Assessment of Pump Failures in Rotterdam: A Five-Year Study (2016-2020)

A Failure Analysis based on statistical modelling

Qiwen Zhang

TU Delft

Assessment of Pump Failures in Rotterdam: A Five-Year Study (2016-2020)

A Failure Analysis based on statistical modelling

by

Qiwen Zhang

Student Name	Qiwen Zhang
Student Number	5458595

Supervisor and Chair:	Dr. Ir. Jeroen Langeveld, Associate Professor
Thesis Committee:	Dr. Riccardo Taormina, Assistant Professor
	Prof. Dr. Ir. Jan Peter van der Hoek, Full Professor
Project Duration:	October, 2022 - July, 2023
Faculty:	Faculty Of Civil Engineering and Geosciences, Delft

Cover Design: Qiwen and his brother

TUDelft

Preface

The completion of this thesis signifies not just the fruit of years of diligent research and dedication, but also the final stride in my journey through the master's program in Water Management at Delft University of Technology. This study is reflective of my personal passion for the subject, as well as the desire to contribute meaningful solutions to the world's water management challenges.

In the early stages of this program, the course 'Water Management in Urban Areas' ignited in me a keen interest in solving practical problems within our water systems. It allowed me a glimpse into the complex world of urban water management, illuminating the interconnected issues we face in this field. The insights I gained from this course were instrumental in guiding me to a thesis subject that resonated with my interests and ambitions.

The choice of thesis subject was vital, considering the significant effort I was prepared to, and indeed did, invest in this project. The personal resonance I found in this topic propelled my enthusiasm, energizing my drive to excel. This, combined with the guidance of my mentors, has shaped the body of work that this thesis has become.

I extend my deepest gratitude to Jeroen Langeveld, whose guidance and consistent support throughout this endeavor have been nothing short of extraordinary. Our weekly conversations were a beacon of light in an often arduous process, illuminating the path and ensuring steady progress. Your astute insights have helped me navigate through challenges, while your constructive feedback constantly pushed me towards excellence. This journey was made easier because you were beside me, guiding me every step of the way, helping to transform an initially daunting task into a manageable, enriching experience.

In the same vein, I am sincerely indebted to Riccardo Taormina, who provided indispensable technical support and expertise. Your skillful scrutiny of the logic and methods employed in my research, as well as your suggestions for improvement, have been essential in elevating the quality of this thesis. You challenged me to address complex technicalities, ensuring that my work remained both robust and credible. Without your technical acumen, this study would not have reached its current level of rigor and precision.

Both of you have been the lighthouses on this academic journey, illuminating my way through the uncertainty and stimulating my intellectual growth. The collaborative mentorship you provided has shaped this thesis and me as a scholar, and for that, I am eternally grateful.

Lastly, I must extend my gratitude to the wider academic community, my friends, and my family. Your continued encouragement and unwavering faith in my abilities have been a pillar of support.

To all, my heartfelt thanks.

Qiwen Zhang
Delft, September 2023

Summary

Sewage systems hold a critical function within the fabric of urban infrastructure, primarily by ensuring efficient wastewater management. Within this network, sewage pumping stations emerge as important components. The performance of such stations is susceptible to being undermined by system failures and malfunctions, which can cause significant implications for both public health and occurrences of Combined Sewer Overflow (CSO) events. Consequently, discerning the patterns and root causes of these failures is of great importance.

The present study explores and seeks to understand the reliability of sewage pumping stations, particularly in the setting of the Department of Public Works in Rotterdam. The scope of this study is confined to data collected between 2016 and 2020. The research encompasses a diverse range of interconnected aspects, such as examining discrepancies in failure data, categorizing types of failures, tracking changes in failure patterns over time, and selecting a fitting statistical model to accurately represent the findings.

The core focus of this study is to formulate an analysis framework to assess the reliability and failure patterns in wastewater pumping stations. In these patterns, various trends can be discerned through interarrival time - the period between two failures. For instance, an increasing trend indicates less frequent failures, suggesting improved reliability of the pumps. The devised framework combines both objective and subjective trend analyses, multiple trend tests, and employs a range of models such as the Homogeneous Poisson Process (HPP), Renewal Process (RP), and Non-Homogeneous Poisson Process (NHPP). This methodological approach is structured to facilitate a better understanding of the patterns of system behavior and shifts.

Analyzing pump performance patterns over five years revealed various trends among 447 pumps studied. Of these, 254 had consistent failure patterns; 98 showed a stable trend, 146 improved over time with fewer failures, and 10 experienced more failures, indicating declining performance. The remaining pumps were divided into various trend groups or segments, demonstrating different patterns in their performance over the five years. The application of statistical methods elucidated these failure patterns, contributing to the evaluation of sewage pump station efficiency.

Notwithstanding these progressions, my investigation has identified certain domains that could derive advantages from additional enhancements. The current methodology, which entailed manual configuration of kernel parameters and dependence on piecewise linear regression, was deemed insufficient for about 50% of the pumps. The study proposes the integration of sophisticated parameter estimation techniques, including Bayesian optimization or grid search, and alternative modeling methodologies to tackle this issue. It is recommended that forthcoming research expands its scope beyond Rotterdam and investigates a wider variety of pump mechanisms to validate the generalizability of the findings presented in this study.

In conclusion, this thesis develops an analytical framework based on examination of failure patterns and system dynamics. It establishes a platform for future progress by outlining a distinct pathway for prospective investigations pertaining to the failures in sewage pumping stations.

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
ACF	Autocorrelation Function
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
BIC	Bayesian Information Criterion
CDF	Cumulative Distribution Function
CSO	Combined Sewer Overflows
ECDF	Empirical Cumulative Distribution Functions
EPR	Evolutionary Polynomial Regression
GP	Gaussian Process
HPP	Homogeneous Poisson Process
K-S Test	Kolmogorov-Smirnov Test
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
MSE	Mean Squared Error
NHPP	Non-Homogeneous Poisson Process
PDF	Probability Density Function
PLP	Crow's Power Law Process models
PWLF	Piecewise Linear Fit (Python Library)
Q-Q Plot	Quantile-Quantile Plot
RBF	Radial Basis Function
RP	Renewal Process
RSS	Residual Sum of Squares
segment	a group of data with consistent interarrival time trend, which can be classified as increasing, decreasing, or constant
TBF	Time Between Failures

1 Introduction

1.1. Background

Modern wastewater management systems heavily rely on sewage pumping stations, which enable the efficient transport of wastewater to treatment facilities. These facilities help protect the environment and public health by overcoming topographical obstacles. They also promote sustainability by preventing CSO events. CSO events are a concern because they can lead to water pollution, potentially negatively impacting public health and local ecosystems. Amidst the backdrop of increasing urbanization and population expansion, it is crucial to prioritize the allocation of resources towards the development, upkeep, and implementation of advanced technology in order to sustain the long-term effectiveness of wastewater management infrastructure.

Factors like component deterioration, manufacturing defects, climate change, complicate the management of urban water systems (Martins et al., 2013). Failures in industries with substantial stakes can lead to severe consequences, including safety risks, and reputational harm (Garmabaki et al., 2016). Effective maintenance reduces the risk of failure by lowering the likelihood or severity of failure, which is crucial for ensuring system reliability. In this context, failure analysis models play a vital role in planning and decision support processes, aiding in prioritizing system rehabilitation measures according to identified risks and potential failure impacts.

Modeling pump failures can inform planning and decision-making processes, identify areas with a higher failure propensity, and provide an objective basis for prioritizing maintenance actions. Pump failures at these stations indicate an inability to function per design specifications, potentially leading to sewer overflow, inundation, and sewer deterioration. Various sewer system defects require ongoing attention and effective maintenance.

However, urban water utilities have only recently acknowledged the importance of maintaining comprehensive, current system inventories and failure record databases (Martins et al., 2013). Therefore, historical failure data for most urban water systems remain limited. The identification or development of failure analysis models is of utmost importance in the context of repairable systems, which must be capable of producing outcomes of superior quality despite the constraints arising from the absence of comprehensive failure data.

