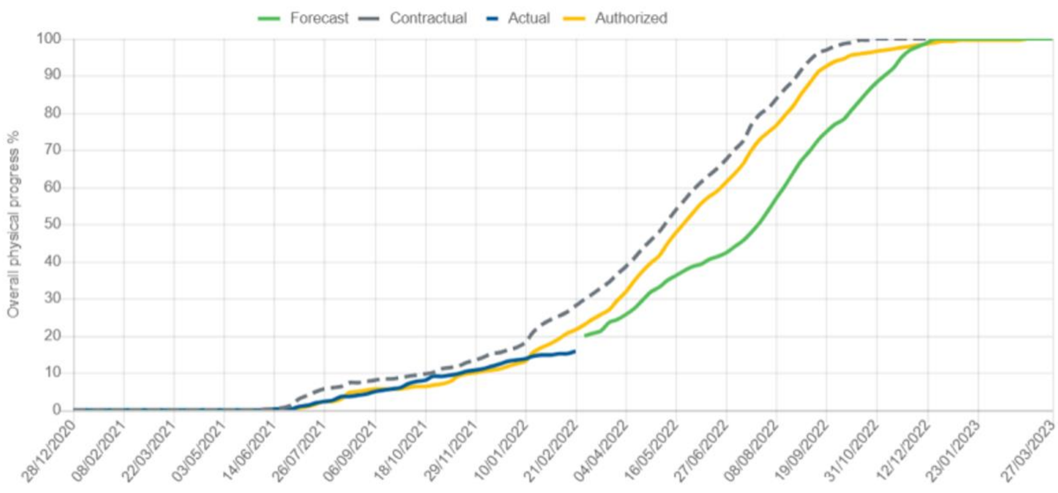



Figure 33 - Traditional approach used to monitor the construction progress using the S curves.

Therefore, the current methodology proposes the adoption of AI techniques to forecast the time-series related to safety and delays events. Differently from the stochastic model, the neural networks model uses both the database of the ongoing project and the past projects. Using the AI technique, the predictions made by the stochastic models can be updated, being more compatible with the context of the ongoing project. Due to dynamic construction context, however, the dataset used in the AI model is constantly updated during the construction phase. Therefore, the "planned" delays are updated every time that a new report is added to the database.

It is important to highlight that beyond the consideration of safety effects into project delays, the proposed methodology is helpful to avoid new accidents since the forecasting models also forecast the safety events. Note that the "zero accidents" mindset results in "planned" safety events always equal to zero during the construction planning process. However, understanding the uniqueness of such industry, this assumption is too optimistic. As verified by the proposed methodology applied to a real-world database of infrastructure projects always results in some new safety events. Therefore, the hope is that future databases could reduce the safety accidents in a manner that the model will be also "optimistic".

Another significant difference between the proposed method and traditional methodologies is that the planning method assumes that safety events will occur given the applied real database prior to the construction phase. However, it should be noted that the construction team will focus on maintaining the "zero accidents" mindset and will not allow the safety event predictions to become a reality.

Table 23 shows the main differences between the traditional and the proposed construction planning method.

Table 23 - Comparison between the traditional and proposed construction planning method.

| | **TRADITIONAL** | **PROPOSED ADDITIONAL ITEMS** |
|---|---|---|
| **PLANNING METHOD** | Probabilistic analysis with Monte Carlo simulation | Stochastic modeling with neural networks time-series forecasting |
| **PURPOSE** | To identify general and common risks that can affect the construction schedule | To capture unexpected uncertainties that can generate risks, beyond the common risks, and can affect the construction schedule, specifically in the relation between safety and delay occurrences. |
| **INPUT** | User previous experiences, lessons learned and unstructured data from previous projects | Structured data from previous projects with safety and delay records. |
| **PREMISES** | Safety events do not have significant impact on the main project's indicators – time, cost, and quality. Thus, it does not need be considered in the planning method. | Safety events play an important role, since they influence the occurrence of delay events and, thus, should be taken into account. |
| **INITIAL STEPS** | Creation of a list of identified risks to be considered in the simulations | Data preprocessing and transformation techniques applied in the raw dataset |
| **PROCESSING** | Definition of the pessimistic, realistic, and optimistic scenarios. | Calculation of the transition rates and the generator matrix in the QBDP. |
| | Number of iterations to simulate with Monte Carlo. | Precision required to converge the stationary distributions |
| **CONSTRUCTION MONITORING** | Using traditional reports and S-curve | Using bivariate time-series forecasting modeled with neural networks |
| **RESULTS** | Most likely construction delays considering the usual risks | Most likely construction delays considering random events |

# 8
# Conclusions

This chapter provides a summary of the developed methodology, emphasizing the results obtained from the proposed hybrid solution for construction safety and project management. It also highlights the main contribution of considering uncertainties related to safety during construction. Finally, the main limitations of the proposed approach and future research directions are provided.

The study of construction planning methods is a scientific demand due to the high level of uncertainty and random events associated with a construction project, such as weather conditions, political and economic context, and so on. In relation to the safety issue, the literature has not fully understood the effects of this variable on project duration. In fact, previous research has shown that safety regulations can reduce construction costs and that using the safest construction methods can shorten construction time. Most planning approaches, however, ignore the fact that safety events will occur and may have an impact on project completion. At the same time, the construction industry's safety record is among the worst in the world. As a result, the current work proposes a new construction planning method that takes into account uncertainties associated with safety events, with the goal of improving decision-making in terms of the main project indicator: time.

Consequently, this study proposes using a stochastic model called the Quasi Birth and Death Process to consider both safety and time random variables in the construction planning method. Using a real-world database, the stochastic model produced satisfactory results for the pre-construction phase, while the neural networks model produced more reliable results for the construction phase. It was also demonstrated that for bivariate time-series forecasting, the LSTM models outperformed statistical methods in the short term. Long term, both statistical and neural networks local models performed well, with the exception of the global neural networks model, which shows an increasing RMSE metric, most likely due to the generalized nature of the trained dataset.

Furthermore, exemplifying with a real-world application in Chapter 7, the current work demonstrated that using traditional planning approaches, construction planners can be assuming a high occurrence of safety events, which is completely contrary to the "zero accidents" mindset commonly promoted in large projects. It should be noted that inserting safety concerns during the preconstruction phases can result in not only a reduction in safety events but also a reduction in "planned" (or acceptable) project delays.

Therefore, as identified by the literature review presented in the initial chapters, this study contributes partially to the research field related to accidents analysis and prevention and bring more attention to the importance of uncertainties consideration in the planning methods, specifically those related to safety. Using a real-world database with 39 energy infrastructure construction projects, the current study showed that the safety occurrences impact directly the construction delays, even this assumption is not assumed in the current planning method, as discussed in Chapter 7.

This study addressed the following topic in relation to the specific objectives:

(1) Two Markov transition diagrams were proposed in Chapter 6 to represent safety and delay events in construction projects, one to analyze the impacts of safety into delay and the other from delay to safety;

(2) In Chapters 5 and 6, two different computational methods based on stochastic process theory were applied to a real-life database to understand/describe the pure-birth process, which is uniquely associated with the evolution of safety events, and the quasi birth and death process, which is linked to the safety and delay states;

(3) Chapter 7 demonstrated the significance of the KDD adoption result with the new knowledge obtained with the treated dataset, after several data preprocessing and transformation techniques, such as binarization and normalization;

(4) This study concluded from the QBDP's stationary distribution that the higher the safety state, the higher the most likely delay state,

which means that the occurrence of safety events affects the occurrence of delay events;

(5) Because of the QBDP's limitations in dealing with time series forecasting, Chapter 6 demonstrated that using the LSTM model, better predictions can be obtained for a short available period of time when compared to statistical methods.

In general, it can be concluded that the methodology proposed in this work fulfills all the objectives presented in Chapter 1. Also, the main hypothesis was verified, that is: a safety event that occurs during the construction phase affects the construction duration, and when it is not considered in the planning method, it results in dangerous construction schedules. Thus, the proposed method is able to assist construction planners in decision making when planning a new project related energy infrastructure construction project.

## 8.1. Limitations and future research

One of the main limitations of this study is that the stochastic model used is based on a complex mathematical model, which makes its application difficult for a typical user who is unfamiliar with the subject. Thus, in future works, a user-interface platform could be developed to make the proposed approach more user-friendly. Furthermore, the neural network techniques used were developed to address the shortcomings of the QBDP solution in dealing with an ongoing construction project, and thus may not be the best AI solution for bivariate time series forecasting. This study encourages future research that combines CNN and LSTM or employs other AI techniques like Prophet.

Furthermore, it is worth noting that many other random variables, such as project cost and quality, can affect and are affected by safety occurrences. The current work, however, did not take this into account because the high dimensionality of the Markov transition diagram would make it difficult to understand and mathematically formulate the generator matrix that generates the stationary distributions. Future research could attempt to combine other random variables to better understand their impact.

Finally, the KDD process required a significant amount of time spent preprocessing and transforming the available datasets from construction projects. This can be addressed by developing an information model that simplifies the system of safety and delay records and organizes the data as required by the QBDP and neural networks models. Thus, integration with BIM methodology is suggested for future work.

## References


ABBASI, Saman; TAGHIZADE, Katayoon; NOORZAI, Esmatullah. BIM-Based Combination of Takt Time and Discrete Event Simulation for Implementing Just in Time in Construction Scheduling under Constraints. **Journal of Construction Engineering and Management**, v. 146, n. 12, p. 04020143, dez. 2020.

ABOURIZK, Simaan M.; HALPIN, Daniel W. Statistical Properties of Construction Duration Data. **Journal of Construction Engineering and Management**, v. 118, n. 3, p. 525–544, set. 1992.

ABOURIZK, Simaan M.; SAWHNEY, Anil. Subjective and Interactive Duration Estimation. **Canadian Journal of Civil Engineering**, v. 20, n. 3, p. 457–470, 1 jun. 1993.

ADEDOKUN, Olufisayo; EGBELAKIN, Temitope; OMOTAYO, Temitope. Random Forest and Path Diagram Taxonomies of Risks Influencing Higher Education Construction Projects. **International Journal of Construction Management**, p. 1–9, 17 maio 2023.

ALBUQUERQUE, José Paulo de Almeida e. **Probabilidade, variáveis aleatórias e processos estocásticos**. [S.l.]: Puc-Rio, 2017.

ALHUSSEIN, Hasnaa; SHEHAB, Lynn; HAMZEH, Farook. Improvisation in Construction Planning: An Agent-Based Simulation Approach. **Buildings**, v. 12, n. 10, p. 1608, 5 out. 2022.

ALJASSMI, Hamad; ABDULJALIL, Yusef; PHILIP, Babitha. Towards Self-Recovering Construction Schedules: A New Method for Periodically Updating Project Plans and Optimizing Recovery Actions. **Journal of Asian Architecture and Building Engineering**, v. 22, n. 4, p. 2335–2347, 4 jul. 2023.

ALSHBOUL, Odey; SHEHADEH, Ali; HAMEDAT, Ola. Governmental Investment Impacts on the Construction Sector Considering the Liquidity Trap. **Journal of Management in Engineering**, v. 38, n. 2, p. 04021099, mar. 2022.

ALVARENGA, Felipe Campos et al. Alterações de custo e prazo em obras públicas. **Ambiente Construído**, v. 21, n. 1, p. 161–180, jan. 2021.

AMARAL, Glenda; BAIÃO, Fernanda; GUIZZARDI, Giancarlo. Foundational Ontologies, ONTOLOGY-DRIVEN Conceptual Modeling, and Their Multiple Benefits to Data Mining. **WIREs Data Mining and Knowledge Discovery**, v. 11, n. 4, p. e1408, jul. 2021.

AMIRI, Mehran; ARDESHIR, Abdollah; FAZEL ZARANDI, Mohammad Hossein. Fuzzy Probabilistic Expert System for Occupational Hazard Assessment in Construction. **Safety Science**, v. 93, p. 16–28, mar. 2017.

ANSARI, Ramin. Dynamic Optimization for Analyzing Effects of Multiple Resource Failures on Project Schedule Robustness. **KSCE Journal of Civil Engineering**, v. 25, n. 5, p. 1515–1532, maio 2021.

BACON, Robert W; BESANT-JONES, John E. Estimating Construction Costs and Schedules. **Energy Policy**, v. 26, n. 4, p. 317–333, mar. 1998.

BELTRAMI, Edward. While the City Burns. **Mathematical Models for Society and Biology**. [S.l.]: Elsevier, 2013. p. 39–62. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/B9780124046245000030>. Acesso em: 15 nov. 2023.

BENJAORAN, Vacharapoom; BHOKHA, Sdhabhon. An Integrated Safety Management with Construction Management Using 4D CAD Model. **Safety Science**, v. 48, n. 3, p. 395–403, mar. 2010.

BI, Lei et al. Real-Time Construction Schedule Analysis of Long-Distance Diversion Tunnels Based on Lithological Predictions Using a Markov Process. **Journal of Construction Engineering and Management**, v. 141, n. 2, p. 04014076, fev. 2015.

BIRUK, Sławomir; RZEPECKI, Łukasz. A Simulation Model of Construction Projects Executed in Random Conditions with the Overlapping Construction Works. **Sustainability**, v. 13, n. 11, p. 5795, 21 maio 2021.

BOLCH, Gunter et al. **Queueing Networks and Markov Chains**. 1. ed. [S.l.]: Wiley, 2006. Disponível em: <https://onlinelibrary.wiley.com/doi/book/10.1002/0471791571>. Acesso em: 15 nov. 2023.

CANCA, David; LAPORTE, Gilbert. Solving Real-Size Stochastic Railway Rapid Transit Network Construction Scheduling Problems. **Computers & Operations Research**, v. 138, p. 105600, fev. 2022.

CARMO, Cristiano Saad Travassos Do; SOTELINO, Elisa Dominguez. Planejamento para o imprevisto em projetos de construção: uma revisão. **Brazilian Journal of Production Engineering**, v. 9, n. 4, p. 107–130, 31 out. 2023.

CARON, F; MARCHET, G; PEREGO, A. Project Logistics: Integrating the Procurement and Construction Processes. **International Journal of Project Management**, v. 16, n. 5, p. 311–319, out. 1998.

CHAKRABORTY, Debaditya et al. A Novel Construction Cost Prediction Model Using Hybrid Natural and Light Gradient Boosting. **Advanced Engineering Informatics**, v. 46, p. 101201, out. 2020.

CHEN, Long et al. Bayesian Monte Carlo Simulation–Driven Approach for Construction Schedule Risk Inference. **Journal of Management in Engineering**, v. 37, n. 2, p. 04020115, mar. 2021.

CHEN, Long; LU, Qiuchen; HAN, Daguang. A Bayesian-Driven Monte Carlo Approach for Managing Construction Schedule Risks of Infrastructures under Uncertainty. **Expert Systems with Applications**, v. 212, p. 118810, fev. 2023.

CHENG, Fengju; ZHANG, Jingzhao. RETRACTED ARTICLE: Research on Rain Slope Stability and Digital Village Construction in Mountainous Areas Based on Remote Sensing Images. **Arabian Journal of Geosciences**, v. 14, n. 18, p. 1881, set. 2021.

CHENG, Min-Yuan et al. ENHANCED TIME-DEPENDENT EVOLUTIONARY FUZZY SUPPORT VECTOR MACHINES INFERENCE MODEL FOR CASH FLOW PREDICTION AND ESTIMATE AT COMPLETION. **International Journal of Information Technology & Decision Making**, v. 12, n. 04, p. 679–710, jul. 2013.

CHOI, Byungjoo et al. Framework for Work-Space Planning Using Four-Dimensional BIM in Construction Projects. **Journal of Construction Engineering and Management**, v. 140, n. 9, p. 04014041, set. 2014.

CHUA, D. K.; GOH, Y. M. Poisson Model of Construction Incident Occurrence. **Journal of Construction Engineering and Management**, v. 131, n. 6, p. 715–722, jun. 2005.

CLARK, Wallace. **The Gantt Chart, a Working Tool of Management**. New York: The Ronald Press Company, 1922.

CRUZ-NÁJERA, Mariel Abigail et al. Short Time Series Forecasting: Recommended Methods and Techniques. **Symmetry**, v. 14, n. 6, p. 1231, 14 jun. 2022.

DE GUNST, Mathisca; MANDJES, Michel; SOLLIE, Birgit. Statistical Inference for a Quasi Birth–Death Model of RNA Transcription. **BMC Bioinformatics**, v. 23, n. 1, p. 105, dez. 2022.

DE MEYER, A.; LOCH, C.H.; PICH, M.T. Managing project uncertainty: from variation to chaos. **IEEE Engineering Management Review**, v. 30, n. 3, p. 91–91, 2002.

DENYER, D; TRANFIELD, D. Producing a systematic review. **The Sage handbook of organizational research methods**. [S.l.]: Sage Publications Ltd, 2009. p. 671–689.

DHALMAHAPATRA, Krantiraditya et al. Exploring Causes of Crane Accidents from Incident Reports Using Decision Tree. In: SATAPATHY, Suresh Chandra; JOSHI, Amit (Org.). . **Information and Communication Technology for Intelligent Systems**. Smart Innovation, Systems and Technologies. Singapore: Springer Singapore, 2019. v. 106. p. 175–183. Disponível em:

<http://link.springer.com/10.1007/978-981-13-1742-2_18>. Acesso em: 15 nov. 2023.

DU, Jing; KIM, Byung-Cheol; ZHAO, Dong. Cost Performance as a Stochastic Process: EAC Projection by Markov Chain Simulation. **Journal of Construction Engineering and Management**, v. 142, n. 6, p. 04016009, jun. 2016.