Ascher and Feingold, as cited in Garmabaki et al. (2016), define repairable systems as those that can be restored to a fully satisfactory level of performance through means other than the complete replacement of the system. Their work focuses on modeling, inferences, misconceptions, and the underlying causes related to repairable systems. In real-world scenarios, an analyst may encounter a set of repairable objects with varying reliability performances. When examining the failure patterns of sewage pumping stations, it is imperative to take into account various reliability processes, including Renewal Processes (RP), Homogeneous Poisson Processes (HPP), and Non-homogeneous Poisson Processes (NHPP) (Xie et al., 2004) . By using these models, one can identify and isolate decreasing and increasing segments of pump station failures, providing critical insights into system reliability and aiding in maintenance and repair decision-making. RP assumes that failures occur randomly and are not correlated with previous events. HPP assumes that failures occur randomly, but the probability of failure is constant over time, while NHPP models take into account changing failure rates over time. Using these models, we can determine the system's risk of failure at different life stages and identify the best time for maintenance and repairs.

A methodical approach can often avoid the incorporation of incorrect assumptions and resolve various challenges in many cases. This strategy should be based on the distinct behavior of data with statistical validity across multiple data sets. The trend behavior of failure data is an effective method for determining the most appropriate statistical procedure for a given data set. Depending on the maintenance strategy, data trends may exhibit different behaviors. For instance, under a perfect repair strategy, failure data lack trends, while they exhibit a monotonic behavior under minimal repair. Several statistical tests have been developed for analyzing trend tests of multiple repairable units. However, the utilization of a univariate trend analysis in isolation does not enable the examination of all possible scenarios. Consequently, a series of analyses should be conducted for each data set to identify the most fitting models. Finally, with a thorough examination of failure patterns in pumping stations and a reliable model of a repairable system, it is possible to make informed decisions regarding system maintenance and repair (Garmabaki et al., 2016).

Through the implementation of maintenance interventions that are prioritized and scheduled based on an analysis of failure trends, urban water utilities can ensure the attainment of optimal system reliability while simultaneously mitigating the potential impact of failures on the environment and public health. This proactive approach to maintenance management helps identify areas for improvement and allows efficient resource allocation. Continuous monitoring and analysis of failure trends facilitate the early identification of system wear and tear, allowing for timely interventions to prevent disruptions. In addition, recognizing and addressing the causes of failure patterns can result in improvements to system design and engineering. The incorporation of sophisticated monitoring technologies and analytical tools can furnish instantaneous insights into the functioning of a system, enabling the identification of emerging patterns and the prediction of potential malfunctions. The adoption of a proactive and data-driven methodology facilitates the creation of maintenance plans that are tailored to specific needs, leading to a decrease in operational expenses and an extension of the lifespan of crucial infrastructure assets.

In essence, the implementation of a systematic methodology for evaluating failure trends in wastewater pumping stations is imperative for optimizing maintenance administration. Through the application of dependable repairable system modeling and trend analysis methodologies, urban water utilities can make well-informed determinations regarding infrastructure investments, maintenance tactics, and the adoption of cutting-edge technologies. This approach ultimately ensures the long-term sustainability of wastewater management systems, mitigating the environmental and public health risks associated with potential failures.

1.2. Problem Statement

The study focuses on analyzing pump failure data and recognizing trends, but faces challenges due to a lack of extensive research and adequate scholarly resources. This absence leads to a limited understanding of pump failure mechanisms, ineffective maintenance strategies, increased operational costs, unexpected system disruptions, and inefficient resource allocation.

The primary focus of this study is relevant to the analysis of pump failure data and the identification of trends in this domain. A prevailing challenge lies in the scarcity of comprehensive research and scholarly articles offering relevant information. This deficiency in scholarly resources consequently leads to an insufficient analysis of interarrival time data, resulting in a limited and generalized understanding of pump failure mechanisms. Without in-depth knowledge of pump failure patterns, it can be difficult to devise effective maintenance strategies. This can lead to increased system downtime and higher operational costs. Besides, limited understanding of failure patterns could result in resources being allocated inefficiently - maintenance efforts may be spent on areas that are less likely to fail while overlooking components that are at a higher risk of failure. Furthermore, lack of in-depth understanding about the nature of pump failures could lead to unexpected system disruptions, impacting the reliability and efficiency of the entire pumping stations.

In addition, the similarities among different instances of pump failure are frequently disregarded as a result of constraints in the analyzed historical data. The exclusion of variables such as weather, temperature, and the age of the pump station from the data sets is a common practice, leading to a limitation in the ability to identify patterns or clusters that exhibit similar attributes. The limited use of comprehensive data mainly results from insufficient visualization techniques that don't clearly highlight trends based on factors like slope and interval.

Furthermore, the verification of identified trends' accuracy and relevance is frequently neglected. Currently, there is a dearth of a framework that integrates reliability engineering principles to validate these visualizations for the purpose of analyzing the trend of pump failure. The process of validation holds significant importance in ascertaining the accuracy of the identified trends pertaining to the occurrences of pump failures.

Additionally, the relationship between pump malfunctions and time is not well-understood, limiting our ability to predict and understand failure patterns. This gap can be addressed by finding an appropriate statistical method to show this relationship.

Ultimately, it is imperative to understand the comprehensive efficiency of the systems being examined. The aforementioned requirement can be fulfilled by determining the ratio of systems exhibiting rising or falling rates of failure, utilizing scrutinized data and the designated model.

Summary of Problem Statement:

- Primary study aim: analyze pump failure data and identify trends.
- Challenge: scarcity of comprehensive research and relevant scholarly articles.

Resultant problems:

- Insufficient analysis of interarrival time data.
- Limited understanding of pump failure mechanisms.
- Ineffective maintenance strategies leading to increased system downtime and costs.
- Inefficient resource allocation.
- Unexpected system disruptions affecting reliability and efficiency.
- Constraints in analyzed historical data often exclude relevant variables (e.g., weather, temperature, pump station age).
- Limitations in visualization techniques to highlight trends effectively.
- Lack of a validation framework for identified trends using reliability engineering principles.
- Inadequate understanding of the relationship between pump malfunctions and time.
- Need for a statistical framework to depict this correlation.

- Emphasis on understanding comprehensive system efficiency.
- Literature review aims to provide insights into these gaps as a foundation for future research.

To shed light on these identified gaps, the following literature review seeks to provide some insights into these issues, serving as a pivotal reference point for further research.

1.3. Literature Review

The present study undertakes a comprehensive examination of the literature on the evaluation of failure analysis of pumps and classifies them into four distinct categories based on their respective methodologies for assessing service. The subsequent explanations furnish a comprehensive summary of every classification, accompanied by instances of pertinent research investigations. Besides, this section presents a collection of statistical models for associated failures. Within this context, the sewer system is considered a repairable system, allowing for the extraction of inter-arrival times between failures based on the recorded failure occurrences. Additionally, this section outlines various data analysis techniques employed in this study with the objective of characterizing the true failure rates.

1.3.1. Integrating Physics-Based and Data-Driven Models for Sewer System Failure Analysis

The present study explores the synergy between physics-based and data-driven models in analyzing sewer system failures.

At the core, physics-oriented models delve into the primary causes of failure, taking into account hydraulic dynamics and operational parameters. These models assess factors like pipe functionality, fluid throughput, and the response of sewage systems to various stressors. For instance, Marquez and Gupta (2006) extensively reviewed models that emphasize hydraulic analysis, offering insights into potential failure points in sewer lines and pumping stations.

On the other hand, the rise of data-driven methodologies leverages statistical analyses and large datasets to predict system failures. These methods often complement the physics-based models by providing a broader perspective and helping in the validation of findings.

A holistic strategy for wastewater infrastructure management, as proposed by Berardi et al. (2008), underscores the amalgamation of hydraulic modeling with data-driven techniques. This integrated approach not only enhances the decision-making process around sewer system performance and maintenance but also augments the precision of predictions, as earlier indicated by works of Bennis et al. (2003) and Ermolin et al. (2002).