EDMUND, Baffoe-Twum; YANG, Huojun; ERIC, Asa. A General Overview of the Impact of Global Financial Crisis on Construction Industries. In: INTERNATIONAL CONFERENCE ON CONSTRUCTION AND REAL ESTATE MANAGEMENT 2018, 8 ago. 2018, Charleston, South Carolina. *Anais*... Charleston, South Carolina: American Society of Civil Engineers, 8 ago. 2018. p. 28–38. Disponível em: <https://ascelibrary.org/doi/10.1061/9780784481745.004>. Acesso em: 15 nov. 2023.

ELENA BRUNI, Maria et al. A Scheduling Methodology for Dealing with Uncertainty in Construction Projects. **Engineering Computations**, v. 28, n. 8, p. 1064–1078, 15 nov. 2011.

ELHAKEEM, Ahmed; HEGAZY, Tarek. Graphical Approach for Manpower Planning in Infrastructure Networks. **Journal of Construction Engineering and Management**, v. 131, n. 2, p. 168–175, fev. 2005.

ENSHASSI, Adnan; MOHAMED, Sherif; ABUSHABAN, Saleh. FACTORS AFFECTING THE PERFORMANCE OF CONSTRUCTION PROJECTS IN THE GAZA STRIP. **JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT**, v. 15, n. 3, p. 269–280, 30 jun. 2009.

ESPINOZA, R. David. Contingency Estimating Using Option Pricing Theory: Closing the Gap between Theory and Practice. **Construction Management and Economics**, v. 29, n. 9, p. 913–927, set. 2011.

FADILOGLU, M.M.; YERALAN, S. Models of Production Lines as Quasi-Birth-Death Processes. **Mathematical and Computer Modelling**, v. 35, n. 7–8, p. 913–930, abr. 2002.

FARSHCHIAN, Mohammad Mahdi; HERAVI, Gholamreza. Probabilistic Assessment of Cost, Time, and Revenue in a Portfolio of Projects Using Stochastic Agent-Based Simulation. **Journal of Construction Engineering and Management**, v. 144, n. 5, p. 04018028, maio 2018.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. **AI MAGAZINE**, 1996. Disponível em: <https://doi.org/10.1609/aimag.v17i3.1230>.

FENG, Kailun et al. A Predictive Environmental Assessment Method for Construction Operations: Application to a Northeast China Case Study. **Sustainability**, v. 10, n. 11, p. 3868, 24 out. 2018.

______. Planning Construction Projects in Deep Uncertainty: A Data-Driven Uncertainty Analysis Approach. **Journal of Construction Engineering and Management**, v. 148, n. 8, p. 04022060, ago. 2022.

FITZSIMMONS, John Patrick et al. Construction Schedule Risk Analysis – a Hybrid Machine Learning Approach. **Journal of Information Technology in Construction**, v. 27, p. 70–93, 13 jan. 2022.

FONDAHL, John W. **A NON-COMPUTER APPROACH TO THE CRITICAL PATH METHOD FOR THE CONSTRUCTION INDUSTRY**. . [S.l.]: Stanford University. , 1962

FONG, Simon James et al. Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak. **International Journal of Interactive Multimedia and Artificial Intelligence**, v. 6, n. 1, p. 132, 2020.

FORD, David N.; LANDER, Diane M.; VOYER, John J. A Real Options Approach to Valuing Strategic Flexibility in Uncertain Construction Projects. **Construction Management and Economics**, v. 20, n. 4, p. 343–351, jun. 2002.

FRANCIS, Adel. Chronographical Spatiotemporal Scheduling Optimization for Building Projects. **Frontiers in Built Environment**, v. 5, p. 36, 2 abr. 2019.

GEISSDOERFER, Martin et al. The Circular Economy – A New Sustainability Paradigm? **Journal of Cleaner Production**, v. 143, p. 757–768, fev. 2017.

GERASSIS, S. et al. Bayesian Decision Tool for the Analysis of Occupational Accidents in the Construction of Embankments. **Journal of Construction Engineering and Management**, v. 143, n. 2, p. 04016093, fev. 2017.

GOH, Yang Miang; ASKAR ALI, Mohamed Jawad. A Hybrid Simulation Approach for Integrating Safety Behavior into Construction Planning: An Earthmoving Case Study. **Accident Analysis & Prevention**, v. 93, p. 310–318, ago. 2016.

GOLABCHI, Alireza et al. An Automated Biomechanical Simulation Approach to Ergonomic Job Analysis for Workplace Design. **Journal of Construction Engineering and Management**, v. 141, n. 8, p. 04015020, ago. 2015.

HADIPRIONO, Fabian C. Fault Tree Network Analysis for Construction. **Civil Engineering Systems**, v. 5, n. 1, p. 42–48, mar. 1988.

HALPIN, Daniel W. Cyclone–Method for Modeling Job Site Processes. **Journal of the Construction Division**, v. 103, n. 3, p. 489–499, set. 1977.

HASSAN, Abbas; EL-RAYES, Khaled; ATTALLA, Mohamed. Stochastic Scheduling Optimization of Repetitive Construction Projects to Minimize Project Duration and Cost. **International Journal of Construction Management**, v. 23, n. 9, p. 1447–1456, 4 jul. 2023.

HAUTPHENNE, Sophie; PATCH, Brendan. Birth-and-death Processes in Python: The BirDePy Package. 2021. Disponível em: <https://arxiv.org/abs/2110.05067>. Acesso em: 15 nov. 2023.

HAYKIN, Simon S. **Neural networks and learning machines**. 3rd ed ed. New York: Prentice Hall, 2009.

HE, Qinghua et al. Mapping the Managerial Areas of Building Information Modeling (BIM) Using Scientometric Analysis. **International Journal of Project Management**, v. 35, n. 4, p. 670–685, maio 2017.

HILL, Sue A. Statistics. **Foundations of Anesthesia**. [S.l.]: Elsevier, 2006. p. 207–217. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/B9780323037075500243>. Acesso em: 15 nov. 2023.

HINZE, Jimmie; DEVENPORT, Jenny N.; GIANG, Geneva. Analysis of Construction Worker Injuries That Do Not Result in Lost Time. **Journal of Construction Engineering and Management**, v. 132, n. 3, p. 321–326, mar. 2006.

HONG, Ying et al. A Graph-Based Approach for Unpacking Construction Sequence Analysis to Evaluate Schedules. **Advanced Engineering Informatics**, v. 52, p. 101625, abr. 2022.

HORVATH, Gabor; TELEK, Miklos. BuTools 2: A Rich Toolbox for Markovian Performance Evaluation. In: 10TH EAI INTERNATIONAL CONFERENCE ON PERFORMANCE EVALUATION METHODOLOGIES AND TOOLS, 2017, Taormina, Italy. *Anais*... Taormina, Italy: ACM, 2017. Disponível em: <http://eudl.eu/doi/10.4108/eai.25-10-2016.2266400>. Acesso em: 15 nov. 2023.

HOSNY, Abdelhady; NIK-BAKHT, Mazdak; MOSELHI, Osama. Physical Distancing Analytics for Construction Planning Using 4D BIM. **Journal of Computing in Civil Engineering**, v. 36, n. 4, p. 04022012, jul. 2022.

______. Workspace Planning in Construction: Non-Deterministic Factors. **Automation in Construction**, v. 116, p. 103222, ago. 2020.

HOSSEN, Muhammed Mufazzal; KANG, Sunkoo; KIM, Jonghyun. Construction Schedule Delay Risk Assessment by Using Combined AHP-RII Methodology for an International NPP Project. **Nuclear Engineering and Technology**, v. 47, n. 3, p. 362–379, abr. 2015.

HSU, Pei-Yuan; ANGELOUDIS, Panagiotis; AURISICCHIO, Marco. Optimal Logistics Planning for Modular Construction Using Two-Stage Stochastic Programming. **Automation in Construction**, v. 94, p. 47–61, out. 2018.

HU, Songbo; FANG, Yihai; MOEHLER, Robert. Estimating and Visualizing the Exposure to Tower Crane Operation Hazards on Construction Sites. **Safety Science**, v. 160, p. 106044, abr. 2023.

HUANG, Rui et al. Accident Prevention Analysis: Exploring the Intellectual Structure of a Research Field. **Sustainability**, v. 14, n. 14, p. 8784, 18 jul. 2022.

HUSIN, Albert Eddy. Time performance upgrade by critical chain project management and BIM 4D integration on top structural work of a high rise building construction project. ARPN Journal of Engineering and Applied Sciences, 2019.

ISAH, Muritala Adebayo; KIM, Byung-Soo. Integrating Schedule Risk Analysis with Multi-Skilled Resource Scheduling to Improve Resource-Constrained Project Scheduling Problems. **Applied Sciences**, v. 11, n. 2, p. 650, 11 jan. 2021.

JALALI, Samireh; WOHLIN, Claes. Systematic Literature Studies: Database Searches vs. Backward Snowballing. In: ESEM '12: 2012 ACM-IEEE INTERNATIONAL SYMPOSIUM ON EMPIRICAL SOFTWARE ENGINEERING AND MEASUREMENT, 19 set. 2012, Lund Sweden. *Anais*... Lund Sweden: ACM, 19 set. 2012. p. 29–38. Disponível em: <https://dl.acm.org/doi/10.1145/2372251.2372257>. Acesso em: 15 nov. 2023.

JANARDAN, K.G. A Stochastic Model for the Number of Accidents and a Class of Distributions for the Number of Fatalities. **Communications in Statistics - Theory and Methods**, v. 27, n. 7, p. 1823–1833, jan. 1998.

JAŚKOWSKI, Piotr; SOBOTKA, Anna. Scheduling Construction Projects Using Evolutionary Algorithm. **Journal of Construction Engineering and Management**, v. 132, n. 8, p. 861–870, ago. 2006.

KAMMOUH, O. et al. Dynamic Control for Construction Project Scheduling On-the-Run. **Automation in Construction**, v. 141, p. 104450, set. 2022.

KARIM, Aras Jalal Mhamad; AHMED, Nawzad Muhammed. Vector Autoregressive Integrating Moving Average (Varima) Model of COVID-19 Pandemic and Oil Price. **International Journal of Professional Business Review**, v. 8, n. 1, p. e0988, 31 jan. 2023.

KARLIN, Samuel; MCGREGOR, James. REPRESENTATION OF A CLASS OF STOCHASTIC Processes. **Proceedings of the National Academy of Sciences**, v. 41, n. 6, p. 387–391, 15 jun. 1955.

KAVANAGH, Donncha P. Siren: A Repetitive Construction Simulation Model. **Journal of Construction Engineering and Management**, v. 111, n. 3, p. 308–323, set. 1985.

KAVEH, Ali; RAJABI, Farivar; MIRVALAD, Sajjad. Many-Objective Optimization for Construction Project Scheduling Using Non-Dominated Sorting Differential Evolution Algorithm Based on Reference Points. **Scientia Iranica**, v. 0, n. 0, p. 0–0, 5 out. 2021.

KEDIR, Nebiyu Siraj et al. Hybridization of Reinforcement Learning and Agent-Based Modeling to Optimize Construction Planning and Scheduling. **Automation in Construction**, v. 142, p. 104498, out. 2022.

KELLEY, James E.; WALKER, Morgan R. Critical-Path Planning and Scheduling. In: PAPERS PRESENTED AT THE DECEMBER 1-3, 1959, EASTERN JOINT IRE-AIEE-ACM COMPUTER CONFERENCE, 1959, Boston, Massachusetts. *Anais*... Boston, Massachusetts: ACM Press, 1959. p. 160–173. Disponível em: <http://portal.acm.org/citation.cfm?doid=1460299.1460318>. Acesso em: 15 nov. 2023.

KENDALL, David G. On the Generalized “Birth-and-Death” Process. **The Annals of Mathematical Statistics**, v. 19, n. 1, p. 1–15, mar. 1948.

KHAN, Khalid S et al. Five Steps to Conducting a Systematic Review. **Journal of the Royal Society of Medicine**, v. 96, n. 3, p. 118–121, mar. 2003.

KHAROUFEH, Jeffrey P. Level-Dependent Quasi-Birth-and-Death Processes. In: COCHRAN, James J. et al. **Wiley Encyclopedia of Operations Research and Management Science**. 1. ed. [S.l.]: Wiley, 2011. . Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/9780470400531.eorms0460>. Acesso em: 15 nov. 2023.

KIM, Jin-Lee; ELLIS, Ralph D. Robust Global and Local Search Approach to Resource-Constrained Project Scheduling. **Canadian Journal of Civil Engineering**, v. 36, n. 3, p. 375–388, mar. 2009.

KOEHN, Enno; MUSSER, Kurt. OSHA Regulations Effects on Construction. **Journal of Construction Engineering and Management**, v. 109, n. 2, p. 233–244, jun. 1983.

KOKKAEW, Nakhon; CHIARA, Nicola. Modelling Completion Risk Using Stochastic Critical Path-envelope Method: A BOT Highway Project Application. **Construction Management and Economics**, v. 28, n. 12, p. 1239–1254, dez. 2010.

KULEJEWSKI, Janusz et al. Cash Flow Optimization for Renewable Energy Construction Projects with a New Approach to Critical Chain Scheduling. **Energies**, v. 14, n. 18, p. 5795, 14 set. 2021.

LAUFER, Alexander; COHENCA, Dora. Factors Affecting Construction-Planning Outcomes. **Journal of Construction Engineering and Management**, v. 116, n. 1, p. 135–156, mar. 1990.

LE ISIDORE, Roy J.; EDWARD BACK, W.; FRY, Gary T. Integrated Probabilistic Schedules and Estimates from Project Simulated Data. **Construction Management and Economics**, v. 19, n. 4, p. 417–426, jul. 2001.

LEE, Dong-Eun. Probability of Project Completion Using Stochastic Project Scheduling Simulation. **Journal of Construction Engineering and Management**, v. 131, n. 3, p. 310–318, mar. 2005.

LEE, Hyun-soo et al. Probabilistic Duration Estimation Model for High-Rise Structural Work. **Journal of Construction Engineering and Management**, v. 135, n. 12, p. 1289–1298, dez. 2009.

LEMPERT, Robert; POPPER, Steven; BANKES, Steven. **Shaping the Next One Hundred Years**. [S.l.]: RAND Corporation, 2003. Disponível em: <https://www.rand.org/pubs/monograph_reports/MR1626.html>. Acesso em: 15 nov. 2023.

LEU, Sou-Sen; CHEN, An-Ting; YANG, Chung-Huei. A GA-Based Fuzzy Optimal Model for Construction Time–Cost Trade-Off. **International Journal of Project Management**, v. 19, n. 1, p. 47–58, jan. 2001.

______. Fuzzy Optimal Model for Resource-Constrained Construction Scheduling. **Journal of Computing in Civil Engineering**, v. 13, n. 3, p. 207–216, jul. 1999.

LEU, Sou-Sen; HUNG, Tzung-Heng. A Genetic Algorithm-Based Optimal Resource-Constrained Scheduling Simulation Model. **Construction Management and Economics**, v. 20, n. 2, p. 131–141, mar. 2002.

LI, Heng et al. Optimizing Construction Planning Schedules by Virtual Prototyping Enabled Resource Analysis. **Automation in Construction**, v. 18, n. 7, p. 912–918, nov. 2009.

______. Stochastic State Sequence Model to Predict Construction Site Safety States through Real-Time Location Systems. **Safety Science**, v. 84, p. 78–87, abr. 2016.

LI, Qian et al. Robust Optimization for Integrated Construction Scheduling and Multiscale Resource Allocation. **Complexity**, v. 2018, p. 1–17, 24 jul. 2018.

LI, Xin et al. MetaInjury: Meta-Learning Framework for Reusing the Risk Knowledge of Different Construction Accidents. **Safety Science**, v. 140, p. 105315, ago. 2021.

LIANG, Huakang; ZHANG, Shoujian; SU, Yikun. The Structure and Emerging Trends of Construction Safety Management Research: A Bibliometric Review. **International Journal of Occupational Safety and Ergonomics**, v. 26, n. 3, p. 469–488, 2 jul. 2020.

LIU, Xun; SHEN, Le; ZHANG, Kun. Estimating the Probability Distribution of Construction Project Completion Times Based on Drum-Buffer-Rope Theory. **Applied Sciences**, v. 11, n. 15, p. 7150, 2 ago. 2021.

MA, Guofeng et al. Improved Critical Chain Project Management Framework for Scheduling Construction Projects. **Journal of Construction Engineering and Management**, v. 140, n. 12, p. 04014055, dez. 2014.

MALCOLM, D. G. et al. Application of a Technique for Research and Development Program Evaluation. **Operations Research**, v. 7, n. 5, p. 646–669, out. 1959.

MANDJES, Michel; SOLLIE, Birgit. A Numerical Approach for Evaluating the Time-Dependent Distribution of a Quasi Birth-Death Process. **Methodology and Computing in Applied Probability**, v. 24, n. 3, p. 1693–1715, set. 2022.

MANIK, Anshu et al. NEURAL NETWORKS SURROGATE MODELS FOR SIMULATING PAYMENT RISK IN PAVEMENT CONSTRUCTION/NEURONINIŲ TINKLŲ PAKAITALO MODELIS MOKĖJIMO RIZIKAI KELIŲ STATYBOJE NUSTATYTI. **JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT**, v. 14, n. 4, p. 235–240, 31 dez. 2008.

MARHAVILAS, P.K.; KOULOURIOTIS, D.E.; SPARTALIS, S.H. Harmonic Analysis of Occupational-Accident Time-Series as a Part of the Quantified Risk Evaluation in Worksites: Application on Electric Power Industry and Construction Sector. **Reliability Engineering & System Safety**, v. 112, p. 8–25, abr. 2013.

MARONATI, G.; PETROVIC, B. Estimating Cost Uncertainties in Nuclear Power Plant Construction through Monte Carlo Sampled Correlated Random Variables. **Progress in Nuclear Energy**, v. 111, p. 211–222, mar. 2019.

MASSEY, Frank J. The Kolmogorov-Smirnov Test for Goodness of Fit. **Journal of the American Statistical Association**, v. 46, n. 253, p. 68–78, mar. 1951.