Further bridging the gap between these methodologies, Glover and Bhatt (2006) introduced an integrated framework for optimizing process sensing and control based on performance metrics. This strategy, when applied in conjunction with statistical and data-driven methods, holds significant promise for comprehensive sewer infrastructure management and proactive maintenance planning.

1.3.2. Using Historical Data to Model Wastewater System Failures

Several studies have utilized descriptive models to comprehend system behavior concerning sewage pump failure and sanitary sewer overflows by modeling specific random failure events using historical data. Korving et al. (2006) developed a probabilistic model to evaluate the likelihood of sewer overflow occurrences. The model incorporated various factors, including precipitation, blockages, and pump failure. The study conducted by Mutlu et al. (2007). aimed to examine the occurrence of sanitary sewer overflows. To achieve this, the researchers developed a statistical model that incorporated various factors such as the age of the sewer system, pipe materials, and maintenance history. The authors Lee et al. (2006) formulated a probabilistic framework for assessing the likelihood of malfunction in sewage pumping stations. Their approach involved analyzing past records of pump failures and taking into account factors such as the age of the pump, its type, and maintenance history. Le Gat and Eisenbeis (2000) developed a stochastic model in a separate investigation to predict sewer blockages by utilizing past data and considering variables such as pipe diameter, material, and maintenance frequency. Savic et al. (2006) introduced a symbolic data-driven approach based on evolutionary polynomial regression. This method exhibits promise for utilization in wastewater infrastructure systems to construct models and anticipate failure occurrences by leveraging past data.

1.3.3. Analyzing Current Pump Condition and Deterioration Data

This section delves into a review of models assessing the current state and deterioration of sewer pipes and pumps. These models consider a spectrum of factors, encompassing physical attributes (e.g., pipe length, diameter, material, age, slope), operational aspects (e.g., hydraulic status, maintenance approaches, sediment accumulation, blockages), and environmental elements (e.g., waste characteristics, infiltration patterns, soil variety, bedding conditions) (Malek Mohammadi et al., 2019).

Ascher and Hansen (1998) introduced a methodology for evaluating failure data from repairable systems that may be applicable to the analysis of pump condition. This approach could potentially aid in the interpretation of present pump data. It is recommended to derive probability distributions pertaining to time intervals between failures and to identify notable trends in failure rates. The framework was utilized by Jin and Mukherjee (2010) to examine blockage rates, thereby demonstrating the viability of this methodology in the analysis of pump data.

The utilization of the Evolutionary Polynomial Regression (EPR) technique is a viable approach that can be implemented, as suggested by Giustolisi and Savic (2006). The application of current regression models, including linear regression, logistic regression, and multinomial regression, could potentially provide significant value in the examination of present pump data. The models under consideration assume a linear correlation between the factors associated with pump operation and the metrics used to evaluate pump performance. As a result, they can be utilized to provide insight into the present condition of pumps.

Finally, the assessment of pump conditions can be conducted through the utilization of rule-based simulation and artificial neural network (ANN) models, as demonstrated in previous studies (Najafi and Kulandaivel, 2005; Ruwanpura et al., 2004). Through the incorporation of diverse influential elements, these models have the capacity to furnish a comprehensive evaluation of pump conditions.

1.3.4. Stochastic Methods

While probabilistic methods broadly estimate uncertainties using probability theory, stochastic methods further specialize in addressing randomness inherent in processes. Within the infrastructure analysis domain, especially when dealing with unpredictable discrete events such as sewer blockages and overflows, stochastic techniques have proven to be invaluable.

Stochastic arrival processes are often employed in these scenarios. Characterized by events that, while interdependent, manifest at consistent intervals, they serve as a vital tool for predicting system behavior. This principle has been articulated in seminal works by Karlin and Taylor (1981) and KINGMAN (1993).

Defining further, a stochastic process comprises a series of random variables denoting the system state at varying time points, as detailed by Meester and Shanthikumar (1993). In the context of infrastructure systems, these variables can signify operational or non-operational network states. Disruptive events, like blockages, can transition a system to a failed state, while interventions such as repairs can restore functionality.

However, real-world data may not always reflect purely random patterns. This can result in overcautious reliability estimates, as observed by Chu and Durango-Cohen (2008). Given that consistent failure rates might not always be observed in practical scenarios, alternative distribution models, including Weibull and gamma distributions, find relevance. Notably, the Weibull distribution has been recognized for capturing interfailure durations with varying rates and intervals, as highlighted by Korving et al. (2006).

1.3.5. Repairable systems

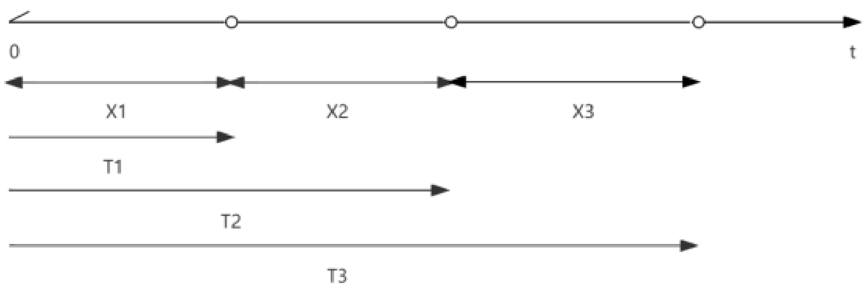


Figure 1.1: Arrival times, Ti, and inter-arrival times, Xi.

Reliability engineering involves analyzing failure patterns to gain insights into the behavior of a system. In real-world situations, determinations regarding failure patterns necessitate dependence on empirical failure data and statistical methodologies. The objective of this study is to investigate techniques for trend analysis in failure data derived from repairable systems. Figure 1.1 illustrates a standard failure process of a solitary repairable system that is initiated at time $t = 0$. The temporal instances of failure, denoted as T1, T2, and so forth, are frequently referred to as arrival times within the literature. In contrast, the time durations between consecutive failures, represented as X1, X2, and so on, are commonly referred to as inter-arrival times. As illustrated in Figure 1.2, each inter-arrival time can be seen as a combination of diagnose time, repair time, and operational time. Failures are regarded as discrete incidents that take place at particular moments in time. The temporal duration between successive occurrences of malfunctions is subject to the influence of multiple factors, including but not limited to the characteristics of the sewage infrastructure, the age of the pumping equipment, and the control mechanisms regulating the pumps. Conversely, the duration of system unavailability due to malfunctions is predominantly influenced by the chosen method of restoration and various managerial considerations, such as workforce accessibility and contingency resources (Korving et al., 2006).

If the analysis of inter-arrival times between failures displays a distinct trend, it can offer valuable insights into a system's reliability and pinpoint areas that require enhancement. Such a trend would suggest that the inter-arrival times are not uniformly distributed, potentially having a significant impact on the system's performance. Evaluating the statistical significance of these variations requires the use of various statistical methodologies, including time series analysis, trend analysis, and statistical process control charts. These techniques facilitate the assessment of the system's stability and the detection of potential patterns. Monotonic trends in failure patterns may indicate either an improvement or a decline in the system's reliability. A rising pattern in inter-arrival intervals suggests an improvement in the system, while a falling pattern signifies a decline. Early detection of such trends is crucial to implement corrective actions and prevent system failure. In addition to monotonic trends, failure