MCCARTHY, J. et al. **A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE**. . [S.l: s.n.]. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 30 nov. 2023. , [s.d.]

MIHIĆ, Matej; VUKOMANOVIĆ, Mladen; ZAVRŠKI, Ivica. Review of Previous Applications of Innovative Information Technologies in Construction Health and Safety. **Organization, Technology and Management in Construction: An International Journal**, v. 11, n. 1, p. 1952–1967, 1 jan. 2019.

MILAT, Martina; KNEZIĆ, Snježana; SEDLAR, Jelena. Application of a Genetic Algorithm for Proactive Resilient Scheduling in Construction Projects. **Designs**, v. 6, n. 1, p. 16, 15 fev. 2022.

MOHAMED, Emad et al. Simulation-Based Approach for Lookahead Scheduling of Onshore Wind Projects Subject to Weather Risk. **Sustainability**, v. 13, n. 18, p. 10060, 8 set. 2021.

MOON, Hyounseok et al. BIM-Based Construction Scheduling Method Using Optimization Theory for Reducing Activity Overlaps. **Journal of Computing in Civil Engineering**, v. 29, n. 3, p. 04014048, maio 2015.

MOORE, Laurence J.; CLAYTON, Edward R. **GERT modeling and simulation**. 1st ed ed. New York: Petrocelli/Charter, 1976. (Petrocelli/Charter modern decision analysis series).

MULHOLLAND, B.; CHRISTIAN, J. Risk Assessment in Construction Schedules. **Journal of Construction Engineering and Management**, v. 125, n. 1, p. 8–15, jan. 1999.

MÜLLER, Andreas C.; GUIDO, Sarah. **Introduction to machine learning with Python**. First edition ed. Sebastopol, CA: O'Reilly Media, Inc, 2016.

NASIR, Daud; MCCABE, Brenda; HARTONO, Loesie. Evaluating Risk in Construction–Schedule Model (ERIC–S): Construction Schedule Risk Model. **Journal of Construction Engineering and Management**, v. 129, n. 5, p. 518–527, out. 2003.

NG, S. Thomas; FANG, Yuan; UGWU, Onuegbu O. Modelling Construction Material Logistics System with Stochastic Petri Nets. **Construction Innovation**, v. 8, n. 1, p. 46–60, 18 jan. 2008.

NICHOLSON, Alan; WONG, Yiik-Diew. Are Accidents Poisson Distributed? A Statistical Test. **Accident Analysis & Prevention**, v. 25, n. 1, p. 91–97, fev. 1993.

OCK, Jong-Ho; HAN, Seung-Heon. Measuring Risk-Associated Activity's Duration: A Fuzzy Set Theory Application. **KSCE Journal of Civil Engineering**, v. 14, n. 5, p. 663–671, set. 2010.

OSOGAMI, Takayuki. **Analysis of multi-server systems via dimensionality reduction of Markov chains**. 2005. Carnegie Mellon University, 2005. Disponível em: <https://www.cs.cmu.edu/~osogami/thesis/>.

OSPINA, Raydonal et al. An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil. **Mathematics**, v. 11, n. 14, p. 3069, 12 jul. 2023.

OZDEMIR, Burak; KUMRAL, Mustafa. Stochastic Assessment of the Material Haulage Efficiency in the Earthmoving Industry. **Journal of Construction Engineering and Management**, v. 143, n. 8, p. 05017013, ago. 2017.

PAN, Nang-Fei. Assessment of Productivity and Duration of Highway Construction Activities Subject to Impact of Rain. **Expert Systems with Applications**, v. 28, n. 2, p. 313–326, fev. 2005.

PAPASTEFANOPOULOS, Vasilis et al. Multivariate Time-Series Forecasting: A Review of Deep Learning Methods in Internet of Things Applications to Smart Cities. **Smart Cities**, v. 6, n. 5, p. 2519–2552, 23 set. 2023.

PARDO, Robert. **Design, Testing, and Optimization of Trading Systems**. New York, NY: Wiley, 1992. (Wiley trader's advantage series).

PAWAN, Paijit; LORTERAPONG, Pasit. A Fuzzy-Based Integrated Framework for Assessing Time Contingency in Construction Projects. **Journal of Construction Engineering and Management**, v. 142, n. 3, p. 04015083, mar. 2016.

PENG, Junlong et al. Research on Extremely Short Construction Period of Engineering Project Based on Labor Balance under Resource Tolerance. **PLOS ONE**, v. 17, n. 3, p. e0266036, 30 mar. 2022.

PÉREZGONZÁLEZ, Jose D. **Construction Safety Management, a Systems Approach**. Knowledge Management ed ed. Dublin: The Author, 2005.

PINTO, Abel; NUNES, Isabel L.; RIBEIRO, Rita A. Occupational Risk Assessment in Construction Industry – Overview and Reflection. **Safety Science**, v. 49, n. 5, p. 616–624, jun. 2011.

PLEBANKIEWICZ, Edyta; ZIMA, Krzysztof; WIECZOREK, Damian. MODELLING OF TIME, COST AND RISK OF CONSTRUCTION WITH

USING FUZZY LOGIC. **JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT**, v. 27, n. 6, p. 412–426, 15 jul. 2021.

RAHMAN, Md Mahbubur et al. An Automated Methodology for Worker Path Generation and Safety Assessment in Construction Projects. **IEEE Transactions on Automation Science and Engineering**, v. 15, n. 2, p. 479–491, abr. 2018.

RAMANI, Prasanna Venkatesan; ARUN KUMAR, Talanti Ravi. Developing a Schedule Integrated Automated Safety Planning Tool for Residential Construction Projects. **International Journal of Occupational Safety and Ergonomics**, v. 29, n. 2, p. 747–755, 3 abr. 2023.

RIVAS, T. et al. Explaining and Predicting Workplace Accidents Using Data-Mining Techniques. **Reliability Engineering & System Safety**, v. 96, n. 7, p. 739–747, jul. 2011.

ROZENFELD, Ophir; SACKS, Rafael; ROSENFELD, Yehiel. 'CHASTE': Construction Hazard Assessment with Spatial and Temporal Exposure. **Construction Management and Economics**, v. 27, n. 7, p. 625–638, jul. 2009.

RUSSELL, Alan D.; WONG, William C. M. New Generation of Planning Structures. **Journal of Construction Engineering and Management**, v. 119, n. 2, p. 196–214, jun. 1993.

SACKS, R.; ROZENFELD, O.; ROSENFELD, Y. Spatial and Temporal Exposure to Safety Hazards in Construction. **Journal of Construction Engineering and Management**, v. 135, n. 8, p. 726–736, ago. 2009.

SALES, Tiago Prince et al. The Common Ontology of Value and Risk. In: TRUJILLO, Juan C. et al. (Org.). . **Conceptual Modeling**. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018. v. 11157. p. 121–135. Disponível em: <http://link.springer.com/10.1007/978-3-030-00847-5_11>. Acesso em: 15 nov. 2023.

SANDHYAVITRI, Ari. Stochastic Analyses for Managing Risk of Delay in Duri Oil Construction Projects, Indonesia. **International Journal of Construction Management**, v. 22, n. 4, p. 711–731, 12 mar. 2022.

SARI, Mehmet et al. Stochastic Modeling of Accident Risks Associated with an Underground Coal Mine in Turkey. **Safety Science**, v. 47, n. 1, p. 78–87, jan. 2009.

SARKAR, Debasis; JHA, K. N.; PATEL, Shraddha. Critical Chain Project Management for a Highway Construction Project with a Focus on Theory of Constraints. **International Journal of Construction Management**, v. 21, n. 2, p. 194–207, 1 fev. 2021.

SARKAR, Sobhan et al. Decision Support System for Prediction of Occupational Accident: A Case Study from a Steel Plant. In: ABRAHAM, Ajith et al. (Org.). . **Emerging Technologies in Data Mining and Information Security**. Advances in Intelligent Systems and Computing. Singapore: Springer Singapore, 2019. v. 813. p. 787–796. Disponível em: <http://link.springer.com/10.1007/978-981-13-1498-8_69>. Acesso em: 15 nov. 2023.

SAWHNEY, Anil; ABOURIZK, Simaan M; HALPIN, Daniel W. Construction Project Simulation Using CYCLONE. **Canadian Journal of Civil Engineering**, v. 25, n. 1, p. 16–25, 1 jan. 1998.

SCOTT L. MILLER; DONALD CHILDERS. **Probability and Random Processes**. [S.l.]: Elsevier, 2012. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/C20100676115>. Acesso em: 15 nov. 2023.

SHAMS ABADI, Seyedeh Tannaz et al. BIM-Based Co-Simulation of Fire and Occupants' Behavior for Safe Construction Rehabilitation Planning. **Fire**, v. 4, n. 4, p. 67, 4 out. 2021.

SHARMA, Virok et al. Machine Learning-Aided Cost Prediction and Optimization in Construction Operations. **Engineering, Construction and Architectural Management**, 28 abr. 2021. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/ECAM-10-2020-0778/full/html>. Acesso em: 15 nov. 2023.

SHIBANI, Abdussalam; HASSAN, Dyaa; SHAKIR, Nehal. The Effects of Pandemic on Construction Industry in the UK. **Mediterranean Journal of Social Sciences**, v. 11, n. 6, p. 48, 21 nov. 2020.

SHIN, Dong-Pil et al. Association Rules Mined from Construction Accident Data. **KSCE Journal of Civil Engineering**, v. 22, n. 4, p. 1027–1039, abr. 2018.

SILVA, Luciana O.; ZÁRATE, Luis E. A brief review of the main approaches for treatment of missing data. **Intelligent Data Analysis**, v. 18, n. 6, p. 1177–1198, 29 out. 2014.

SOLTANI, A.R.; FERNANDO, T. A Fuzzy Based Multi-Objective Path Planning of Construction Sites. **Automation in Construction**, v. 13, n. 6, p. 717–734, nov. 2004.

STEPCHENKO, Arthur et al. Nonlinear, Non-Stationary and Seasonal Time Series Forecasting Using Different Methods Coupled with Data Preprocessing. **Procedia Computer Science**, v. 104, p. 578–585, 2017.

SU, Yi; ISAAC, Shabtai; LUCKO, Gunnar. Integrated Temporal-Spatial Model for Construction Plans with Boolean Logic Operators. **Journal of Construction Engineering and Management**, v. 144, n. 4, p. 04018009, abr. 2018.

SUKUMARAN, Praveen et al. Validation of a Model for Predicting Schedule Changes in Highway Work Zones—Case Studies. **Journal of Transportation Engineering**, v. 132, n. 8, p. 638–648, ago. 2006.

SZCZESNY, Kamil; KÖNIG, Markus. Reactive Scheduling Based on Actual Logistics Data by Applying Simulation-Based Optimization. **Visualization in Engineering**, v. 3, n. 1, p. 10, dez. 2015.

TAGHADDOS, Hosein; HEYDARI, Mohammad Hosein; ASGARI, AmirHosein. A Hybrid Simulation Approach for Site Layout Planning in Construction Projects. **Construction Innovation**, v. 21, n. 3, p. 417–440, 29 jul. 2021.

TAM, C.M.; ZENG, S.X.; DENG, Z.M. Identifying Elements of Poor Construction Safety Management in China. **Safety Science**, v. 42, n. 7, p. 569–586, ago. 2004.

TEIZER, Jochen et al. Autonomous Pro-Active Real-Time Construction Worker and Equipment Operator Proximity Safety Alert System. **Automation in Construction**, v. 19, n. 5, p. 630–640, ago. 2010.

TESFAYE, Erimas et al. A Simulation of Project Completion Probability Using Different Probability Distribution Functions. In: ABRAHAM, Ajith; KRÖMER, Pavel; SNASEL, Vaclav (Org.). . **Afro-European Conference for Industrial Advancement**. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, 2015. v. 334. p. 133–145. Disponível em: <https://link.springer.com/10.1007/978-3-319-13572-4_11>. Acesso em: 15 nov. 2023.

TOMCZAK, Michał. Modeling of the Harmonization Method for Executing a Multi-Unit Construction Project. **Open Engineering**, v. 9, n. 1, p. 282–291, 25 jul. 2019.

TOURAN, Ali. Computer Packages for Construction Data Acquisition and Processing. **Construction Management and Economics**, v. 4, n. 3, p. 233–243, dez. 1986.

TRAN, Duc Hoc; LONG, Luong Duc. Project Scheduling with Time, Cost and Risk Trade-off Using Adaptive Multiple Objective Differential Evolution. **Engineering, Construction and Architectural Management**, v. 25, n. 5, p. 623–638, 18 jun. 2018.

TRILLO CABELLO, Antonio et al. Occupational Accident Analysis According to Professionals of Different Construction Phases Using Association Rules. **Safety Science**, v. 144, p. 105457, dez. 2021.

TSENG, Chung-Li; ZHAO, Tong; FU, Chung C. Contingency Estimation Using a Real Options Approach. **Construction Management and Economics**, v. 27, n. 11, p. 1073–1087, nov. 2009.

TURING, A. M. I.—COMPUTING MACHINERY AND INTELLIGENCE. **Mind**, v. LIX, n. 236, p. 433–460, 1 out. 1950.

VAN ECK, Nees Jan; WALTMAN, Ludo. Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. **Scientometrics**, v. 84, n. 2, p. 523–538, ago. 2010.

VAN LEEUWAARDEN, l.S.H.; WINANDS, E.M.M. **Quasi-birth-death processes with an explicit rate matrix**. , Reports in Statistics, Probability and Operations Research., nº 2004–08. [S.l.]: Technische Universiteit Eindhoven, 2004. Disponível em: <https://research.tue.nl/en/publications/quasi-birth-death-processes-with-an-explicit-rate-matrix>.

VAZIRI, Kabeh; CARR, Paul G.; NOZICK, Linda K. Project Planning for Construction under Uncertainty with Limited Resources. **Journal of Construction Engineering and Management**, v. 133, n. 4, p. 268–276, abr. 2007.

WAEHRER, Geetha M. et al. Costs of Occupational Injuries in Construction in the United States. **Accident Analysis & Prevention**, v. 39, n. 6, p. 1258–1266, nov. 2007.

WALKER, Warren E.; LEMPERT, Robert J.; KWAKKEL, Jan H. Deep Uncertainty. In: GASS, Saul I.; FU, Michael C. (Org.). . **Encyclopedia of Operations Research and Management Science**. Boston, MA: Springer US, 2013. p. 395–402. Disponível em: <http://link.springer.com/10.1007/978-1-4419-1153-7_1140>. Acesso em: 15 nov. 2023.

WALLACE, Victor L. **The solution of quasi birth and death processes arising from multiple access computer systems**. 1969. The University of Michigan, 1969. Disponível em: <https://deepblue.lib.umich.edu/handle/2027.42/8180>.

WANG, Bo; WANG, Ying; CUI, Li Jie. Decision of Mechanical Allocation in Subgrade Earthwork Construction under Uncertainty. **International Journal of Information and Communication Technology**, v. 15, n. 3, p. 275, 2019.

WANG, Chao et al. A Real-Time Online Structure-Safety Analysis Approach Consistent with Dynamic Construction Schedule of Underground Caverns. **Journal of Construction Engineering and Management**, v. 142, n. 9, p. 04016042, set. 2016.

WANG, Shuo; FENG, Kailun; WANG, Yaowu. Modeling Performance and Uncertainty of Construction Planning under Deep Uncertainty: A Prediction Interval Approach. **Buildings**, v. 13, n. 1, p. 254, 16 jan. 2023.

WILLIAMS, Terry. Allocation of Contingency in Activity Duration Networks. **Construction Management and Economics**, v. 17, n. 4, p. 441–447, jul. 1999.

WOHLIN, Claes. Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In: EASE ’14: 18TH INTERNATIONAL CONFERENCE ON EVALUATION AND ASSESSMENT IN SOFTWARE ENGINEERING, 13 maio 2014, London England United Kingdom. *Anais*... London England United Kingdom: ACM, 13 maio 2014. p. 1–10. Disponível em: <https://dl.acm.org/doi/10.1145/2601248.2601268>. Acesso em: 15 nov. 2023.

WOOLERY, John C.; CRANDALL, Keith C. Stochastic Network Model for Planning Scheduling. **Journal of Construction Engineering and Management**, v. 109, n. 3, p. 342–354, set. 1983.

XU, Xiaoxiao; ZOU, Patrick X.W. Discovery of New Safety Knowledge from Mining Large Injury Dataset in Construction. **Safety Science**, v. 144, p. 105481, dez. 2021.

YANG, I-Tung. Chance-Constrained Time–Cost Tradeoff Analysis Considering Funding Variability. **Journal of Construction Engineering and Management**, v. 131, n. 9, p. 1002–1012, set. 2005.

YANG, I-Tung; CHANG, Chi-Yi. Stochastic Resource-Constrained Scheduling for Repetitive Construction Projects with Uncertain Supply of Resources and Funding. **International Journal of Project Management**, v. 23, n. 7, p. 546–553, out. 2005.

ZAALOUK, Ahmed; HAN, SangHyeok. Parameterized Design Optimization Framework for Worker-Friendly Workplaces in Modular Construction. **Journal of Construction Engineering and Management**, v. 147, n. 5, p. 04021030, maio 2021.

ZAHRAIE, Banafsheh; TAVAKOLAN, Mehdi. Stochastic Time-Cost-Resource Utilization Optimization Using Nondominated Sorting Genetic Algorithm and Discrete Fuzzy Sets. **Journal of Construction Engineering and Management**, v. 135, n. 11, p. 1162–1171, nov. 2009.

ZHANG, Fan et al. Construction Site Accident Analysis Using Text Mining and Natural Language Processing Techniques. **Automation in Construction**, v. 99, p. 238–248, mar. 2019.