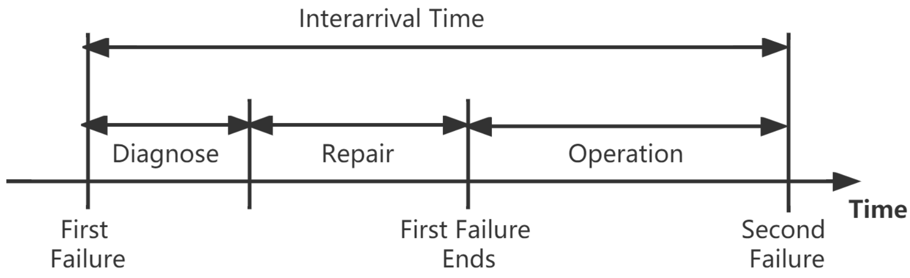


Figure 1.2: Composition of Interarrival Time

patterns can also exhibit non-monotonic trends, such as cyclic and bathtub trends. In detailed, non-monotonic trends describe failure patterns that vary in a non-linear and inconsistent manner over time. These trends can be complex, often involving multiple factors and stages. One type, the cyclic trend, exhibits failures at predictable intervals. For instance, a machine might fail periodically due to operational cycles, or certain equipment might show increased failures seasonally, like air conditioning units failing more in summer months. Another prevalent non-monotonic trend is the bathtub curve, which captures three distinct phases: the infant mortality phase, where new products may have a high failure rate due to defects; a steady-state phase where the failure rate is relatively constant and at its lowest; and the wear-out phase, where older products or components see a rise in failures due to aging and prolonged usage. Tests based on TTT methodology for detecting trends in data from repairable systems have been conducted by authors such as Kvaløy and Lindqvist (1998). It is important to note that this study focuses solely on monotonic trends in failure patterns. Non-monotonic trends, such as cyclic and bathtub trends, are not considered. By limiting the analysis to monotonic trends, researchers can more accurately evaluate the direction and statistical significance of changes in the inter-arrival times of failures, and identify possible improvements or deteriorations in the system's reliability.

In many reliability analyses, it is commonly assumed that a component or system failure triggers a renewal process, resulting in a state of restored functionality comparable to a newly manufactured item. In such cases, failures are considered to be independent and identically distributed, allowing the use of standard failure interval histograms and traditional statistical distribution fitting techniques. This enables the optimization of maintenance strategies.

However, numerous maintenance scenarios involve repairable systems that do not adhere to the renewal assumption. These systems undergo limited maintenance, leading to a "bad-as-old" state and a consequent decline in reliability, ultimately resulting in increased rates of failure over time. In these situations, traditional statistical distribution fitting is inappropriate due to the non-identically distributed and non-independent nature of successive failures.

Applying Homogenous Poisson Process and Renewal Process Models

To capture the stochastic behavior of diverse systems, the Homogeneous Poisson Process (HPP) and Renewal Process (RP) models are employed. The application of these methodologies involves understanding event frequency, paving the way for an in-depth exploration of system mechanics.

The HPP model is applied due to its feature of complete memorylessness, implying that the likelihood of an event occurrence doesn't rely on preceding system incidents or its physical condition. As per Crow (1975)'s assertion, the system neither improves nor deteriorates over time, but rather maintains a constant level of failure. This approach assumes a consistent failure rate, negating any aging effects on the system. The parameter λ denotes the average event occurrence rate, which is fundamental in interpreting and modeling the behavior of repairable systems. This rate serves as an indicator of the mean system failure occurrences. Events within a given time interval t are designated as Nt and modeled via the Poisson distribution with a mean value of λt.

The mathematical formulation of this Poisson distribution is as follows:

$$P(Nt = r) = \frac{(\lambda t)^r \cdot e^{-\lambda t}}{r!} \quad (r = 0, 1, 2, \ldots) \tag{1.1}$$

Inter-event intervals, represented as $X1$, $X2$, etc., are considered independent random variables that follow an exponential distribution.

Another crucial step in the methodology involves focusing on the time, T, between a chosen temporal reference point and the subsequent system event. By analyzing these intervals, significant insights into the temporal distribution of failures are obtained. To this end, the probability density function (PDF) and cumulative distribution function (CDF) of T, as defined by Cox and Lewis (1966), are used:

$$F_T(T) = 1 - e^{-\lambda T} \quad (T \geq 0) \tag{1.2}$$

$$f_T(T) = \lambda e^{-\lambda T} \quad (T \geq 0) \tag{1.3}$$

In evaluating the temporal behavior of repairable systems, the point of reference for T, the elapsed time, is a moment arbitrarily chosen post the previous failure. However, conventionally, it's set as the immediate moment following the last malfunction for ease of understanding. The representation of intervals between occurrences is done using $T1$, $T2$, etc., which are independent random variables and adhere to an exponential distribution as depicted by the probability density function (PDF), denoted as Equation 2.3.

The process of analyzing repairable systems significantly benefits from taking into account the time series of inter-arrival times. Ignoring this critical aspect might lead to inaccurate conclusions. For instance, if the data seems to fit an exponential distribution, it may misleadingly suggest that the system is aptly represented by a Homogeneous Poisson Process (HPP). But this may not hold true if the data possesses non-stationarity.

To enhance the precision of the modeling and provide a trustworthy depiction of the system's dynamics, implementing an apt trend test before fitting a distribution model for the inter-arrival times is highly recommended. This step enables the inclusion of any non-stationary elements in the data, which further leads to the selection of a more accurate model that mirrors the true behavior of the system. Enhanced modeling strategies lead to informed decision-making and efficient maintenance techniques for repairable systems.

While the simplicity of the HPP model is commendable, it might not fully encapsulate the intricacies of real-world systems. In such scenarios, employing a more comprehensive Renewal Process (RP) model could be beneficial. The RP model's ability to handle a variety of inter-arrival time distributions, including Weibull and gamma distributions, provides a more versatile approach to analyze event occurrences. Especially useful for instances where the frequency of events begins low and gradually rises over time, followed by a process reset after a specific event, the RP model offers a broader scope to illustrate different failure patterns.

The Weibull distribution, due to its adaptability, has found applications across a multitude of scenarios, including representing product lifetimes and strength distribution of certain materials. Its ability to take on various shapes and empirical compatibility with diverse data types have bolstered its popularity among engineers, a trend significantly attributed to Waloddi Weibull's contributions in 1951. In systems operating on a "weakest link" framework, where the unit failure coincides with the first component failure, the Weibull distribution serves as a fitting statistical model. An example of this can be seen in a capacitor, where its lifespan is primarily determined by the durability of the least robust component of its dielectric. The Weibull distribution provides the probability density function (PDF) and cumulative distribution function (CDF) for inter-arrival times.

The mathematical expressions for the Weibull distribution's PDF and CDF are given as follows:

$$f_T(T) = \frac{k}{\lambda}\left(\frac{T}{\lambda}\right)^{k-1} e^{-\left(\frac{T}{\lambda}\right)^k} \quad (T \geq 0) \tag{1.4}$$

$$F_T(T) = 1 - e^{-\left(\frac{T}{\lambda}\right)^k} \quad (T \geq 0) \tag{1.5}$$

Here, $\lambda > 0$ is the scale parameter, and $k > 0$ is the shape parameter. The Weibull distribution reduces to the exponential distribution when $k = 1$, making it a more general model that can represent a wide range of failure patterns.

Another notable distribution often used in RP models is the gamma distribution, defined by its PDF and CDF as follows:

$$f_T(T) = \frac{T^{k-1} e^{-\frac{T}{\theta}}}{\theta^k \Gamma(k)} \quad (T \geq 0) \tag{1.6}$$

$$F_T(T) = \frac{1}{\Gamma(k)} \int_0^{\frac{T}{\theta}} x^{k-1} e^{-x} dx \quad (T \geq 0) \tag{1.7}$$

In this case, $\theta > 0$ is the scale parameter, $k > 0$ is the shape parameter, and $\Gamma(k)$ is the gamma function.

Prior to the implementation of said models, it is imperative to conduct suitable trend detection examinations in order to prevent erroneous inferences from being drawn from the data. Section 3.3 will discuss the Laplace test as a means of determining the validity of the HPP model.

To summarize, the Homogeneous Poisson Process and Renewal Process models are robust statistical methodologies that can be utilized to examine the incidence of events in various systems.

Non-homogenous Poisson process models

The Non-homogenous Poisson Process (NHPP) models present a more advanced and adaptable framework for delineating the variability and patterns in a system's failure intensity. NHPP, in contrast to the Homogeneous Poisson Process (HPP), incorporates a time-varying mean failure rate, stepping beyond the constant mean failure rate assumed in HPP. This capacity to adjust enables a more effective capture and analysis of the dynamic behavior of repairable systems, yielding accurate predictions and providing actionable insights for maintenance planning and system improvement.