ZHANG, Hong; LIN, Yilei. Prediction of Ergonomic Risks and Impacts on Construction Schedule through Agent-Based Simulation. **Journal of Construction Engineering and Management**, v. 149, n. 3, p. 04022181, mar. 2023.

ZHANG, Shen; WANG, Xingyu. Dynamic Probability Analysis for Construction Schedule Using Subset Simulation. **Advances in Civil Engineering**, v. 2021, p. 1–13, 8 set. 2021.

ZHANG, Sijie et al. BIM-Based Fall Hazard Identification and Prevention in Construction Safety Planning. **Safety Science**, v. 72, p. 31–45, fev. 2015.

ZHANG, Zhengyu et al. A General Channel Model for Integrated Sensing and Communication Scenarios. **IEEE Communications Magazine**, v. 61, n. 5, p. 68–74, maio 2023.

ZHANG, Zhenhao; LI, Wenbiao; YANG, Jianyu. ANALYSIS OF STOCHASTIC PROCESS TO MODEL SAFETY RISK IN CONSTRUCTION INDUSTRY. **JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT**, v. 27, n. 2, p. 87–99, 10 fev. 2021.

ZHENG, Esther Zhi Hong; CARVALHO, Marly Monteiro De. Managing Uncertainty in Projects: A Review, Trends and Gaps. **Revista de Gestão e Projetos**, v. 07, n. 02, p. 95–109, 1 ago. 2016.

ZHONG, DengHua et al. Robustness Analysis of Underground Powerhouse Construction Simulation Based on Markov Chain Monte Carlo Method. **Science China Technological Sciences**, v. 59, n. 2, p. 252–264, fev. 2016.

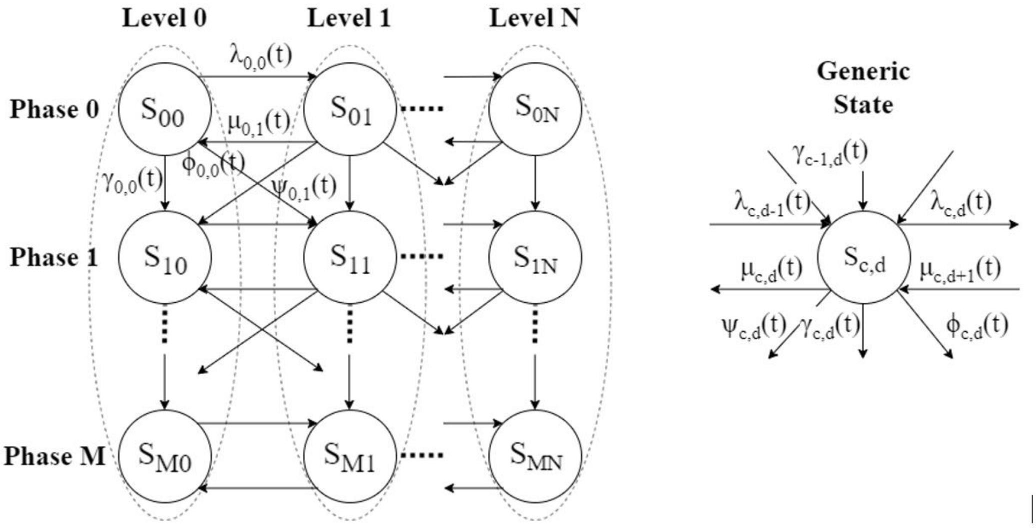
ZHOU, Zhipeng; GOH, Yang Miang; LI, Qiming. Overview and Analysis of Safety Management Studies in the Construction Industry. **Safety Science**, v. 72, p. 337–350, fev. 2015.

ZHOU, Zhipeng; IRIZARRY, Javier; LI, Qiming. Applying Advanced Technology to Improve Safety Management in the Construction Industry: A Literature Review. **Construction Management and Economics**, v. 31, n. 6, p. 606–622, jun. 2013.

ZOHREHVANDI, Shakib; KHALILZADEH, Mohammad. APRT-FMEA Buffer Sizing Method in Scheduling of a Wind Farm Construction Project. **Engineering, Construction and Architectural Management**, v. 26, n. 6, p. 1129–1150, 10 jul. 2019.

ZOLFAGHARIAN, Samaneh et al. Automated Safety Planning Approach for Residential Construction Sites in Malaysia. **International Journal of Construction Management**, v. 14, n. 3, p. 134–147, 3 jul. 2014.

# Appendix A
# QBDP-based solution considering delay events as levels.




Formally, the function map that defines the stochastic process is $f(s_i, d_i, t)$ and the possible states, $\boldsymbol{S}$(t), are:

$$\Omega = \boldsymbol{S}(t) = \{(s, d) \mid a = 0,1,2,3, \dots ; s = 0,1,2,3, \dots\}$$

It should be noted that the adopted QBDP assumes only states from the set of non-negative integer numbers, i.e., $\{s, d \in \mathbb{Z}^{+}\}$, as a matter of simplification.

Therefore, at time $t$ and given a time increment δ, the following transitions are possible:

$$\boldsymbol{S}(t+\delta) = \begin{cases} \boldsymbol{S}(t) \\ \boldsymbol{S}(t) + (0,1) \\ \boldsymbol{S}(t) + (0,-1) \\ \boldsymbol{S}(t) + (1,0) \\ \boldsymbol{S}(t) + (1,1) \\ \boldsymbol{S}(t) + (1,-1) \end{cases}$$

Similarly, as previously described, the transition probabilities $p_{(i,j),(m,n)}(dt)$, which denotes a transition from a state $(i,j)$ to a state $(m,n)$ in the infinitesimal time $dt$, are given by:

$$p_{(i,j),(m,n)}(dt) = \begin{cases} \lambda_{ij}(t)dt + o(dt), & if\ (m,n) = (i,j) + (0,1) \\ \mu_{ij}(t)dt + o(dt), & if\ (m,n) = (i,j) + (0,-1) \\ \gamma_{ij}(t)dt + o(dt), & if\ (m,n) = (i,j) + (1,0) \\ \phi_{ij}(t)dt + o(dt), & if\ (m,n) = (i,j) + (1,1) \\ \psi_{ij}(t)dt + o(dt), & if\ (m,n) = (i,j) + (1,-1) \\ \left[1 - \left(\lambda_{ij}(t) + \mu_{ij}(t) + \gamma_{ij}(t) + \phi_{ij}(t) + \psi_{ij}(t)\right)\right] dt + o(dt), & if\ (m,n) = (i,j) \\ 0, otherwise & \end{cases}$$

And the transition probabilities at time $t$ given an increment $dt$ are:

$$\begin{aligned} p_{(a,b),(c,d)}(t + dt) &= \lambda_{c,d-1}(t)p_{(a,b),(c,d-1)}(t)dt \\ &+ \mu_{c,d+1}(t)p_{(a,b),(c,d+1)}(t)dt \\ &+ \gamma_{c-1,d}(t)p_{(a,b),(c-1,d)}(t)dt \\ &+ \phi_{c-1,d-1}(t)p_{(a,b),(c-1,d-1)}(t)dt \\ &+ \psi_{c-1,d+1}(t)p_{(a,b),(c-1,d+1)}(t)dt \\ &+ [\ -\lambda_{c,d}(t) - \mu_{c,d}(t) - \gamma_{c,d}(t) - \phi_{c,d}(t) \\ &- \psi_{c,d}(t)]p_{(a,b),(c,d)}(t)dt + o(dt) \end{aligned}$$

As a result, the forward Kolmogorov equations can be obtained:

$$\begin{aligned} \frac{dp_{(a,b),(c,d)}(t)}{dt} &= \lambda_{c,d-1}(t)p_{(a,b),(c,d-1)}(t) \\ &+ \mu_{c,d+1}(t)p_{(a,b),(c,d+1)}(t) + \gamma_{c-1,d}(t)p_{(a,b),(c-1,d)}(t) \\ &+ \phi_{c-1,d-1}(t)p_{(a,b),(c-1,d-1)}(t) \\ &+ \psi_{c-1,d+1}(t)p_{(a,b),(c-1,d+1)}(t) \\ &- [\lambda_{c,d}(t) + \mu_{c,d}(t) + \gamma_{c,d}(t) + \phi_{c,d}(t) \\ &+ \psi_{c,d}(t)]p_{(a,b),(c,d)}(t) \end{aligned}$$

Rewriting the equations in matrix format, one obtains:

$$\frac{dp_{(a,b),(c,d)}(t)}{dt}$$
$$= \begin{bmatrix} p_{(a,b),(c-1,d-1)}(t) & p_{(a,b),(c-1,d)}(t) & p_{(a,b),(c-1,d+1)}(t) \\ p_{(a,b),(c,d-1)}(t) & p_{(a,b),(c,d)}(t) & p_{(a,b),(c,d+1)}(t) \end{bmatrix}$$
$$\times \begin{bmatrix} \phi_{c-1,d-1}(t) & \lambda_{c,d-1}(t) \\ \gamma_{c-1,d}(t) & -[\lambda_{c,d}(t) + \mu_{c,d}(t) + \gamma_{c,d}(t) + \phi_{c,d}(t) + \psi_{c,d}(t)] \\ \psi_{c-1,d+1}(t) & \mu_{c,d+1}(t) \end{bmatrix}$$

Due to the number of variables, some notations are simplified by dropping subscripts and birth and death rates are treated as constants, as follows: $p_{(a,b),(i,j)} = p_{(i,j)}$ ; $\lambda_{i,j}(t) = \lambda$ ; $\mu_{i,j}(t) = \mu$ ; $\gamma_{i,j}(t) = \gamma$ ; $\phi_{i,j}(t) = \phi$ ; $\psi_{i,j}(t) = \psi$.

The steady-state equations are calculated to obtain the generator matrix that describes the QBDP process, as described in Fadiloglu and Yeralan (2002). These equations define the balance conditions and represent the border states in the states diagram. For example, it is known that there are only two possible transitions to the first state $(0,0)$, coming from the state $(0,1)$ or from the same state $(0,0)$. As a result, the sum of both possible transitions must be zero.

$$-(\lambda + \gamma + \phi)\, p_{(0,0)} + \mu\, p_{(0,1)} = 0$$
$$-(\lambda + \gamma + \phi)\, p_{(i,0)} + \gamma\, p_{(i-1,0)} + \mu\, p_{(i,1)} + \psi p_{(i-1,1)} = 0$$
$$-\lambda\, p_{(M,0)} + \gamma\, p_{(M-1,0)} + \mu\, p_{(M,1)} + \psi p_{(M-1,1)} = 0$$
$$-(\lambda + \mu + \gamma + \phi + \psi)\, p_{(0,j)} + \lambda\, p_{(0,j-1)} + \mu\, p_{(0,j+1)} = 0$$
$$-(\lambda + \mu + \gamma + \phi + \psi)\, p_{(0,N)} + \lambda\, p_{(0,N-1)} = 0$$
$$-(\lambda + \mu + \gamma + \phi + \psi)\, p_{(i,j)} + \lambda\, p_{(i,j-1)} + \mu\, p_{(i,j+1)} + \phi\, p_{(i-1,j-1)} + \gamma\, p_{(i-1,j)} + \psi\, p_{(i-1,j+1)} = 0$$
$$-(\mu + \gamma + \psi)\, p_{(i,N)} + \lambda\, p_{(i,N-1)} + \phi\, p_{(i-1,N-1)} + \gamma\, p_{(i-1,N)} = 0$$
$$-(\lambda + \mu)\, p_{(M,j)} + \lambda\, p_{(M,j-1)} + \mu\, p_{(M,j+1)} + \phi\, p_{(M-1,j-1)} + \gamma\, p_{(M-1,j)} + \psi\, p_{(M-1,j+1)} = 0$$
$$-\mu\, p_{(M,N)} + \lambda\, p_{(M,N-1)} + \phi\, p_{(M-1,N-1)} + \gamma\, p_{(M-1,N)} = 0$$

According to Osogami (2005), the generator matrix $\boldsymbol{Q}$ has the following format, a block tridiagonal matrix of submatrices:

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{L}_{00} & \boldsymbol{F}_{01} & 0 & \\ \boldsymbol{B}_{10} & \boldsymbol{L}_{11} & \boldsymbol{F}_{12} & \cdots \\ 0 & \boldsymbol{B}_{21} & \boldsymbol{L}_{22} & \\ & \vdots & & \ddots \end{bmatrix}$$

Where,

$L_{ij}$ refers to the transition rate submatrix related to from state $(i, j)$ to state $(k, j)$ for $i \neq k$;

$F_{ij}$ refers to the transition rate submatrix related to from state $(i, j)$ to state $(k, j + 1)$ for any value of $i$ and $k$;

$B_{ij}$ refers to the transition rate submatrix related to from state $(i, j)$ to state $(k, j - 1)$ for any value of $i$ and $k$.

Thus, to solve the submatrices, matrix geometric procedures can be applied, according to Feldman et al. (1993). This results in the following submatrices:

$$L_{00} = \begin{bmatrix} -(\lambda+\gamma+\phi) & \gamma & \cdots & 0 \quad 0 & \\ 0 & -(\lambda+\gamma+\phi) & & 0 \quad 0 & \\ \vdots & & \ddots & \vdots & \\ 0 \quad 0 & & \cdots & -(\lambda+\gamma+\phi) & \gamma \\ 0 \quad 0 & & & 0 & -\lambda \end{bmatrix}_{N\times N}$$

$$F_{01} = \begin{bmatrix} \lambda & \phi & \cdots & 0 & 0 \\ 0 & \lambda & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \lambda & \phi \\ 0 & 0 & & 0 & \lambda \end{bmatrix}_{N\times N} \quad F_{12} = \begin{bmatrix} \lambda & \phi & \cdots & 0 & 0 \\ 0 & \lambda & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \lambda & \phi \\ 0 & 0 & & 0 & \lambda \end{bmatrix}_{N\times N}$$

$$B_{10} = \begin{bmatrix} \mu & \psi & \cdots & 0 & 0 \\ 0 & \mu & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \mu & \psi \\ 0 & 0 & & 0 & \mu \end{bmatrix}_{N\times N} \quad B_{21} = \begin{bmatrix} \mu & \psi & \cdots & 0 & 0 \\ 0 & \mu & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \mu & \psi \\ 0 & 0 & & 0 & \mu \end{bmatrix}_{N\times N}$$

$$L_{11} = \begin{bmatrix} -(\lambda+\gamma+\phi+\mu+\psi) & \gamma & \cdots & 0 \quad 0 & \\ 0 & -(\lambda+\gamma+\phi+\mu+\psi) & & 0 \quad 0 & \\ \vdots & & \ddots & \vdots & \\ 0 \quad 0 & & \cdots & -(\lambda+\gamma+\phi+\mu+\psi) & \gamma \\ 0 \quad 0 & & & 0 & -(\lambda+\mu) \end{bmatrix}_{N\times N}$$

$$L_{NN} = \begin{bmatrix} -(\gamma+\mu+\psi) & \gamma & \cdots & 0 \quad 0 & \\ 0 & -(\gamma+\mu+\psi) & & 0 \quad 0 & \\ \vdots & & \ddots & \vdots & \\ 0 \quad 0 & & \cdots & -(\gamma+\mu+\psi) & \gamma \\ 0 \quad 0 & & & 0 & -\mu \end{bmatrix}_{N\times N}$$