One instance of an NHPP is the Crow's model, also known as the PLP model, wherein the intensity function follows a power law Crow, 1975:

$$\lambda(t) = b_0 b_1 t^{b-1} \quad b_0 > 0, b_1 > 0 \tag{1.8}$$

In this formulation, $\lambda(t)$ symbolizes the time-dependent mean failure rate, b_0 is a scale parameter, and b_1 serves as a growth parameter. When $b_1 > 1$, the failure rate increases over time, signifying system deterioration. Conversely, when $b_1 < 1$, the failure rate decreases, indicating system improvement. The Maximum Likelihood (ML) estimators for the parameters α and β in the Crow model are expressed as:

$$\hat{\beta} = \frac{n}{n \log(t) - \sum(\log(t_i))} \tag{1.9}$$

$$\hat{\alpha} = \frac{n}{t\hat{\beta}} \tag{1.10}$$

Here, $\hat{\beta}$ signifies the ML estimator of the growth parameter β; n represents the total number of observations; t denotes the end of the observation period; t_i refers to the time of the i^{th} failure; and $\hat{\alpha}$ is the ML estimator of the scale parameter α.

On the other hand, Cox and Lewis put forth another NHPP model called the log-linear model, originally developed to portray improving systems:

$$\lambda(t) = \exp(b_0 + b_1 t) \tag{1.11}$$

Here, b_0 is the scale parameter, and b_1 is the growth parameter. For $b_1 > 0$, the failure rate increases over time, while for $b_1 < 0$, the failure rate decreases, suggesting system enhancement. The parameters of both models can be estimated using maximum likelihood techniques, rendering them applicable to real-world situations. To account for more complex system behavior, this model has been extended to include quadratic and higher-degree polynomial elements. The Maximum Likelihood (ML) estimator for parameter β_1 in the Cox-Lewis model can be determined by solving:

$$\sum \left[t_i - \left(\frac{n}{\hat{\beta}_1} \right) \right] \left[\frac{(\exp(\hat{\beta}_1 t) - 1)}{\exp(\hat{\beta}_1 t)} \right] = 0 \tag{1.12}$$

In this context, $\hat{\beta}_1$ represents the ML estimator for the growth parameter β_1; n denotes the total number of observations; t indicates the conclusion of the observation period; and t_i refers to the occurrence of the i^{th} observed failure. Following this, β_0 can be ascertained using:

$$\hat{\beta}_0 = \log \left[\frac{n}{\sum(\exp(\hat{\beta}_1 t) - 1)} \right] \tag{1.13}$$

Here, $\hat{\beta}_0$ signifies the ML estimator of the scale parameter β_0.

By adopting the above methodologies, it is possible to effectively implement the NHPP models, either the Crow's or Cox-Lewis, for any given repairable system. These methodologies offer comprehensive approaches to modeling the failure intensity of repairable systems. It's crucial to remember, however, that accurate parameter estimation and selection of the most fitting model is key to yielding the most insightful and actionable results.

Branching Poisson Process Models

The Branching Poisson Process (BPP) is a captivating stochastic process that emerges as a natural extension to traditional Poisson processes. In essence, a BPP describes a system where each event can independently give rise to several secondary events, or ”offspring.” This is in sharp contrast to traditional Poisson processes where events take place independently of one another.

Cox Process

In recent studies on system failure analysis, the Cox process has been recognized as an innovative tool to address traditional methodologies' limitations, especially when confronting rapid fluctuations in failure rates. The uniqueness of the Cox process lies in its capability to account for failures that might cluster or occur periodically, positioning it as a more apt analytical tool for dynamic systems.

Complementing the advancements of the Cox process is the integration of Gaussian Processes (GPs). Revered as a potent machine learning tool, GPs excel in processing intricate, non-linear relationships within datasets. Their non-parametric and probabilistic nature not only encapsulates uncertainties but also offers a smoother approximation of underlying functions, making them invaluable for modeling time series data with variable intensity (Schulz et al., 2018). Libraries like GPy have amplified the benefits of GPs. Their user-centric interfaces, combined with robust modeling capabilities, provide a streamlined approach to GP modeling. Within such libraries, the prominence of the radial basis function (RBF) kernel stands out, given its adeptness at capturing complex, non-linear data correlations.

The world of modeling is always changing, and lately, there's been a lot of focus on predicting the intensity function to understand system patterns better. By creating new sets of data values, current methods can guess this function with more accuracy. The intensity function's gradient, as calculated through today's techniques, offers a glimpse into system patterns. Such patterns, evident in the function's trends, shed light on the intervals between system failures, paving the way for more targeted interventions.

1.3.6. Analytical Trend Test and Underlying Concepts

The utilization of the analytical trend test is imperative in statistical analysis, particularly in the investigation of systems' behavior. Several statistical tests have been devised to evaluate the existence of trends and underlying patterns in the data, including the Laplace Trend Test, Military Handbook Test, and Mann Test. These assessments aid in ascertaining whether the observed conduct adheres to a particular pattern or diverges from the null hypothesis. The present section explores the theoretical underpinnings and practical implementation of trend tests, furnishing a thorough comprehension of their importance in scrutinizing the behavior of system failures.

Laplace Trend Test

The Laplace Trend Test is a statistical technique utilized to examine the null hypothesis (H0) positing that a system adheres to a homogeneous Poisson process (HPP), in contrast to the alternative hypothesis (H1) suggesting the system follows a non-homogeneous Poisson process (NHPP) characterized by a monotonic intensity function (Kvaløy & Lindqvist, 1998). This test is particularly pertinent in assessing the evolution of failure rates and performance patterns in repairable systems.

The methodology of the Laplace Trend Test involves contrasting the arithmetic mean of failure times with the median point of the observation interval. In the event that a process is scrutinized within the time interval $(a, b]$, and h is delineated as:

$$h = \begin{cases} h & \text{if the process is time truncated,} \\ h-1 & \text{if the process is failure truncated} \end{cases} \tag{1.14}$$

the test statistic L can be computed accordingly:

$$L = \frac{T_{\text{mean}} - \frac{1}{2} \cdot h \cdot (b+a)}{\sqrt{h \cdot (b-a)^2/12}} \tag{1.15}$$

Under the presumption of the null hypothesis (HPP), the test statistic L adheres to an asymptotically standard normal distribution. This approximation to the normal distribution is remarkably accurate, with a general guideline indicating that a sample size of $n \geq 3$ is sufficient (Kvaløy & Lindqvist, 1998). The Laplace Trend Test is optimally designed for the null hypothesis (HPP) when confronted with an NHPP alternative characterized by a log-linear intensity function, assuming the precise null distribution of L is employed.

At the core of the Laplace Trend Test lies the principle that, under the null hypothesis of an HPP, the order statistics $(T_1, ..., T_n)$ are uniformly distributed over the interval $(a, b]$. As a consequence, the mean of these statistics possesses an expected value of $h \cdot (b+a)/2$ and a variance of $h \cdot (b-a)^2/12$. The test statistic L offers insights into the direction of the trend: a negative L value signifies a diminishing trend, whereas a positive L value indicates an ascending trend.

The Laplace Trend Test's capacity to discern between HPP and NHPP processes, as well as identify the direction of trends, renders it an invaluable tool for reliability analysts and engineers. By leveraging this test, analysts can gain insights into system performance, identify periods of increased or decreased reliability, and implement targeted maintenance strategies or system improvements accordingly.

In summary, the Laplace Trend Test offers a method for examining the null hypothesis of a homogeneous Poisson process against an alternative hypothesis of a non-homogeneous Poisson process with a monotonic intensity function. This test is particularly relevant for assessing failure rates and performance patterns in repairable systems, allowing analysts to determine the direction of trends and make informed decisions based on system performance data. With its ability to distinguish between HPP and NHPP processes, as well as ascertain the direction of trends, the Laplace Trend Test constitutes an indispensable asset for reliability analysts and engineers seeking to optimize system performance and reliability.