# Appendix B
# Experiment plan related to the bivariate time series forecasting

| # | Training Project | Walk forward validation % Train* | Trainig data size (instances, attributes) | VARIMA parameters p | VARIMA parameters q | Neural networks Block sequences | VAR (Lag = 1) results MAE | VAR (Lag = 1) results MAPE | VAR (Lag = 1) results RMSE | VARIMA results MAE | VARIMA results MAPE | VARIMA results RMSE | CNN results MAE | CNN results MAPE | CNN results RMSE | LSTM results MAE | LSTM results MAPE | LSTM results RMSE | **Global CNN results** MAE | **Global CNN results** MAPE | **Global CNN results** RMSE | **Global LSTM results** MAE | **Global LSTM results** MAPE | **Global LSTM results** RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **...** | **...2** | **...3** | **...5** | **...6** | **...7** | **...8** | **...11** | **...12** | ...13 | ...14 | ...15 | ...16 | ...17 | ...18 | ...19 | ...20 | ...21 | ...22 | **...23** | **...24** | **...25** | **...26** | **...27** | **...28** |
| 1 | T238 | 0,3 | (13, 2) | 1 | 1 | - | 5,73 | 19,7% | 6,44 | 5,70 | 18,2% | 6,41 | - | - | - | - | - | - | - | - | - | - | - | - |
| 2 | T238 | 0,3 | (13, 2) | 1 | 2 | - | 5,73 | 19,7% | 6,44 | 5,85 | 20,6% | 6,54 | - | - | - | - | - | - | - | - | - | - | - | - |
| 3 | T238 | 0,3 | (13, 2) | 2 | 1 | - | 5,73 | 0,20 | 6,44 | 5,53 | 0,05 | 6,09 | - | - | - | - | - | - | - | - | - | - | - | - |
| 4 | T238 | 0,3 | (13, 2) | 2 | 2 | - | 5,73 | 19,7% | 6,44 | 6,11 | 29,7% | 6,62 | - | - | - | - | - | - | - | - | - | - | - | - |
| 5 | T238 | 0,5 | (23, 2) | 1 | 1 | - | 2,65 | 2,2% | 3,03 | 4,54 | 19,8% | 5,14 | - | - | - | - | - | - | - | - | - | - | - | - |
| 6 | T238 | 0,5 | (23, 2) | 1 | 2 | - | 2,65 | 2,2% | 3,03 | 4,61 | 20,9% | 5,19 | - | - | - | - | - | - | - | - | - | - | - | - |
| 7 | T238 | 0,5 | (23, 2) | 2 | 1 | - | 2,65 | 2,2% | 3,03 | 4,35 | 22,8% | 5,06 | - | - | - | - | - | - | - | - | - | - | - | - |
| 8 | T238 | 0,5 | (23, 2) | 2 | 2 | - | 2,65 | 2,2% | 3,03 | 4,59 | 24,1% | 5,21 | - | - | - | - | - | - | - | - | - | - | - | - |
| 9 | T238 | 0,7 | (32, 2) | 1 | 1 | - | 0,36 | 0,3% | 0,60 | 1,89 | 19,1% | 2,26 | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | T238 | 0,7 | (32, 2) | 1 | 2 | - | 0,36 | 0,3% | 0,60 | 2,43 | 21,3% | 2,84 | - | - | - | - | - | - | - | - | - | - | - | - |
| 11 | T238 | 0,7 | (32, 2) | 2 | 1 | - | 0,36 | 0,3% | 0,60 | 0,79 | 11,2% | 1,01 | - | - | - | - | - | - | - | - | - | - | - | - |
| 12 | T238 | 0,7 | (32, 2) | 2 | 2 | - | 0,36 | 0,00 | 0,60 | 0,68 | 0,11 | 0,88 | - | - | - | - | - | - | - | - | - | - | - | - |
| 13 | S843 | 0,3 | (8, 2) | 1 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 14 | S843 | 0,3 | (8, 2) | 1 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 15 | S843 | 0,3 | (8, 2) | 2 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 16 | S843 | 0,3 | (8, 2) | 2 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 17 | S843 | 0,5 | (14, 2) | 1 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 18 | S843 | 0,5 | (14, 2) | 1 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 19 | S843 | 0,5 | (14, 2) | 2 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 20 | S843 | 0,5 | (14, 2) | 2 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 21 | S843 | 0,7 | (20, 2) | 1 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 22 | S843 | 0,7 | (20, 2) | 1 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 23 | S843 | 0,7 | (20, 2) | 2 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 24 | S843 | 0,7 | (20, 2) | 2 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 25 | C434 | 0,3 | (37, 2) | 1 | 1 | - | 7,33 | 9,2% | 9,10 | 18,19 | 28,0% | 22,09 | - | - | - | - | - | - | - | - | - | - | - | - |
| 26 | C434 | 0,3 | (37, 2) | 1 | 2 | - | 7,33 | 9,2% | 9,10 | 18,55 | 28,4% | 22,43 | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | C434 | 0,3 | (37, 2) | 2 | 1 | - | 7,33 | 9,2% | 9,10 | 17,56 | 26,7% | 21,11 | - | - | - | - | - | - | - | - | - | - | - | - |
| 28 | C434 | 0,3 | (37, 2) | 2 | 2 | - | 7,33 | 9,2% | 9,10 | 16,58 | 25,2% | 20,14 | - | - | - | - | - | - | - | - | - | - | - | - |
| 29 | C434 | 0,5 | (62, 2) | 1 | 1 | - | 12,00 | 17,0% | 13,80 | 19,45 | 27,8% | 22,49 | - | - | - | - | - | - | - | - | - | - | - | - |
| 30 | C434 | 0,5 | (62, 2) | 1 | 2 | - | 12,00 | 17,0% | 13,80 | 18,93 | 26,9% | 22,07 | - | - | - | - | - | - | - | - | - | - | - | - |
| 31 | C434 | 0,5 | (62, 2) | 2 | 1 | - | 12,00 | 17,0% | 13,80 | 20,71 | 29,4% | 23,81 | - | - | - | - | - | - | - | - | - | - | - | - |
| 32 | C434 | 0,5 | (62, 2) | 2 | 2 | - | 12,00 | 17,0% | 13,80 | 19,47 | 27,6% | 22,41 | - | - | - | - | - | - | - | - | - | - | - | - |
| 33 | C434 | 0,7 | (86, 2) | 1 | 1 | - | 6,03 | 6,8% | 7,23 | 16,12 | 20,2% | 18,48 | - | - | - | - | - | - | - | - | - | - | - | - |
| 34 | C434 | 0,7 | (86, 2) | 1 | 2 | - | 6,03 | 6,8% | 7,23 | 14,72 | 18,4% | 17,14 | - | - | - | - | - | - | - | - | - | - | - | - |
| 35 | C434 | 0,7 | (86, 2) | 2 | 1 | - | 6,03 | 6,8% | 7,23 | 11,11 | 13,5% | 13,13 | - | - | - | - | - | - | - | - | - | - | - | - |
| 36 | C434 | 0,7 | (86, 2) | 2 | 2 | - | 6,03 | 6,8% | 7,23 | 14,41 | 18,1% | 16,57 | - | - | - | - | - | - | - | - | - | - | - | - |
| 37 | V104 | 0,3 | (45, 2) | 1 | 1 | - | 39,78 | 43,9% | 48,66 | 18,63 | 31,8% | 22,25 | - | - | - | - | - | - | - | - | - | - | - | - |
| 38 | V104 | 0,3 | (45, 2) | 1 | 2 | - | 39,78 | 43,9% | 48,66 | 18,73 | 32,8% | 22,21 | - | - | - | - | - | - | - | - | - | - | - | - |
| 39 | V104 | 0,3 | (45, 2) | 2 | 1 | - | 39,78 | 43,9% | 48,66 | 18,90 | 33,0% | 22,31 | - | - | - | - | - | - | - | - | - | - | - | - |
| 40 | V104 | 0,3 | (45, 2) | 2 | 2 | - | 39,78 | 43,9% | 48,66 | 18,71 | 32,5% | 22,21 | - | - | - | - | - | - | - | - | - | - | - | - |
| 41 | V104 | 0,5 | (75, 2) | 1 | 1 | - | 15,99 | 17,4% | 21,62 | 21,25 | 31,4% | 23,71 | - | - | - | - | - | - | - | - | - | - | - | - |
| 42 | V104 | 0,5 | (75, 2) | 1 | 2 | - | 15,99 | 17,4% | 21,62 | 22,13 | 33,7% | 24,45 | - | - | - | - | - | - | - | - | - | - | - | - |
| 43 | V104 | 0,5 | (75, 2) | 2 | 1 | - | 15,99 | 17,4% | 21,62 | 22,47 | 34,3% | 24,57 | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 44 | V104 | 0,5 | (75, 2) | 2 | 2 | - | 15,99 | 17,4% | 21,62 | 21,53 | 32,5% | 23,96 | - | - | - | - | - | - | - | - | - | - | - | - |
| 45 | V104 | 0,7 | (105, 2) | 1 | 1 | - | 6,57 | 7,5% | 8,98 | 13,60 | 18,0% | 14,82 | - | - | - | - | - | - | - | - | - | - | - | - |
| 46 | V104 | 0,7 | (105, 2) | 1 | 2 | - | 6,57 | 7,5% | 8,98 | 14,26 | 19,0% | 15,61 | - | - | - | - | - | - | - | - | - | - | - | - |
| 47 | V104 | 0,7 | (105, 2) | 2 | 1 | - | 6,57 | 7,5% | 8,98 | 16,16 | 22,0% | 17,73 | - | - | - | - | - | - | - | - | - | - | - | - |
| 48 | V104 | 0,7 | (105, 2) | 2 | 2 | - | 6,57 | 7,5% | 8,98 | 16,88 | 22,9% | 18,54 | - | - | - | - | - | - | - | - | - | - | - | - |
| 49 | A172 | 0,3 | (16, 2) | 1 | 1 | - | 11,40 | 29,4% | 13,86 | 9,03 | 38,7% | 10,20 | - | - | - | - | - | - | - | - | - | - | - | - |
| 50 | A172 | 0,3 | (16, 2) | 1 | 2 | - | 11,40 | 29,4% | 13,86 | 9,10 | 39,1% | 10,25 | - | - | - | - | - | - | - | - | - | - | - | - |
| 51 | A172 | 0,3 | (16, 2) | 2 | 1 | - | 11,40 | 29,4% | 13,86 | 9,66 | 40,5% | 10,79 | - | - | - | - | - | - | - | - | - | - | - | - |
| 52 | A172 | 0,3 | (16, 2) | 2 | 2 | - | 11,40 | 29,4% | 13,86 | 9,98 | 41,5% | 11,06 | - | - | - | - | - | - | - | - | - | - | - | - |
| 53 | A172 | 0,5 | (28, 2) | 1 | 1 | - | 2,82 | 7,1% | 3,15 | 8,34 | 26,4% | 9,70 | - | - | - | - | - | - | - | - | - | - | - | - |
| 54 | A172 | 0,5 | (28, 2) | 1 | 2 | - | 2,82 | 7,1% | 3,15 | 7,05 | 21,1% | 8,43 | - | - | - | - | - | - | - | - | - | - | - | - |
| 55 | A172 | 0,5 | (28, 2) | 2 | 1 | - | 2,82 | 7,1% | 3,15 | 5,63 | 23,4% | 6,28 | - | - | - | - | - | - | - | - | - | - | - | - |
| 56 | A172 | 0,5 | (28, 2) | 2 | 2 | - | 2,82 | 7,1% | 3,15 | 7,30 | 29,9% | 8,16 | - | - | - | - | - | - | - | - | - | - | - | - |
| 57 | A172 | 0,7 | (39, 2) | 1 | 1 | - | 1,03 | 2,2% | 1,34 | 3,62 | 7,1% | 4,31 | - | - | - | - | - | - | - | - | - | - | - | - |
| 58 | A172 | 0,7 | (39, 2) | 1 | 2 | - | 1,03 | 2,2% | 1,34 | 2,56 | 5,3% | 3,08 | - | - | - | - | - | - | - | - | - | - | - | - |
| 59 | A172 | 0,7 | (39, 2) | 2 | 1 | - | 1,03 | 2,2% | 1,34 | 4,21 | 15,1% | 4,87 | - | - | - | - | - | - | - | - | - | - | - | - |
| 60 | A172 | 0,7 | (39, 2) | 2 | 2 | - | 1,03 | 2,2% | 1,34 | 4,41 | 15,4% | 5,06 | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 61 | M941 | 0,3 | (30, 2) | 1 | 1 | - | 40,08 | 68,2% | 53,92 | 11,52 | 32,1% | 13,13 | - | - | - | - | - | - | - | - | - | - | - | - |
| 62 | M941 | 0,3 | (30, 2) | 1 | 2 | - | 40,08 | 68,2% | 53,92 | 11,63 | 33,0% | 13,26 | - | - | - | - | - | - | - | - | - | - | - | - |
| 63 | M941 | 0,3 | (30, 2) | 2 | 1 | - | 40,08 | 68,2% | 53,92 | 11,16 | 29,5% | 13,26 | - | - | - | - | - | - | - | - | - | - | - | - |
| 64 | M941 | 0,3 | (30, 2) | 2 | 2 | - | 40,08 | 68,2% | 53,92 | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 65 | M941 | 0,5 | (50, 2) | 1 | 1 | - | 10,32 | 16,4% | 11,60 | 9,19 | 21,6% | 10,53 | - | - | - | - | - | - | - | - | - | - | - | - |
| 66 | M941 | 0,5 | (50, 2) | 1 | 2 | - | 10,32 | 16,4% | 11,60 | 8,77 | 22,6% | 10,31 | - | - | - | - | - | - | - | - | - | - | - | - |
| 67 | M941 | 0,5 | (50, 2) | 2 | 1 | - | 10,32 | 16,4% | 11,60 | 12,73 | 23,2% | 13,87 | - | - | - | - | - | - | - | - | - | - | - | - |
| 68 | M941 | 0,5 | (50, 2) | 2 | 2 | - | 10,32 | 16,4% | 11,60 | 13,31 | 25,5% | 14,40 | - | - | - | - | - | - | - | - | - | - | - | - |
| 69 | M941 | 0,7 | (70, 2) | 1 | 1 | - | 4,15 | 5,2% | 5,06 | 12,23 | 23,3% | 13,09 | - | - | - | - | - | - | - | - | - | - | - | - |
| 70 | M941 | 0,7 | (70, 2) | 1 | 2 | - | 4,15 | 5,2% | 5,06 | 11,15 | 19,4% | 12,05 | - | - | - | - | - | - | - | - | - | - | - | - |
| 71 | M941 | 0,7 | (70, 2) | 2 | 1 | - | 4,15 | 5,2% | 5,06 | 12,35 | 21,0% | 13,06 | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 72 | M941 | 0,7 | (70, 2) | 2 | 2 | - | 4,15 | 5,2% | 5,06 | 11,45 | 22,3% | 12,30 | - | - | - | - | - | - | - | - | - | - | - | - |
| 73 | L406 | 0,3 | (23, 2) | 1 | 1 | - | 8,91 | 27,3% | 10,51 | 12,74 | 24,8% | 14,76 | - | - | - | - | - | - | - | - | - | - | - | - |
| 74 | L406 | 0,3 | (23, 2) | 1 | 2 | - | 8,91 | 27,3% | 10,51 | 12,06 | 23,7% | 13,95 | - | - | - | - | - | - | - | - | - | - | - | - |
| 75 | L406 | 0,3 | (23, 2) | 2 | 1 | - | 8,91 | 27,3% | 10,51 | 16,84 | 38,2% | 18,61 | - | - | - | - | - | - | - | - | - | - | - | - |
| 76 | L406 | 0,3 | (23, 2) | 2 | 2 | - | 8,91 | 27,3% | 10,51 | 16,47 | 37,1% | 18,27 | - | - | - | - | - | - | - | - | - | - | - | - |
| 77 | L406 | 0,5 | (38, 2) | 1 | 1 | - | 3,58 | 7,6% | 4,13 | 11,65 | 24,7% | 13,31 | - | - | - | - | - | - | - | - | - | - | - | - |
| 78 | L406 | 0,5 | (38, 2) | 1 | 2 | - | 3,58 | 7,6% | 4,13 | 11,23 | 24,1% | 12,87 | - | - | - | - | - | - | - | - | - | - | - | - |
| 79 | L406 | 0,5 | (38, 2) | 2 | 1 | - | 3,58 | 7,6% | 4,13 | 12,97 | 30,6% | 15,54 | - | - | - | - | - | - | - | - | - | - | - | - |
| 80 | L406 | 0,5 | (38, 2) | 2 | 2 | - | 3,58 | 7,6% | 4,13 | 13,47 | 31,6% | 16,08 | - | - | - | - | - | - | - | - | - | - | - | - |
| 81 | L406 | 0,7 | (53, 2) | 1 | 1 | - | 1,06 | 2,9% | 1,35 | 8,92 | 15,5% | 9,96 | - | - | - | - | - | - | - | - | - | - | - | - |
| 82 | L406 | 0,7 | (53, 2) | 1 | 2 | - | 1,06 | 2,9% | 1,35 | 9,69 | 16,9% | 10,93 | - | - | - | - | - | - | - | - | - | - | - | - |
| 83 | L406 | 0,7 | (53, 2) | 2 | 1 | - | 1,06 | 2,9% | 1,35 | 3,48 | 6,3% | 4,27 | - | - | - | - | - | - | - | - | - | - | - | - |
| 84 | L406 | 0,7 | (53, 2) | 2 | 2 | - | 1,06 | 2,9% | 1,35 | 2,90 | 4,9% | 3,71 | - | - | - | - | - | - | - | - | - | - | - | - |
| 85 | C907 | 0,3 | (17, 2) | 1 | 1 | - | 20,77 | 83,2% | 29,20 | 9,65 | 31,8% | 11,43 | - | - | - | - | - | - | - | - | - | - | - | - |
| 86 | C907 | 0,3 | (17, 2) | 1 | 2 | - | 20,77 | 83,2% | 29,20 | 9,60 | 31,9% | 11,33 | - | - | - | - | - | - | - | - | - | - | - | - |
| 87 | C907 | 0,3 | (17, 2) | 2 | 1 | - | 20,77 | 83,2% | 29,20 | 10,74 | 36,0% | 12,31 | - | - | - | - | - | - | - | - | - | - | - | - |
| 88 | C907 | 0,3 | (17, 2) | 2 | 2 | - | 20,77 | 83,2% | 29,20 | 10,32 | 33,7% | 11,95 | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 89 | C907 | 0,5 | (29, 2) | 1 | 1 | - | 8,59 | 27,4% | 10,52 | 10,02 | 27,8% | 10,98 | - | - | - | - | - | - | - | - | - | - | - | - |
| 90 | C907 | 0,5 | (29, 2) | 1 | 2 | - | 8,59 | 27,4% | 10,52 | 9,81 | 26,9% | 10,88 | - | - | - | - | - | - | - | - | - | - | - | - |
| 91 | C907 | 0,5 | (29, 2) | 2 | 1 | - | 8,59 | 27,4% | 10,52 | 6,71 | 12,8% | 7,58 | - | - | - | - | - | - | - | - | - | - | - | - |
| 92 | C907 | 0,5 | (29, 2) | 2 | 2 | - | 8,59 | 27,4% | 10,52 | 6,50 | 10,8% | 7,45 | - | - | - | - | - | - | - | - | - | - | - | - |
| 93 | C907 | 0,7 | (40, 2) | 1 | 1 | - | 4,17 | 11,9% | 6,01 | 8,72 | 18,0% | 9,22 | - | - | - | - | - | - | - | - | - | - | - | - |
| 94 | C907 | 0,7 | (40, 2) | 1 | 2 | - | 4,17 | 11,9% | 6,01 | 8,58 | 18,0% | 9,07 | - | - | - | - | - | - | - | - | - | - | - | - |
| 95 | C907 | 0,7 | (40, 2) | 2 | 1 | - | 4,17 | 11,9% | 6,01 | 1,33 | 2,8% | 1,79 | - | - | - | - | - | - | - | - | - | - | - | - |
| 96 | C907 | 0,7 | (40, 2) | 2 | 2 | - | 4,17 | 11,9% | 6,01 | 1,47 | 3,7% | 2,04 | - | - | - | - | - | - | - | - | - | - | - | - |
| 97 | M174 | 0,3 | (24, 2) | 1 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 98 | M174 | 0,3 | (24, 2) | 1 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 99 | M174 | 0,3 | (24, 2) | 2 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 100 | M174 | 0,3 | (24, 2) | 2 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 101 | M174 | 0,5 | (41, 2) | 1 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 102 | M174 | 0,5 | (41, 2) | 1 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 103 | M174 | 0,5 | (41, 2) | 2 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 104 | M174 | 0,5 | (41, 2) | 2 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 105 | M174 | 0,7 | (58, 2) | 1 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 106 | M174 | 0,7 | (58, 2) | 1 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 107 | M174 | 0,7 | (58, 2) | 2 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 108 | M174 | 0,7 | (58, 2) | 2 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 109 | S242 | 0,3 | (48, 2) | 1 | 1 | - | 8,69 | 16,4% | 10,38 | 18,46 | 37,5% | 20,87 | - | - | - | - | - | - | - | - | - | - | - | - |
| 110 | S242 | 0,3 | (48, 2) | 1 | 2 | - | 8,69 | 16,4% | 10,38 | 19,17 | 38,6% | 21,51 | - | - | - | - | - | - | - | - | - | - | - | - |
| 111 | S242 | 0,3 | (48, 2) | 2 | 1 | - | 8,69 | 16,4% | 10,38 | 15,40 | 31,5% | 18,20 | - | - | - | - | - | - | - | - | - | - | - | - |
| 112 | S242 | 0,3 | (48, 2) | 2 | 2 | - | 8,69 | 16,4% | 10,38 | 14,72 | 31,0% | 17,65 | - | - | - | - | - | - | - | - | - | - | - | - |
| 113 | S242 | 0,5 | (80, 2) | 1 | 1 | - | 11,50 | 20,5% | 14,02 | 18,40 | 33,4% | 20,29 | - | - | - | - | - | - | - | - | - | - | - | - |
| 114 | S242 | 0,5 | (80, 2) | 1 | 2 | - | 11,50 | 20,5% | 14,02 | 18,57 | 33,7% | 20,58 | - | - | - | - | - | - | - | - | - | - | - | - |
| 115 | S242 | 0,5 | (80, 2) | 2 | 1 | - | 11,50 | 20,5% | 14,02 | 6,28 | 11,5% | 8,55 | - | - | - | - | - | - | - | - | - | - | - | - |
| 116 | S242 | 0,5 | (80, 2) | 2 | 2 | - | 11,50 | 20,5% | 14,02 | 6,58 | 12,0% | 8,75 | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 117 | S242 | 0,7 | (112, 2) | 1 | 1 | - | 10,03 | 16,2% | 11,34 | 13,80 | 23,3% | 15,70 | - | - | - | - | - | - | - | - | - | - | - | - |
| 118 | S242 | 0,7 | (112, 2) | 1 | 2 | - | 10,03 | 16,2% | 11,34 | 13,57 | 23,3% | 15,46 | - | - | - | - | - | - | - | - | - | - | - | - |
| 119 | S242 | 0,7 | (112, 2) | 2 | 1 | - | 10,03 | 16,2% | 11,34 | 12,85 | 21,5% | 14,66 | - | - | - | - | - | - | - | - | - | - | - | - |
| 120 | S242 | 0,7 | (112, 2) | 2 | 2 | - | 10,03 | 16,2% | 11,34 | 12,66 | 21,1% | 14,46 | - | - | - | - | - | - | - | - | - | - | - | - |
| 121 | L868 | 0,3 | (30, 2) | 1 | 1 | - | 4,97 | 0,17 | 5,85 | 15,96 | 0,43 | 18,02 | - | - | - | - | - | - | - | - | - | - | - | - |
| 122 | L868 | 0,3 | (30, 2) | 1 | 2 | - | 4,97 | 0,17 | 5,85 | 15,74 | 0,42 | 17,85 | - | - | - | - | - | - | - | - | - | - | - | - |
| 123 | L868 | 0,3 | (30, 2) | 2 | 1 | - | 4,97 | 0,17 | 5,85 | 17,86 | 0,48 | 19,72 | - | - | - | - | - | - | - | - | - | - | - | - |
| 124 | L868 | 0,3 | (30, 2) | 2 | 2 | - | 4,97 | 0,17 | 5,85 | 16,23 | 0,44 | 18,22 | - | - | - | - | - | - | - | - | - | - | - | - |
| 125 | L868 | 0,5 | (50, 2) | 1 | 1 | - | 5,13 | 15,0% | 5,89 | 10,21 | 30,1% | 11,35 | - | - | - | - | - | - | - | - | - | - | - | - |
| 126 | L868 | 0,5 | (50, 2) | 1 | 2 | - | 5,13 | 15,0% | 5,89 | 10,25 | 29,5% | 11,41 | - | - | - | - | - | - | - | - | - | - | - | - |
| 127 | L868 | 0,5 | (50, 2) | 2 | 1 | - | 5,13 | 15,0% | 5,89 | 10,21 | 27,5% | 11,76 | - | - | - | - | - | - | - | - | - | - | - | - |
| 128 | L868 | 0,5 | (50, 2) | 2 | 2 | - | 5,13 | 15,0% | 5,89 | 10,96 | 28,9% | 12,81 | - | - | - | - | - | - | - | - | - | - | - | - |
| 129 | L868 | 0,7 | (70, 2) | 1 | 1 | - | 4,56 | 14,7% | 5,93 | 3,71 | 7,0% | 4,14 | - | - | - | - | - | - | - | - | - | - | - | - |
| 130 | L868 | 0,7 | (70, 2) | 1 | 2 | - | 4,56 | 14,7% | 5,93 | 5,03 | 9,7% | 5,72 | - | - | - | - | - | - | - | - | - | - | - | - |
| 131 | L868 | 0,7 | (70, 2) | 2 | 1 | - | 4,56 | 14,7% | 5,93 | 2,98 | 9,0% | 3,18 | - | - | - | - | - | - | - | - | - | - | - | - |
| 132 | L868 | 0,7 | (70, 2) | 2 | 2 | - | 4,56 | 14,7% | 5,93 | 3,69 | 10,9% | 4,08 | - | - | - | - | - | - | - | - | - | - | - | - |
| 133 | A739 | 0,3 | (51, 2) | 1 | 1 | - | 31,76 | 53,0% | 40,18 | 27,23 | 38,0% | 29,70 | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 134 | A739 | 0,3 | (51, 2) | 1 | 2 | - | 31,76 | 53,0% | 40,18 | 28,57 | 42,4% | 31,07 | - | - | - | - | - | - | - | - | - | - | - | - |
| 135 | A739 | 0,3 | (51, 2) | 2 | 1 | - | 31,76 | 53,0% | 40,18 | 27,78 | 40,0% | 30,09 | - | - | - | - | - | - | - | - | - | - | - | - |
| 136 | A739 | 0,3 | (51, 2) | 2 | 2 | - | 31,76 | 53,0% | 40,18 | 28,21 | 41,1% | 30,59 | - | - | - | - | - | - | - | - | - | - | - | - |
| 137 | A739 | 0,5 | (85, 2) | 1 | 1 | - | 5,61 | 9,2% | 6,79 | 19,97 | 33,7% | 23,65 | - | - | - | - | - | - | - | - | - | - | - | - |
| 138 | A739 | 0,5 | (85, 2) | 1 | 2 | - | 5,61 | 9,2% | 6,79 | 17,41 | 30,5% | 20,77 | - | - | - | - | - | - | - | - | - | - | - | - |
| 139 | A739 | 0,5 | (85, 2) | 2 | 1 | - | 5,61 | 9,2% | 6,79 | 7,20 | 16,4% | 8,34 | - | - | - | - | - | - | - | - | - | - | - | - |
| 140 | A739 | 0,5 | (85, 2) | 2 | 2 | - | 5,61 | 9,2% | 6,79 | 9,99 | 21,2% | 12,33 | - | - | - | - | - | - | - | - | - | - | - | - |
| 141 | A739 | 0,7 | (118, 2) | 1 | 1 | - | 0,71 | 1,0% | 0,96 | 8,70 | 15,0% | 10,53 | - | - | - | - | - | - | - | - | - | - | - | - |
| 142 | A739 | 0,7 | (118, 2) | 1 | 2 | - | 0,71 | 1,0% | 0,96 | 8,19 | 15,4% | 9,85 | - | - | - | - | - | - | - | - | - | - | - | - |
| 143 | A739 | 0,7 | (118, 2) | 2 | 1 | - | 0,71 | 1,0% | 0,96 | 3,55 | 7,6% | 4,35 | - | - | - | - | - | - | - | - | - | - | - | - |
| 144 | A739 | 0,7 | (118, 2) | 2 | 2 | - | 0,71 | 1,0% | 0,96 | 3,10 | 7,2% | 3,83 | - | - | - | - | - | - | - | - | - | - | - | - |
| 145 | C624 | 0,3 | (30, 2) | 1 | 1 | - | 7,23 | 20,4% | 8,85 | 10,07 | 36,3% | 11,70 | - | - | - | - | - | - | - | - | - | - | - | - |
| 146 | C624 | 0,3 | (30, 2) | 1 | 2 | - | 7,23 | 20,4% | 8,85 | 10,03 | 36,3% | 11,68 | - | - | - | - | - | - | - | - | - | - | - | - |
| 147 | C624 | 0,3 | (30, 2) | 2 | 1 | - | 7,23 | 20,4% | 8,85 | 9,68 | 35,6% | 11,29 | - | - | - | - | - | - | - | - | - | - | - | - |
| 148 | C624 | 0,3 | (30, 2) | 2 | 2 | - | 7,23 | 20,4% | 8,85 | 9,54 | 34,5% | 11,21 | - | - | - | - | - | - | - | - | - | - | - | - |
| 149 | C624 | 0,5 | (50, 2) | 1 | 1 | - | 6,57 | 14,3% | 7,55 | 9,27 | 19,5% | 10,56 | - | - | - | - | - | - | - | - | - | - | - | - |
| 150 | C624 | 0,5 | (50, 2) | 1 | 2 | - | 6,57 | 14,3% | 7,55 | 9,46 | 18,9% | 10,82 | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 151 | C624 | 0,5 | (50, 2) | 2 | 1 | - | 6,57 | 14,3% | 7,55 | 9,75 | 24,8% | 10,92 | - | - | - | - | - | - | - | - | - | - | - | - |
| 152 | C624 | 0,5 | (50, 2) | 2 | 2 | - | 6,57 | 14,3% | 7,55 | 8,85 | 25,0% | 10,12 | - | - | - | - | - | - | - | - | - | - | - | - |
| 153 | C624 | 0,7 | (70, 2) | 1 | 1 | - | 1,19 | 2,4% | 1,56 | 10,26 | 22,5% | 10,92 | - | - | - | - | - | - | - | - | - | - | - | - |
| 154 | C624 | 0,7 | (70, 2) | 1 | 2 | - | 1,19 | 2,4% | 1,56 | 10,89 | 25,2% | 11,46 | - | - | - | - | - | - | - | - | - | - | - | - |
| 155 | C624 | 0,7 | (70, 2) | 2 | 1 | - | 1,19 | 2,4% | 1,56 | 9,47 | 19,0% | 10,08 | - | - | - | - | - | - | - | - | - | - | - | - |
| 156 | C624 | 0,7 | (70, 2) | 2 | 2 | - | 1,19 | 2,4% | 1,56 | 9,73 | 21,2% | 10,41 | - | - | - | - | - | - | - | - | - | - | - | - |
| 157 | A676 | 0,3 | (1, 2) | 1 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 158 | A676 | 0,3 | (1, 2) | 1 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 159 | A676 | 0,3 | (1, 2) | 2 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 160 | A676 | 0,3 | (1, 2) | 2 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 161 | A676 | 0,5 | (2, 2) | 1 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 162 | A676 | 0,5 | (2, 2) | 1 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 163 | A676 | 0,5 | (2, 2) | 2 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 164 | A676 | 0,5 | (2, 2) | 2 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 165 | A676 | 0,7 | (3, 2) | 1 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 166 | A676 | 0,7 | (3, 2) | 1 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 167 | A676 | 0,7 | (3, 2) | 2 | 1 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 168 | A676 | 0,7 | (3, 2) | 2 | 2 | - | Error | Error | Error | Error | Error | Error | - | - | - | - | - | - | - | - | - | - | - | - |
| 169 | S165 | 0,3 | (16, 2) | 1 | 1 | - | 6,84 | 32,6% | 7,79 | 6,75 | 35,3% | 7,73 | - | - | - | - | - | - | - | - | - | - | - | - |
| 170 | S165 | 0,3 | (16, 2) | 1 | 2 | - | 6,84 | 32,6% | 7,79 | 6,86 | 37,0% | 7,81 | - | - | - | - | - | - | - | - | - | - | - | - |
| 171 | S165 | 0,3 | (16, 2) | 2 | 1 | - | 6,84 | 32,6% | 7,79 | 6,96 | 37,6% | 7,92 | - | - | - | - | - | - | - | - | - | - | - | - |
| 172 | S165 | 0,3 | (16, 2) | 2 | 2 | - | 6,84 | 32,6% | 7,79 | 6,72 | 36,6% | 7,74 | - | - | - | - | - | - | - | - | - | - | - | - |
| 173 | S165 | 0,5 | (27, 2) | 1 | 1 | - | 1,76 | 9,3% | 2,19 | 3,37 | 22,2% | 4,45 | - | - | - | - | - | - | - | - | - | - | - | - |
| 174 | S165 | 0,5 | (27, 2) | 1 | 2 | - | 1,76 | 9,3% | 2,19 | 2,74 | 19,4% | 3,57 | - | - | - | - | - | - | - | - | - | - | - | - |
| 175 | S165 | 0,5 | (27, 2) | 2 | 1 | - | 1,76 | 9,3% | 2,19 | 3,26 | 22,4% | 4,37 | - | - | - | - | - | - | - | - | - | - | - | - |
| 176 | S165 | 0,5 | (27, 2) | 2 | 2 | - | 1,76 | 0,09 | 2,19 | 1,89 | 0,17 | 2,42 | - | - | - | - | - | - | - | - | - | - | - | - |
| 177 | S165 | 0,7 | (37, 2) | 1 | 1 | - | 1,06 | 9,1% | 1,28 | 1,76 | 13,4% | 2,13 | - | - | - | - | - | - | - | - | - | - | - | - |
| 178 | S165 | 0,7 | (37, 2) | 1 | 2 | - | 1,06 | 9,1% | 1,28 | 2,09 | 13,7% | 2,50 | - | - | - | - | - | - | - | - | - | - | - | - |
| 179 | S165 | 0,7 | (37, 2) | 2 | 1 | - | 1,06 | 9,1% | 1,28 | 1,88 | 13,6% | 2,37 | - | - | - | - | - | - | - | - | - | - | - | - |
| 180 | S165 | 0,7 | (37, 2) | 2 | 2 | - | 1,06 | 9,1% | 1,28 | 1,71 | 12,6% | 2,10 | - | - | - | - | - | - | - | - | - | - | - | - |
| 181 | D316 | 0,3 | (47, 2) | 1 | 1 | - | 27,10 | 45,8% | 35,64 | 24,71 | 42,1% | 27,67 | - | - | - | - | - | - | - | - | - | - | - | - |
| 182 | D316 | 0,3 | (47, 2) | 1 | 2 | - | 27,10 | 45,8% | 35,64 | 24,97 | 43,3% | 27,99 | - | - | - | - | - | - | - | - | - | - | - | - |
| 183 | D316 | 0,3 | (47, 2) | 2 | 1 | - | 27,10 | 45,8% | 35,64 | 25,38 | 44,8% | 28,21 | - | - | - | - | - | - | - | - | - | - | - | - |
| 184 | D316 | 0,3 | (47, 2) | 2 | 2 | - | 27,10 | 45,8% | 35,64 | 25,13 | 44,0% | 28,06 | - | - | - | - | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 185 | D316 | 0,5 | (79, 2) | 1 | 1 | - | 28,86 | 45,7% | 39,62 | 25,87 | 38,6% | 27,52 | - | - | - | - | - | - | - | - | - | - | - | - |
| 186 | D316 | 0,5 | (79, 2) | 1 | 2 | - | 28,86 | 45,7% | 39,62 | 20,51 | 28,3% | 22,11 | - | - | - | - | - | - | - | - | - | - | - | - |
| 187 | D316 | 0,5 | (79, 2) | 2 | 1 | - | 28,86 | 45,7% | 39,62 | 15,71 | 17,9% | 16,90 | - | - | - | - | - | - | - | - | - | - | - | - |
| 188 | D316 | 0,5 | (79, 2) | 2 | 2 | - | 28,86 | 45,7% | 39,62 | 23,30 | 33,8% | 25,01 | - | - | - | - | - | - | - | - | - | - | - | - |
| 189 | D316 | 0,7 | (111, 2) | 1 | 1 | - | 3,67 | 5,1% | 4,94 | 18,24 | 23,5% | 19,20 | - | - | - | - | - | - | - | - | - | - | - | - |
| 190 | D316 | 0,7 | (111, 2) | 1 | 2 | - | 3,67 | 5,1% | 4,94 | 18,22 | 23,4% | 19,14 | - | - | - | - | - | - | - | - | - | - | - | - |
| 191 | D316 | 0,7 | (111, 2) | 2 | 1 | - | 3,67 | 5,1% | 4,94 | 14,30 | 16,7% | 15,03 | - | - | - | - | - | - | - | - | - | - | - | - |
| 192 | D316 | 0,7 | (111, 2) | 2 | 2 | - | 3,67 | 5,1% | 4,94 | 18,17 | 23,5% | 19,16 | - | - | - | - | - | - | - | - | - | - | - | - |
| 193 | T238 | 0,3 | (18, 2) | - | - | 4 | - | - | - | - | - | - | 13,22 | 22,9% | 13,34 | 13,22 | 22,9% | 13,22 | - | - | - | - | - | - |
| 194 | T238 | 0,3 | (18, 2) | - | - | 7 | - | - | - | - | - | - | 8,54 | 56,5% | 8,57 | 8,54 | 56,5% | 8,54 | - | - | - | - | - | - |
| 195 | T238 | 0,5 | (30, 2) | - | - | 4 | - | - | - | - | - | - | 20,19 | 41,7% | 20,19 | 20,19 | 41,7% | 20,19 | - | - | - | - | - | - |
| 196 | T238 | 0,5 | (30, 2) | - | - | 7 | - | - | - | - | - | - | 8,23 | 79,6% | 8,26 | 8,23 | 79,6% | 8,23 | - | - | - | - | - | - |
| 197 | T238 | 0,7 | (42, 2) | - | - | 4 | - | - | - | - | - | - | 4,23 | 28,1% | 4,24 | 4,23 | 28,1% | 4,23 | - | - | - | - | - | - |
| 198 | T238 | 0,7 | (42, 2) | - | - | 7 | - | - | - | - | - | - | 0,83 | 25,3% | 0,91 | 0,83 | 25,3% | 0,83 | - | - | - | - | - | - |
| 199 | S843 | 0,3 | (11, 2) | - | - | 4 | - | - | - | - | - | - | 24,17 | 376,1% | 24,21 | 24,17 | 376,1% | 24,17 | - | - | - | - | - | - |
| 200 | S843 | 0,3 | (11, 2) | - | - | 7 | - | - | - | - | - | - | 9,60 | 338,5% | 9,63 | 9,60 | 338,5% | 9,60 | - | - | - | - | - | - |
| 201 | S843 | 0,5 | (19, 2) | - | - | 4 | - | - | - | - | - | - | 7,87 | 184,8% | 7,89 | 7,87 | 184,8% | 7,87 | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 202 | S843 | 0,5 | (19, 2) | - | - | 7 | - | - | - | - | - | - | 3,83 | 23,9% | 3,97 | 3,83 | 23,9% | 3,83 | - | - | - | - | - | - |
| 203 | S843 | 0,7 | (26, 2) | - | - | 4 | - | - | - | - | - | - | 11,75 | 19,1% | 11,87 | 11,75 | 19,1% | 11,75 | - | - | - | - | - | - |
| 204 | S843 | 0,7 | (26, 2) | - | - | 7 | - | - | - | - | - | - | 10,40 | 352,8% | 10,42 | 10,40 | 352,8% | 10,40 | - | - | - | - | - | - |
| 205 | C434 | 0,3 | (42, 2) | - | - | 4 | - | - | - | - | - | - | 15,65 | 29,7% | 17,69 | 15,65 | 29,7% | 15,65 | - | - | - | - | - | - |
| 206 | C434 | 0,3 | (42, 2) | - | - | 7 | - | - | - | - | - | - | 29,08 | 47,4% | 32,60 | 29,08 | 47,4% | 29,08 | - | - | - | - | - | - |
| 207 | C434 | 0,5 | (71, 2) | - | - | 4 | - | - | - | - | - | - | 19,63 | 28,6% | 21,22 | 19,63 | 28,6% | 19,63 | - | - | - | - | - | - |
| 208 | C434 | 0,5 | (71, 2) | - | - | 7 | - | - | - | - | - | - | 17,90 | 25,1% | 19,09 | 17,90 | 25,1% | 17,90 | - | - | - | - | - | - |
| 209 | C434 | 0,7 | (99, 2) | - | - | 4 | - | - | - | - | - | - | 20,29 | 25,4% | 20,68 | 20,29 | 25,4% | 20,29 | - | - | - | - | - | - |
| 210 | C434 | 0,7 | (99, 2) | - | - | 7 | - | - | - | - | - | - | 15,97 | 17,0% | 16,24 | 15,97 | 17,0% | 15,97 | - | - | - | - | - | - |
| 211 | V104 | 0,3 | (46, 2) | - | - | 4 | - | - | - | - | - | - | 20,17 | 32,4% | 23,79 | 20,17 | 32,4% | 20,17 | - | - | - | - | - | - |
| 212 | V104 | 0,3 | (46, 2) | - | - | 7 | - | - | - | - | - | - | 18,79 | 31,5% | 22,22 | 18,79 | 31,5% | 18,79 | - | - | - | - | - | - |
| 213 | V104 | 0,5 | (77, 2) | - | - | 4 | - | - | - | - | - | - | 30,77 | 40,3% | 33,74 | 30,77 | 40,3% | 30,77 | - | - | - | - | - | - |
| 214 | V104 | 0,5 | (77, 2) | - | - | 7 | - | - | - | - | - | - | 22,01 | 33,7% | 23,69 | 22,01 | 33,7% | 22,01 | - | - | - | - | - | - |
| 215 | V104 | 0,7 | (108, 2) | - | - | 4 | - | - | - | - | - | - | 19,20 | 25,2% | 19,56 | 19,20 | 25,2% | 19,20 | - | - | - | - | - | - |
| 216 | V104 | 0,7 | (108, 2) | - | - | 7 | - | - | - | - | - | - | 10,15 | 13,0% | 10,21 | 10,15 | 13,0% | 10,15 | - | - | - | - | - | - |
| 217 | A172 | 0,3 | (18, 2) | - | - | 4 | - | - | - | - | - | - | 25,69 | 56,8% | 26,71 | 25,69 | 56,8% | 25,69 | - | - | - | - | - | - |
| 218 | A172 | 0,3 | (18, 2) | - | - | 7 | - | - | - | - | - | - | 12,51 | 46,8% | 13,18 | 12,51 | 46,8% | 12,51 | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 219 | A172 | 0,5 | (31, 2) | - | - | 4 | - | - | - | - | - | - | 23,56 | 70,7% | 23,88 | 23,56 | 70,7% | 23,56 | - | - | - | - | - | - |
| 220 | A172 | 0,5 | (31, 2) | - | - | 7 | - | - | - | - | - | - | 20,81 | 84,0% | 21,06 | 20,81 | 84,0% | 20,81 | - | - | - | - | - | - |
| 221 | A172 | 0,7 | (43, 2) | - | - | 4 | - | - | - | - | - | - | 12,47 | 38,4% | 12,61 | 12,47 | 38,4% | 12,47 | - | - | - | - | - | - |
| 222 | A172 | 0,7 | (43, 2) | - | - | 7 | - | - | - | - | - | - | 9,46 | 36,4% | 9,64 | 9,46 | 36,4% | 9,46 | - | - | - | - | - | - |
| 223 | M941 | 0,3 | (30, 2) | - | - | 4 | - | - | - | - | - | - | 12,71 | 35,9% | 13,01 | 12,71 | 35,9% | 12,71 | - | - | - | - | - | - |
| 224 | M941 | 0,3 | (30, 2) | - | - | 7 | - | - | - | - | - | - | 10,72 | 30,1% | 11,00 | 10,72 | 30,1% | 10,72 | - | - | - | - | - | - |
| 225 | M941 | 0,5 | (51, 2) | - | - | 4 | - | - | - | - | - | - | 8,54 | 28,1% | 8,67 | 8,54 | 28,1% | 8,54 | - | - | - | - | - | - |
| 226 | M941 | 0,5 | (51, 2) | - | - | 7 | - | - | - | - | - | - | 10,24 | 27,3% | 10,31 | 10,24 | 27,3% | 10,24 | - | - | - | - | - | - |
| 227 | M941 | 0,7 | (71, 2) | - | - | 4 | - | - | - | - | - | - | 5,09 | 13,0% | 5,14 | 5,09 | 13,0% | 5,09 | - | - | - | - | - | - |
| 228 | M941 | 0,7 | (71, 2) | - | - | 7 | - | - | - | - | - | - | 5,73 | 17,7% | 5,87 | 5,73 | 17,7% | 5,73 | - | - | - | - | - | - |
| 229 | L406 | 0,3 | (23, 2) | - | - | 4 | - | - | - | - | - | - | 21,75 | 38,9% | 22,18 | 21,75 | 38,9% | 21,75 | - | - | - | - | - | - |
| 230 | L406 | 0,3 | (23, 2) | - | - | 7 | - | - | - | - | - | - | 12,94 | 21,7% | 13,64 | 12,94 | 21,7% | 12,94 | - | - | - | - | - | - |
| 231 | L406 | 0,5 | (39, 2) | - | - | 4 | - | - | - | - | - | - | 9,53 | 22,8% | 9,74 | 9,53 | 22,8% | 9,53 | - | - | - | - | - | - |
| 232 | L406 | 0,5 | (39, 2) | - | - | 7 | - | - | - | - | - | - | 2,58 | 0,05 | 2,89 | 2,58 | 0,05 | 2,58 | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 233 | L406 | 0,7 | (54, 2) | - | - | 4 | - | - | - | - | - | - | 6,55 | 19,8% | 6,70 | 6,55 | 19,8% | 6,55 | - | - | - | - | - | - |
| 234 | L406 | 0,7 | (54, 2) | - | - | 7 | - | - | - | - | - | - | 10,35 | 19,1% | 10,40 | 10,35 | 19,1% | 10,35 | - | - | - | - | - | - |
| 235 | C907 | 0,3 | (20, 2) | - | - | 4 | - | - | - | - | - | - | 9,79 | 45,6% | 10,59 | 9,79 | 45,6% | 9,79 | - | - | - | - | - | - |
| 236 | C907 | 0,3 | (20, 2) | - | - | 7 | - | - | - | - | - | - | 18,71 | 67,1% | 18,94 | 18,71 | 67,1% | 18,71 | - | - | - | - | - | - |
| 237 | C907 | 0,5 | (34, 2) | - | - | 4 | - | - | - | - | - | - | 11,29 | 20,3% | 11,35 | 11,29 | 20,3% | 11,29 | - | - | - | - | - | - |
| 238 | C907 | 0,5 | (34, 2) | - | - | 7 | - | - | - | - | - | - | 9,93 | 30,2% | 10,00 | 9,93 | 30,2% | 9,93 | - | - | - | - | - | - |
| 239 | C907 | 0,7 | (48, 2) | - | - | 4 | - | - | - | - | - | - | 5,97 | 26,4% | 6,19 | 5,97 | 26,4% | 5,97 | - | - | - | - | - | - |
| 240 | C907 | 0,7 | (48, 2) | - | - | 7 | - | - | - | - | - | - | 5,46 | 22,0% | 5,65 | 5,46 | 22,0% | 5,46 | - | - | - | - | - | - |
| 241 | M174 | 0,3 | (26, 2) | - | - | 4 | - | - | - | - | - | - | 20,01 | 29,3% | 20,87 | 20,01 | 29,3% | 20,01 | - | - | - | - | - | - |
| 242 | M174 | 0,3 | (26, 2) | - | - | 7 | - | - | - | - | - | - | 16,21 | 34,1% | 16,67 | 16,21 | 34,1% | 16,21 | - | - | - | - | - | - |
| 243 | M174 | 0,5 | (43, 2) | - | - | 4 | - | - | - | - | - | - | 15,50 | 28,7% | 15,58 | 15,50 | 28,7% | 15,50 | - | - | - | - | - | - |
| 244 | M174 | 0,5 | (43, 2) | - | - | 7 | - | - | - | - | - | - | 15,76 | 27,7% | 15,79 | 15,76 | 27,7% | 15,76 | - | - | - | - | - | - |
| 245 | M174 | 0,7 | (60, 2) | - | - | 4 | - | - | - | - | - | - | 13,59 | 20,0% | 13,59 | 13,59 | 20,0% | 13,59 | - | - | - | - | - | - |
| 246 | M174 | 0,7 | (60, 2) | - | - | 7 | - | - | - | - | - | - | 0,65 | 0,01 | 0,77 | 0,65 | 0,01 | 0,65 | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 247 | S242 | 0,3 | (49, 2) | - | - | 4 | - | - | - | - | - | - | 23,46 | 45,9% | 24,49 | 23,46 | 45,9% | 23,46 | - | - | - | - | - | - |
| 248 | S242 | 0,3 | (49, 2) | - | - | 7 | - | - | - | - | - | - | 16,86 | 37,9% | 18,15 | 16,86 | 37,9% | 16,86 | - | - | - | - | - | - |
| 249 | S242 | 0,5 | (82, 2) | - | - | 4 | - | - | - | - | - | - | 14,50 | 31,0% | 15,31 | 14,50 | 31,0% | 14,50 | - | - | - | - | - | - |
| 250 | S242 | 0,5 | (82, 2) | - | - | 7 | - | - | - | - | - | - | 11,54 | 24,8% | 12,22 | 11,54 | 24,8% | 11,54 | - | - | - | - | - | - |
| 251 | S242 | 0,7 | (114, 2) | - | - | 4 | - | - | - | - | - | - | 9,64 | 17,4% | 10,08 | 9,64 | 17,4% | 9,64 | - | - | - | - | - | - |
| 252 | S242 | 0,7 | (114, 2) | - | - | 7 | - | - | - | - | - | - | 2,08 | 3,6% | 2,40 | 2,08 | 3,6% | 2,08 | - | - | - | - | - | - |
| 253 | L868 | 0,3 | (31, 2) | - | - | 4 | - | - | - | - | - | - | 17,33 | 49,1% | 18,35 | 17,33 | 49,1% | 17,33 | - | - | - | - | - | - |
| 254 | L868 | 0,3 | (31, 2) | - | - | 7 | - | - | - | - | - | - | 17,37 | 51,6% | 18,25 | 17,37 | 51,6% | 17,37 | - | - | - | - | - | - |