Military Handbook Test

The Military Handbook Test is a widely-used statistical method for evaluating system performance and reliability. It is particularly suitable for equipment operating continuously under diverse environmental conditions. The primary goal of this test is to identify periods of increased or decreased system reliability by detecting changes in failure rates over time (Rausand & Hoyland, 2003).

This test compares the null hypothesis of a Homogeneous Poisson Process (HPP) against the alternative hypothesis of a Non-Homogeneous Poisson Process (NHPP) with a monotonic trend. The test statistic Z is calculated as:

$$Z = 2 \cdot \sum_{i=1}^{k} \left(\frac{\ln(\tau)}{t_i} \right) \tag{1.16}$$

Here, k is equal to n (number of failures) for time-truncated data ($\tau = T$) and $(n-1)$ for failure-truncated data ($\tau = t_n$). Low values of Z indicate system deterioration, while high values suggest system performance improvement.

The null hypothesis (H_{02}) is rejected based on the following criteria:

$$Z = \begin{cases} Z < \chi^2_{2n,1-\alpha/2} \text{ or } Z > \chi^2_{2n,\alpha/2} & \text{for time-truncated data} \\ Z < \chi^2_{2(n-1),1-\alpha/2} \text{ or } Z > \chi^2_{2(n-1),\alpha/2} & \text{for failure-truncated data} \end{cases} \tag{1.17}$$

The Military Handbook Test is optimally designed for the power law intensity function, providing a valuable tool for assessing system performance and reliability by detecting monotonic trends in failure data. This test is particularly suited for equipment operating under continuous and varying conditions. It allows analysts to identify periods of increased or decreased system reliability and make informed decisions regarding maintenance and system improvements. The test's ability to differentiate between HPP and NHPP processes and determine the direction of trends makes it a crucial tool for reliability analysts and engineers.

Kendall-Mann Test

The Kendall-Mann Test, also known as the Kendall Rank Correlation Test, is a non-parametric statistical method designed to assess the null hypothesis (H03) of no association between two variables against an alternative hypothesis of a monotonic relationship (Hamed & Rao, 1998).

The Kendall-Mann Test is calculated by counting the concordant and discordant pairs among the paired observations of two variables. The test statistic τ is computed as follows:

$$\tau = \frac{(C-D)}{\sqrt{(C+D)\cdot(C+D-n)}} \tag{1.18}$$

where C and D represent the number of concordant and discordant pairs, respectively, and n is the total number of paired observations.

Under the null hypothesis (H03), the Kendall-Mann Test statistic τ is approximately distributed as a standard normal distribution. The null hypothesis is rejected at the α% level if $|\tau| > z_{\alpha/2}$.

The Kendall-Mann Test provides a robust and reliable method for detecting trends in system performance and failure data. By evaluating the null hypothesis of no association against a monotonic relationship, this test allows analysts to determine if the system's performance is improving, deteriorating, or remaining constant over time.

Furthermore, the non-parametric characteristic of the Kendall-Mann Test renders it highly appropriate for scrutinizing data that exhibit non-normal distributions or limited sample sizes. The test's flexibility allows for its application in various industries and contexts, including but not limited to manufacturing, engineering, finance, and healthcare.

Apart from its flexibility, the Kendall-Mann Test presents various other benefits in comparison to conventional parametric approaches. An advantage of this method is its lack of reliance on a predetermined probability distribution, rendering it less susceptible to the influence of extreme values and other irregularities in the data. In addition, the Kendall-Mann Test is relatively simple to calculate and interpret, making it accessible to professionals with varying degrees of statistical expertise.

1.3.7. Literature Review Conclusion

In conclusion, while the literature on the topic remains limited, existing studies indicate that stochastic methods can offer valuable insights into pump failure analysis. These methods, though not extensively covered, appear promising in guiding decision-makers, maintenance planners, and researchers. Korving et al. (2006), with its multiple benefits, is essential in shaping the trajectory of my research. The methodology presented by Korving et al. offers a well-structured foundation for this research, which allows for an extension of their findings and an adaption of the approach to meet unique challenges and conditions identified in the study. Moreover, the classification of pump failures into thermal failures and a combined category of mechanical, electrical, and safety fuse failures, as proposed by Korving et al., facilitates a comprehensive understanding of the diverse factors that can lead to pump malfunctions. Therefore, research conducted by Korving et al. (2006) provides significant contextual information for my study, as it pertains to Rotterdam and covers a distinct time frame within the same geographical area. This allows for a broad understanding of the city's infrastructure systems over time.

Table 1.1: Comparison of methodologies for sewer infrastructure system evaluation.

Methodology	**Characteristics**
Physics-Based Models	- Emphasizes on physics-oriented models analyzing reasons behind failure occurrences from hydraulic and operational variables. - Models often examine pipe functionality, fluid throughput, and sewer network reactions to different stressors.
Utilization of Historical Data	- Descriptive models using historical data to comprehend system behavior. - Addresses specific random failure events and integrates various factors like precipitation, blockages, and pump failures.
Current Pump Condition Analysis	- Focuses on evaluating the current condition of sewer pumps and pipes. - Considers variables like physical parameters, operational factors, and environmental influences.
Stochastic Methods	- Essential for assessing infrastructure breakdowns from unpredictable events. - Uses stochastic arrival processes marked by events occurring independently at consistent rates. - Also considers alternative distribution models.

1.4. Research Motivation

The purpose of this research is to delve into the existing pump failure data to equip state and municipal agencies' decision-makers with knowledge and strategies for improving maintenance practices in sewage pumping stations. The ambition is to design an innovative framework for selecting pump failure models in systems that can be repaired, utilizing a five-year dataset (2016-2020) from the Department of Public Works of Rotterdam. By using statistical techniques, this study seeks to discern trends in the intervals between pump failures, aiming to determine whether these trends are ascending, descending, or stable. This in-depth analysis is set to significantly bolster the ability to make informed decisions about maintenance prioritization, subsequently optimizing sewer system management. Furthermore, the development and application of a data-reliant, comprehensive framework is expected to contribute to the creation of precise maintenance strategies, enhancing the sustainability and resilience of urban water infrastructure. This research motivation forms the cornerstone of the thesis's discourse.

1.5. Research Question

The primary aim of this investigation is to evaluate the performance of pump systems by understanding the nature, patterns, and changes in pump failures over time. The main research question driving this study is:

How has the performance of pumps evolved over a given period, such as 5 years or a specific failure duration, and do the patterns of failures in current data align or deviate from historical findings, especially when compared to the data from Korving et al. from 20 years earlier?

To dissect and address these overarching questions in depth, the study delves into the following sub-questions:

(1)What are the distinct characteristics present in the failure data?

(2)What defines a failure, and which modes of failure are the most significant?

(3)What are the various types of failures that can be classified?

(4)Is it possible to identify any changes in failure patterns over a period of time, and what standards should be employed to evaluate them? Are there any identifiable trends in the failure data?

(5)What is the most appropriate statistical model to represent the data, and what methods can be utilized to estimate the parameters of the model?

1.6. Research Objective

In line with the guiding questions of this study, I am dedicated to crafting an analytical framework grounded in a rigorous examination of pump failure patterns and system dynamics. This structured approach has been formulated with the following objectives in mind:

(1) Identify Distinct Characteristics in the Failure Data: To unpack the intricacies of pump failure datasets, revealing nuances that might otherwise go unnoticed.

(2) Define and Identify Prevalent Failure Modes: To provide an understanding of the most common and impactful system failures.

(3) Classify Different Types of Failures: To systematize my findings and present them in a structured, comprehensible manner.

(4) Detect Shifts in Failure Patterns Over Time: To offer insights into how pump systems evolve and how their reliability might be impacted over prolonged periods.

(5) Determine the Most Appropriate Statistical Model for Data Representation: To ensure that my analysis is grounded in fitting statistical methods that truly capture the essence of the data.

The dominant aim, through the application of this framework, is not just to analyze failure trends using suitable statistical models but also to enrich my comprehension of pump system efficiency. Through this endeavor, I anticipate my findings will substantially augment the existing body of knowledge, offering invaluable insights for both the academic community and practitioners in the field.