| 255 | L868 | 0,5 | (52, 2) | - | - | 4 | - | - | - | - | - | - | 12,47 | 43,7% | 12,83 | 12,47 | 43,7% | 12,47 | - | - | - | - | - | - |
| 256 | L868 | 0,5 | (52, 2) | - | - | 7 | - | - | - | - | - | - | 7,86 | 31,1% | 8,15 | 7,86 | 31,1% | 7,86 | - | - | - | - | - | - |
| 257 | L868 | 0,7 | (73, 2) | - | - | 4 | - | - | - | - | - | - | 3,82 | 13,5% | 3,91 | 3,82 | 13,5% | 3,82 | - | - | - | - | - | - |
| 258 | L868 | 0,7 | (73, 2) | - | - | 7 | - | - | - | - | - | - | 9,24 | 30,6% | 9,28 | 9,24 | 30,6% | 9,24 | - | - | - | - | - | - |
| 259 | A739 | 0,3 | (51, 2) | - | - | 4 | - | - | - | - | - | - | 8,01 | 19,7% | 8,13 | 8,01 | 19,7% | 8,01 | - | - | - | - | - | - |
| 260 | A739 | 0,3 | (51, 2) | - | - | 7 | - | - | - | - | - | - | 5,24 | 14,6% | 5,41 | 5,24 | 14,6% | 5,24 | - | - | - | - | - | - |
| 261 | A739 | 0,5 | (86, 2) | - | - | 4 | - | - | - | - | - | - | 3,72 | 7,9% | 3,81 | 3,72 | 7,9% | 3,72 | - | - | - | - | - | - |
| 262 | A739 | 0,5 | (86, 2) | - | - | 7 | - | - | - | - | - | - | 3,69 | 7,7% | 3,81 | 3,69 | 7,7% | 3,69 | - | - | - | - | - | - |
| 263 | A739 | 0,7 | (121, 2) | - | - | 4 | - | - | - | - | - | - | 1,31 | 1,8% | 1,50 | 1,31 | 1,8% | 1,31 | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 264 | A739 | 0,7 | (121, 2) | - | - | 7 | - | - | - | - | - | - | 0,84 | 1,3% | 1,05 | 0,84 | 1,3% | 0,84 | - | - | - | - | - | - |
| 265 | C624 | 0,3 | (32, 2) | - | - | 4 | - | - | - | - | - | - | 7,14 | 30,9% | 7,30 | 7,14 | 30,9% | 7,14 | - | - | - | - | - | - |
| 266 | C624 | 0,3 | (32, 2) | - | - | 7 | - | - | - | - | - | - | 7,30 | 35,7% | 7,38 | 7,30 | 35,7% | 7,30 | - | - | - | - | - | - |
| 267 | C624 | 0,5 | (53, 2) | - | - | 4 | - | - | - | - | - | - | 3,59 | 30,3% | 3,65 | 3,59 | 30,3% | 3,59 | - | - | - | - | - | - |
| 268 | C624 | 0,5 | (53, 2) | - | - | 7 | - | - | - | - | - | - | 3,27 | 35,7% | 3,33 | 3,27 | 35,7% | 3,27 | - | - | - | - | - | - |
| 269 | C624 | 0,7 | (74, 2) | - | - | 4 | - | - | - | - | - | - | 2,66 | 22,6% | 2,73 | 2,66 | 22,6% | 2,66 | - | - | - | - | - | - |
| 270 | C624 | 0,7 | (74, 2) | - | - | 7 | - | - | - | - | - | - | 2,54 | 21,7% | 2,59 | 2,54 | 21,7% | 2,54 | - | - | - | - | - | - |
| 271 | A676 | 0,3 | (22, 2) | - | - | 4 | - | - | - | - | - | - | 21,67 | Error | 21,72 | 21,67 | Error | 21,67 | - | - | - | - | - | - |
| 272 | A676 | 0,3 | (22, 2) | - | - | 7 | - | - | - | - | - | - | 4,26 | Error | 4,38 | 4,26 | Error | 4,26 | - | - | - | - | - | - |
| 273 | A676 | 0,5 | (37, 2) | - | - | 4 | - | - | - | - | - | - | 3,37 | Error | 3,58 | 3,37 | Error | 3,37 | - | - | - | - | - | - |
| 274 | A676 | 0,5 | (37, 2) | - | - | 7 | - | - | - | - | - | - | 4,85 | 13,9% | 5,22 | 4,85 | 13,9% | 4,85 | - | - | - | - | - | - |
| 275 | A676 | 0,7 | (52, 2) | - | - | 4 | - | - | - | - | - | - | 2,89 | Error | 3,01 | 2,89 | Error | 2,89 | - | - | - | - | - | - |
| 276 | A676 | 0,7 | (52, 2) | - | - | 7 | - | - | - | - | - | - | 1,72 | 19,9% | 2,18 | 1,72 | 19,9% | 1,72 | - | - | - | - | - | - |
| 277 | S165 | 0,3 | (17, 2) | - | - | 4 | - | - | - | - | - | - | 18,92 | 93,8% | 19,03 | 18,92 | 93,8% | 18,92 | - | - | - | - | - | - |
| 278 | S165 | 0,3 | (17, 2) | - | - | 7 | - | - | - | - | - | - | 9,70 | 58,7% | 9,84 | 9,70 | 58,7% | 9,70 | - | - | - | - | - | - |
| 279 | S165 | 0,5 | (28, 2) | - | - | 4 | - | - | - | - | - | - | 19,26 | 87,2% | 19,30 | 19,26 | 87,2% | 19,26 | - | - | - | - | - | - |
| 280 | S165 | 0,5 | (28, 2) | - | - | 7 | - | - | - | - | - | - | 7,70 | 58,8% | 7,80 | 7,70 | 58,8% | 7,70 | - | - | - | - | - | - |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 281 | S165 | 0,7 | (39, 2) | - | - | 4 | - | - | - | - | - | - | 3,07 | 22,7% | 3,21 | 3,07 | 22,7% | 3,07 | - | - | - | - | - | - |
| 282 | S165 | 0,7 | (39, 2) | - | - | 7 | - | - | - | - | - | - | 5,45 | 35,6% | 5,54 | 5,45 | 35,6% | 5,45 | - | - | - | - | - | - |
| 283 | D316 | 0,3 | (50, 2) | - | - | 4 | - | - | - | - | - | - | 21,70 | 42,6% | 22,46 | 21,70 | 42,6% | 21,70 | - | - | - | - | - | - |
| 284 | D316 | 0,3 | (50, 2) | - | - | 7 | - | - | - | - | - | - | 17,40 | 36,7% | 18,00 | 17,40 | 36,7% | 17,40 | - | - | - | - | - | - |
| 285 | D316 | 0,5 | (83, 2) | - | - | 4 | - | - | - | - | - | - | 17,26 | 31,8% | 17,37 | 17,26 | 31,8% | 17,26 | - | - | - | - | - | - |
| 286 | D316 | 0,5 | (83, 2) | - | - | 7 | - | - | - | - | - | - | 8,23 | 15,4% | 8,28 | 8,23 | 15,4% | 8,23 | - | - | - | - | - | - |
| 287 | D316 | 0,7 | (116, 2) | - | - | 4 | - | - | - | - | - | - | 2,37 | 4,5% | 2,46 | 2,37 | 4,5% | 2,37 | - | - | - | - | - | - |
| 288 | D316 | 0,7 | (116, 2) | - | - | 7 | - | - | - | - | - | - | 1,33 | 2,1% | 1,57 | 1,33 | 2,1% | 1,33 | - | - | - | - | - | - |
| 289 | T238 | 0,3 | (18, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 6,730769 | 272,3% | 6,776427 | 6,730769 | 272,3% | 6,730769 |
| 290 | T238 | 0,3 | (18, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 9,222222 | 375,8% | 9,275246 | 9,222222 | 375,8% | 9,222222 |
| 291 | T238 | 0,5 | (30, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 7,351852 | 310,3% | 7,42892 | 7,351852 | 310,3% | 7,351852 |
| 292 | T238 | 0,5 | (30, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 7,375 | 343,1% | 7,426191 | 7,375 | 343,1% | 7,375 |
| 293 | T238 | 0,7 | (42, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 10 | 386,9% | 10,0241 | 10 | 386,9% | 10 |
| 294 | T238 | 0,7 | (42, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 9,291667 | 325,3% | 9,339337 | 9,291667 | 325,3% | 9,291667 |
| 295 | S843 | 0,3 | (11, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 7,608696 | 518,7% | 7,63629 | 7,608696 | 518,7% | 7,608696 |
| 296 | S843 | 0,3 | (11, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 9,2 | 596,7% | 9,216627 | 9,2 | 596,7% | 9,2 |
| 297 | S843 | 0,5 | (19, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 9,9 | 646,3% | 9,913451 | 9,9 | 646,3% | 9,9 |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 298 | S843 | 0,5 | (19, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 7,833333 | 492,5% | 7,876603 | 7,833333 | 492,5% | 7,833333 |
| 299 | S843 | 0,7 | (26, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 6,625 | 298,0% | 6,670163 | 6,625 | 298,0% | 6,625 |
| 300 | S843 | 0,7 | (26, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 8,8 | 351,5% | 8,843909 | 8,8 | 351,5% | 8,8 |
| 301 | C434 | 0,3 | (42, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 20,16146 | 33,0% | 24,03345 | 20,16146 | 33,0% | 20,16146 |
| 302 | C434 | 0,3 | (42, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 27,44624 | 40,9% | 33,64691 | 27,44624 | 40,9% | 27,44624 |
| 303 | C434 | 0,5 | (71, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 24,23881 | 38,3% | 25,94568 | 24,23881 | 38,3% | 24,23881 |
| 304 | C434 | 0,5 | (71, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 31,53125 | 43,6% | 35,26818 | 31,53125 | 43,6% | 31,53125 |
| 305 | C434 | 0,7 | (99, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 45,58974 | 52,4% | 47,07377 | 45,58974 | 52,4% | 45,58974 |
| 306 | C434 | 0,7 | (99, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 43,72222 | 51,9% | 44,88753 | 43,72222 | 51,9% | 43,72222 |
| 307 | V104 | 0,3 | (46, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 28,0619 | 44,7% | 35,36942 | 28,0619 | 44,7% | 28,0619 |
| 308 | V104 | 0,3 | (46, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 20,89706 | 36,3% | 24,01337 | 20,89706 | 36,3% | 20,89706 |
| 309 | V104 | 0,5 | (77, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 28,13514 | 41,5% | 31,16148 | 28,13514 | 41,5% | 28,13514 |
| 310 | V104 | 0,5 | (77, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 26,38028 | 41,6% | 28,19667 | 26,38028 | 41,6% | 26,38028 |
| 311 | V104 | 0,7 | (108, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 42,98837 | 53,2% | 43,89467 | 42,98837 | 53,2% | 42,98837 |
| 312 | V104 | 0,7 | (108, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 56,3 | 65,0% | 57,70204 | 56,3 | 65,0% | 56,3 |
| 313 | A172 | 0,3 | (18, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 4,6875 | 27,1% | 5,402686 | 4,6875 | 27,1% | 4,6875 |
| 314 | A172 | 0,3 | (18, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 5,243243 | 23,1% | 6,469523 | 5,243243 | 23,1% | 5,243243 |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 315 | A172 | 0,5 | (31, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 7,259259 | 29,7% | 7,899484 | 7,259259 | 29,7% | 7,259259 |
| 316 | A172 | 0,5 | (31, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 6,979167 | 27,5% | 7,544881 | 6,979167 | 27,5% | 6,979167 |
| 317 | A172 | 0,7 | (43, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 8,333333 | 27,7% | 8,50838 | 8,333333 | 27,7% | 8,333333 |
| 318 | A172 | 0,7 | (43, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 9,041667 | 28,9% | 9,159714 | 9,041667 | 28,9% | 9,041667 |
| 319 | M941 | 0,3 | (30, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 8,169118 | 24,9% | 8,689738 | 8,169118 | 24,9% | 8,169118 |
| 320 | M941 | 0,3 | (30, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 9,907692 | 27,1% | 10,39606 | 9,907692 | 27,1% | 9,907692 |
| 321 | M941 | 0,5 | (51, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 11,43617 | 28,7% | 11,53939 | 11,43617 | 28,7% | 11,43617 |
| 322 | M941 | 0,5 | (51, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 11,69318 | 30,9% | 11,75887 | 11,69318 | 30,9% | 11,69318 |
| 323 | M941 | 0,7 | (71, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 12,88889 | 33,3% | 12,92158 | 12,88889 | 33,3% | 12,88889 |
| 324 | M941 | 0,7 | (71, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 14,5 | 34,5% | 14,52628 | 14,5 | 34,5% | 14,5 |
| 325 | L406 | 0,3 | (23, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 10,18627 | 24,8% | 11,02303 | 10,18627 | 24,8% | 10,18627 |
| 326 | L406 | 0,3 | (23, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 9,5625 | 24,5% | 10,31125 | 9,5625 | 24,5% | 9,5625 |
| 327 | L406 | 0,5 | (39, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 9,7 | 25,5% | 9,942806 | 9,7 | 25,5% | 9,7 |
| 328 | L406 | 0,5 | (39, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 11,51563 | 29,0% | 11,71844 | 11,51563 | 29,0% | 11,51563 |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 329 | L406 | 0,7 | (54, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 12,55 | 27,6% | 12,62434 | 12,55 | 27,6% | 12,55 |
| 330 | L406 | 0,7 | (54, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 12,55882 | 28,5% | 12,60672 | 12,55882 | 28,5% | 12,55882 |
| 331 | C907 | 0,3 | (20, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 3,311111 | 15,6% | 3,758934 | 3,311111 | 15,6% | 3,311111 |
| 332 | C907 | 0,3 | (20, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 5,5 | 16,9% | 6,325168 | 5,5 | 16,9% | 5,5 |
| 333 | C907 | 0,5 | (34, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 5,209677 | 17,0% | 5,437796 | 5,209677 | 17,0% | 5,209677 |
| 334 | C907 | 0,5 | (34, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 5,25 | 15,6% | 5,477147 | 5,25 | 15,6% | 5,25 |
| 335 | C907 | 0,7 | (48, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 7,323529 | 23,7% | 7,387716 | 7,323529 | 23,7% | 7,323529 |
| 336 | C907 | 0,7 | (48, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 5,035714 | 14,6% | 5,280978 | 5,035714 | 14,6% | 5,035714 |
| 337 | M174 | 0,3 | (26, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 18,92982 | 35,0% | 20,07659 | 18,92982 | 35,0% | 18,92982 |
| 338 | M174 | 0,3 | (26, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 19,57407 | 36,5% | 20,50346 | 19,57407 | 36,5% | 19,57407 |
| 339 | M174 | 0,5 | (43, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 23,5625 | 40,3% | 24,11692 | 23,5625 | 40,3% | 23,5625 |
| 340 | M174 | 0,5 | (43, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 21,21622 | 38,1% | 21,51321 | 21,21622 | 38,1% | 21,21622 |
| 341 | M174 | 0,7 | (60, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 25,32609 | 41,6% | 25,33897 | 25,32609 | 41,6% | 25,32609 |
| 342 | M174 | 0,7 | (60, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 22,25 | 37,2% | 22,25551 | 22,25 | 37,2% | 22,25 |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 343 | S242 | 0,3 | (49, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 21,0045 | 44,8% | 24,031 | 21,0045 | 44,8% | 21,0045 |
| 344 | S242 | 0,3 | (49, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 16,33796 | 37,5% | 17,74991 | 16,33796 | 37,5% | 16,33796 |
| 345 | S242 | 0,5 | (82, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 20,0641 | 42,4% | 20,88921 | 20,0641 | 42,4% | 20,0641 |
| 346 | S242 | 0,5 | (82, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 30,76 | 52,7% | 34,33654 | 30,76 | 52,7% | 30,76 |
| 347 | S242 | 0,7 | (114, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 27,32609 | 49,6% | 28,37755 | 27,32609 | 49,6% | 27,32609 |
| 348 | S242 | 0,7 | (114, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 36,26744 | 60,1% | 37,89541 | 36,26744 | 60,1% | 36,26744 |
| 349 | L868 | 0,3 | (31, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 7,3 | 27,7% | 8,622572 | 7,3 | 27,7% | 7,3 |
| 350 | L868 | 0,3 | (31, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 7,962687 | 29,2% | 9,175155 | 7,962687 | 29,2% | 7,962687 |
| 351 | L868 | 0,5 | (52, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 8,22449 | 31,4% | 8,629509 | 8,22449 | 31,4% | 8,22449 |
| 352 | L868 | 0,5 | (52, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 9,967391 | 34,0% | 10,75392 | 9,967391 | 34,0% | 9,967391 |
| 353 | L868 | 0,7 | (73, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 13,01786 | 40,0% | 13,17723 | 13,01786 | 40,0% | 13,01786 |
| 354 | L868 | 0,7 | (73, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 12,84 | 38,8% | 12,91585 | 12,84 | 38,8% | 12,84 |
| 355 | A739 | 0,3 | (51, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 10,57627 | 25,1% | 10,69583 | 10,57627 | 25,1% | 10,57627 |
| 356 | A739 | 0,3 | (51, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 10,79565 | 23,7% | 10,93739 | 10,79565 | 23,7% | 10,79565 |
| 357 | A739 | 0,5 | (86, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 14,71084 | 32,8% | 14,74641 | 14,71084 | 32,8% | 14,71084 |
| 358 | A739 | 0,5 | (86, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 11,8625 | 26,8% | 11,91028 | 11,8625 | 26,8% | 11,8625 |
| 359 | A739 | 0,7 | (121, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 14,53125 | 30,0% | 14,55624 | 14,53125 | 30,0% | 14,53125 |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 360 | A739 | 0,7 | (121, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 14,6 | 26,8% | 14,62655 | 14,6 | 26,8% | 14,6 |
| 361 | C624 | 0,3 | (32, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 2,73 | 0,29 | 2,94 | 2,73 | 0,29 | 2,73 |
| 362 | C624 | 0,3 | (32, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 6,147059 | 49,9% | 6,395541 | 6,147059 | 49,9% | 6,147059 |
| 363 | C624 | 0,5 | (53, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 3,52 | 0,36 | 3,65 | 3,52 | 0,36 | 3,52 |
| 364 | C624 | 0,5 | (53, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 4,510638 | 33,8% | 4,621923 | 4,510638 | 33,8% | 4,510638 |
| 365 | C624 | 0,7 | (74, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 2,87931 | 24,1% | 2,968051 | 2,87931 | 24,1% | 2,87931 |
| 366 | C624 | 0,7 | (74, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 5,230769 | 31,9% | 5,335834 | 5,230769 | 31,9% | 5,230769 |
| 367 | A676 | 0,3 | (22, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 7,183673 | Error | 7,295585 | 7,183673 | Error | 7,183673 |
| 368 | A676 | 0,3 | (22, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 8,228261 | Error | 8,353917 | 8,228261 | Error | 8,228261 |
| 369 | A676 | 0,5 | (37, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 7,632353 | Error | 7,737555 | 7,632353 | Error | 7,632353 |
| 370 | A676 | 0,5 | (37, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 9,370968 | Error | 9,466076 | 9,370968 | Error | 9,370968 |
| 371 | A676 | 0,7 | (52, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 8,342105 | Error | 8,480202 | 8,342105 | Error | 8,342105 |
| 372 | A676 | 0,7 | (52, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 8,25 | Error | 8,450899 | 8,25 | Error | 8,25 |
| 373 | S165 | 0,3 | (17, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 3,819444 | 46,3% | 3,928351 | 3,819444 | 46,3% | 3,819444 |
| 374 | S165 | 0,3 | (17, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 3,818182 | 52,2% | 3,990194 | 3,818182 | 52,2% | 3,818182 |
| 375 | S165 | 0,5 | (28, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 4,38 | 53,3% | 4,521335 | 4,38 | 53,3% | 4,38 |
| 376 | S165 | 0,5 | (28, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 5,090909 | 57,5% | 5,308162 | 5,090909 | 57,5% | 5,090909 |