2 Methodology

This section presents a collection of statistical models for associated failures. Within this context, the sewer system is considered a repairable system, allowing for the extraction of inter-arrival times between failures based on the recorded failure occurrences. Additionally, this section outlines various data analysis techniques employed in this study with the objective of characterizing the true failure rates.

2.1. Data & Failure Analysis

The present study examines data derived from several log files, each of which pertains to a discrete year spanning the period from 2016 to 2020. The process of analysis entails the execution of a sequence of tasks aimed at processing and scrutinizing the data, thereby facilitating the derivation of significant insights concerning pump failures. In order to furnish a more comprehensive elucidation, we shall delve into the particulars of each individual step.

The first step involves importing the data files for each year, which are then combined into a unified dataset. The dataset has been subjected to a cleaning and organization process that involves the formatting of datetime columns to ensure uniformity and facilitate ease of manipulation. Furthermore, the pertinent columns have been rebranded to enhance legibility and comprehension. The present study involves the identification and subsequent categorization of specific error types, including pressure pipe failure, software failure, and basin failure, among others. These identified error types are then compiled and preserved in a distinct dataset for further analysis. Commonly, the analysis of pump failure focuses on two primary categories of issues: $general$ and $thermal$ issues.

$General$ problems encompass a variety of challenges. Specifically, these include mechanical wear, seal failures, and bearing damage. Such issues can culminate in reduced efficiency, heightened energy consumption, and, in extreme cases, pump failure. This categorization draws from a broad understanding of common issues, but in the context of our study, especially with the Rotterdam dataset, these problems are paramount.

On the other hand, $thermal$ problems pertain to issues of temperature regulation and heat management in the system. These problems may cause overheating, leading to damage in critical components, a shortened pump lifespan, and potential system shutdowns.

These issues are deemed significant due to various reasons. Simultaneous examination of these issues is imperative in order to obtain a comprehensive comprehension of pump failure patterns. The justification for adopting this methodology is explicated in the following section:

(1) Prevalence: General and thermal problems are among the most frequently occurring failures in pump systems. Analyzing these issues together ensures that the study addresses a significant portion of the total failure incidents, providing a representative overview of the challenges faced by pump operators and maintenance teams.

(2) Impact on Pump Performance: Both general and thermal problems can severely affect the performance and efficiency of pump systems.

(3) Interconnected Causes and Effects: General and thermal problems often share common root causes, and their effects can be mutually reinforcing. For example, a general problem like bearing damage could lead to increased friction, which then generates excessive heat and contributes to thermal problems (Tuzson, 2000). By analyzing these issues together, a more holistic understanding of the relationships between different failure types and their root causes can be achieved, enabling the development of more effective preventive maintenance strategies and system improvements.

(4) Cost Implications: Pump failures caused by general and thermal problems can have significant financial consequences, as they may necessitate expensive repairs or even the complete replacement of affected equipment. Identifying patterns and trends in these failure types allows operators to develop targeted maintenance and monitoring plans to minimize downtime and reduce the overall costs associated with pump system failures (Lobanoff & Ross, 2013).

Next, the analysis iterates through unique pump stations and pump numbers, generating a new dataset for each combination. For every pump station and pump number combination, it is guaranteed that only one error occurs per time period. This is achieved by adjusting the end date of a failure if it overlaps with the start date of the subsequent failure. This ensured that no more than a single failure event occurred in a given unit of time. The duration of each failure is calculated by subtracting the start date from the end date, providing essential information on the severity and impact of each failure. For the selection of an appropriate age unit and time scale, the existing literature is referred to, specifically Zacks (Zacks, 2012).

Furthermore, the interarrival times between failures are computed by taking the difference between the start dates of consecutive failures. This information is crucial in understanding the distribution and frequency of failures, allowing for the identification of patterns and trends that can inform future maintenance strategies and system improvements. However, the analysis of pump failure data can pose a challenge owing to the possibility of correlated observations. To effectively describe the interarrival times among these failures using a continuous-time distribution like Poisson, certain modifications were necessary to guarantee that only one arrival transpired within a given unit of time increment. Due to the presence of multiple failures within a single day in the data set, it was not feasible to measure the interarrival times in terms of days. Conversely, the hour was selected as the minimum unit of time. This premise was considered to be more rational than aggregating all instances of failure occurring on a particular day into a singular event (Jin & Mukherjee, 2010).

In addition, upon completion of processing the data for each distinct combination of pump station and pump number, any redundant or incomplete datasets are eliminated from the dataset inventory.This ensures that the analysis is performed only on unique and non-empty datasets, maintaining the integrity of the study and preventing any unnecessary duplication or erroneous conclusions.

2.2. Trend Analysis

In the empirical assessment of pump failure trends, a recurring observation underscores the inadequacy of representing a pump's failure trajectory through a single trend line. Consider the following data for station 566 pump 2 from 2016 to 2020 as an example. While broadly informative, the global trend line falls short of reflecting detailed, micro-level changes. This representational flaw not only threatens misinterpretation of the pump's failure dynamics, but it also parallels the analytical flaws identified in Korving et al. (2006). In the referenced study, the failure patterns of each pumping station were equally characterised by a single trend line—a methodological limitation that could lead to analytical imprecision. Consequently, it necessitates the exploration of segmenting failure data. Through the use of segmented analysis and subsequent reliability modeling employing techniques such as NHPP, RP, or HPP, one anticipates a more rigorous and nuanced understanding, thereby augmenting the reliability of the derived insights.

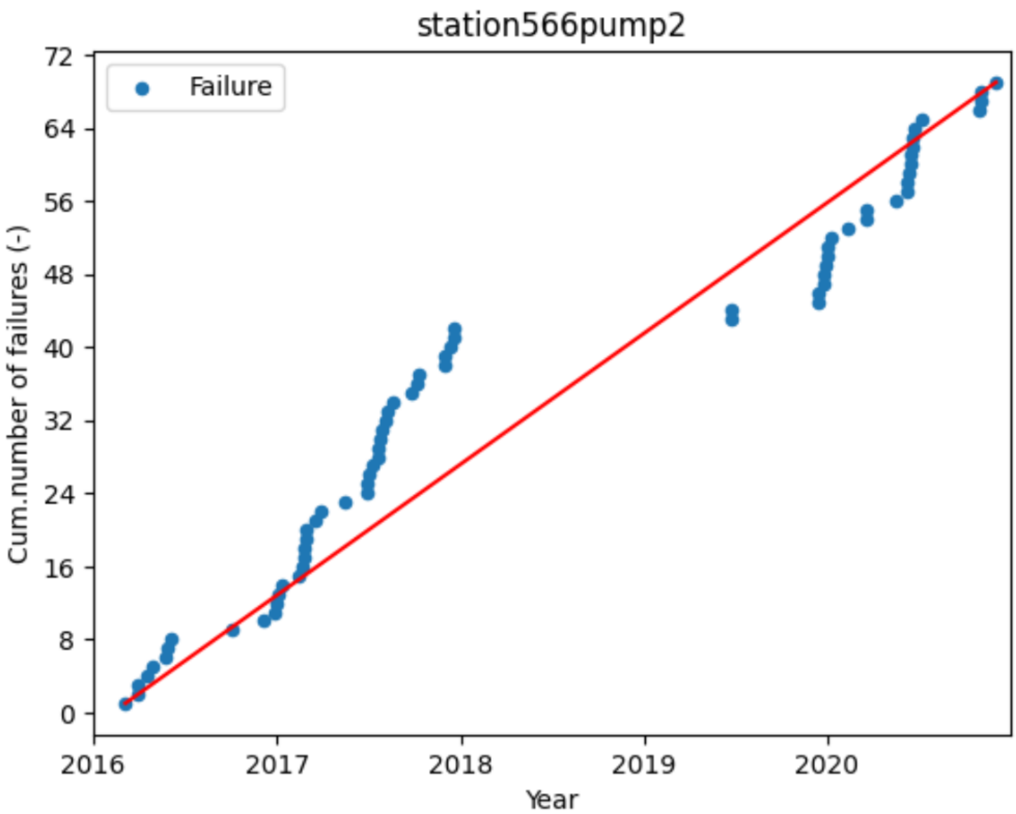


Figure 2.1: Station 566 Pump 2 with Global Trend Line

Expanding upon my prior methodological insights, an enhanced perspective on system reliability becomes imperative. This approach emphasizes integrating both subjective expertise and objective data-driven methods. Especially relevant for wastewater pumping stations, this nuanced method is essential. Failures in these stations extend beyond mere technical glitches; they have broader consequences for public health and the environment. Thus, the focus should not only be on identifying these irregularities but also on comprehensively understanding their nuances. This ensures that maintenance is forward-thinking, effectively preempting and addressing potential risks.