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 377 | S165 | 0,7 | (39, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 2,964286 | 24,3% | 3,104803 | 2,964286 | 24,3% | 2,964286 |
| 378 | S165 | 0,7 | (39, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 2,50 | 0,22 | 2,68 | 2,50 | 0,22 | 2,50 |
| 379 | D316 | 0,3 | (50, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 17,4646 | 35,6% | 18,30505 | 17,4646 | 35,6% | 17,4646 |
| 380 | D316 | 0,3 | (50, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 15,19545 | 29,8% | 15,87615 | 15,19545 | 29,8% | 15,19545 |
| 381 | D316 | 0,5 | (83, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 15,95 | 33,5% | 16,05758 | 15,95 | 33,5% | 15,95 |
| 382 | D316 | 0,5 | (83, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 14,35714 | 29,5% | 14,60163 | 14,35714 | 29,5% | 14,35714 |
| 383 | D316 | 0,7 | (116, 2) | - | - | 4 | - | - | - | - | - | - | - | - | - | - | - | - | 16,32979 | 32,4% | 16,43299 | 16,32979 | 32,4% | 16,32979 |
| 384 | D316 | 0,7 | (116, 2) | - | - | 7 | - | - | - | - | - | - | - | - | - | - | - | - | 17,98864 | 33,8% | 18,00824 | 17,98864 | 33,8% | 17,98864 |
| | | | | | | | **9,76** | **19,5%** | 12,33 | 11,67 | 25,2% | 13,22 | 11,39 | 43,4% | 11,83 | 11,39 | 43,4% | 11,39 | **13,71** | **84,5%** | **14,42** | **13,71** | **84,5%** | **13,71** |
| | | | | | | | VAR | | | VARIMA | | | Local CNN | | | Local LSTM | | | **Global CNN** | | | **Global LSTM** | | |
| | | | | | | | **0,36** | **0,3%** | 0,60 | 0,68 | 2,8% | 0,88 | 0,65 | 0,6% | 0,77 | 0,65 | 0,6% | 0,65 | **2,50** | **14,6%** | **2,68** | **2,50** | **14,6%** | **2,50** |