Two main factors dictate the categorization of trend groups for a pumping station: time between consecutive failures and significant shifts in failure frequency. Extended "interarrival times" (durations between failures) may suggest a station's prolonged inactivity, questioning the need for a continuous trend analysis. Meanwhile, drastic changes in failure rate, evident from the cumulative failure-event time graph, signal potential maintenance strategy reassessments.

In addition to these factors, other variables potentially impacting trend analysis should be considered, including seasonal variations, changes in operating conditions, and the introduction of new maintenance methods. Incorporating these additional elements into the analysis can lead to a more comprehensive understanding of system effectiveness and improve the accuracy of trend group identification.

Subjectively determining trend groups using expert judgment and domain knowledge can provide important context and insights into the pumping station's behavior. However, this method is vulnerable to biases and inconsistencies arising from subjective human judgment. Objective determination, in contrast, uses statistical techniques and algorithms to identify patterns and clusters in the data. Combining subjective and objective methods allows us to benefit from each approach's strengths, leading to a more robust and reliable evaluation of the pumping station's malfunctioning patterns.

To operationalize this integrated approach, a systematic framework is required to facilitate the combination of subjective and objective data streams. The proposed framework should include guidelines for systematically collecting, examining, and understanding failure data, along with criteria for recognizing and validating trend clusters. Additionally, the framework should encourage interdisciplinary collaboration and effective communication among domain experts, data analysts, and maintenance planners. This ensures the analysis benefits from a broad range of perspectives and expertise.

The integration of subjective and objective trend analysis can be realized through a combination of optimization techniques and linear modeling, such as differential evolution and piecewise linear fit, respectively. These algorithms can be trained to identify patterns and trends in the data, considering both quantitative metrics such as interval time and failure rate, and qualitative insights from domain experts. Through iterative refinement of the algorithm's performance via feedback and validation, a more accurate and reliable model for identifying trend groups in the pumping station's failure data can be developed.

2.2.1. Objective trend analysis with Piecewise Linear Fit

The implementation of the piecewise linear regression model in objective trend analysis involves the utilization of a Python library known as 'pwlf' (Piecewise Linear Fit). The purpose of this library is to enable the user to perform piecewise linear function fitting on data points, wherein the user is required to indicate the positions of breakpoints. The theoretical foundation and computational methodology of this library are based on the development of a problem of fitting a piecewise linear function with least squares, which has been extensively discussed in Jekel and Venter (2019) and other relevant scholarly works. The problem of fitting a piecewise linear function through a set of observed data points can be formulated as a least squares optimization problem. Specifically, the objective is to minimize the sum of the squared differences between the observed data points and the piecewise linear function, while keeping the breakpoints of the function fixed as a constraint. The mathematical formulation of this optimization problem is as follows:

Consider n to be the number of data points, y_i and x_i the observed data points, and $f(x)$ the piecewise linear function. The sum of squared residuals can be minimized using the following equation:

$$\text{minimize} \quad \sum_{i=1}^{n}(y_i - f(x_i))^2 \tag{2.1}$$

In this formulation, the term $(y_i - f(x_i))^2$ represents the square of the residual for the i-th data point. The aim is to find the function $f(x)$ that minimizes the sum of these squared residuals. This mathematical representation facilitates an understanding of the underlying mechanism in the 'pwlf' library and how it conducts piecewise linear regression modeling.

Given knowledge of the locations of the breakpoints, it is possible to solve the least squares problem through the application of linear algebra techniques, as expounded upon in several sources, including the aforementioned literature. The task becomes more arduous when the locations of the breakpoints are uncertain, as it necessitates the joint optimization of both the parameters of the linear models and the breakpoint locations.

To address this challenge, the 'pwlf' library employs a global optimization method known as differential evolution, which is readily accessible in the SciPy library. Differential evolution is an optimization algorithm that operates on a population of candidate solutions, using the concepts of mutation, crossover, and selection to explore the optimal solution in a continuous parameter space. The algorithm is particularly adept at handling optimization problems that are non-linear and multi-dimensional in nature (Price, 2013). As such, it is deemed to be an optimal choice for the current piecewise linear regression problem. The algorithm's efficacy in exploring the optimal solution in a continuous parameter space is enhanced by a series of fundamental steps. The subsequent sections provide a comprehensive explanation of each stage, delving into the underlying principles and mechanisms.

(1)Initialization: The initialization phase is a critical step in the algorithm's exploration of the parameter space. A set of potential solutions is stochastically produced within the defined constraints of the exploration space. Every solution within the population represents a prospective combination of breakpoint positions and linear model coefficients for the problem of piecewise linear regression. The step of initialization plays a pivotal role in ensuring sufficient coverage and diversity in the parameter space, thereby enhancing the algorithm's ability to find the optimal solution.

(2)Mutation: Mutation is a vital operation within the differential evolution algorithm, serving to introduce diversity and encourage exploration during the search process. This procedure involves the generation of a mutant vector for each member of the population, achieved through the linear combination of parameter values from other randomly selected individuals. The mutation process is regulated by a mutation factor that determines the degree of diversity incorporated into the mutated vector. Through mutation, the algorithm can explore various regions of the parameter space, thereby reducing the probability of premature convergence to local optima.

(3)Crossover: The differential evolution algorithm incorporates the crossover operation as a key component, aiming to promote inter-individual information exchange within the population. In the crossover process, a new trial vector is generated by merging the parameter values of the target individual and the mutant vector. This procedure is subject to a crossover probability, which dictates the relative proportion of parameter values that are derived from both the target individual and the mutant vector. The crossover operation facilitates the exchange of information among individuals, thereby promoting diversity within the population and facilitating convergence towards the global optimum.

(4)Selection: The selection process is of paramount importance in the differential evolution algorithm, as it plays a crucial role in steering the evolutionary process towards improved solutions. During this stage, a comparative analysis is conducted between the trial vector and the target individual, taking into account their respective objective function values. In the context of the piecewise linear regression problem, these values represent the sum of squared residuals. The selection of individuals for the next generation is based on their objective function value, with those possessing a lower value considered superior and thus chosen to advance. Through a process of iterative selection of the fittest individuals, the algorithm ensures that the population gradually progresses towards better solutions over time.

(5)Termination: The differential evolution algorithm employs a termination criterion to signal the conclusion of the optimization process. The iterative process of the algorithm involves executing the mutation, crossover, and selection operations until a predetermined stopping condition is satisfied. Commonly used halting conditions include meeting a predetermined threshold for the number of iterations, achieving a targeted degree of convergence, or reaching a designated level of improvement in the value of the objective function. The choice of the termination criterion can significantly impact the algorithm's efficacy, as it establishes the balance between computational efficiency and solution quality.

To sum up, The 'pwlf' library allows users to customize the settings of the differential evolution algorithm, thereby enabling them to adjust the algorithm's performance to their specific problem requirements. The algorithm's search behavior can be fine-tuned by adjusting parameters such as the mutation factor, crossover probability, and termination criteria, enabling users to strike the desired balance between computational efficiency and solution quality. The integration of the differential evolution algorithm into the 'pwlf' library facilitates the precise identification of reliable trend groups, thereby enhancing the understanding of complex systems, such as effluent pumping stations, and aiding in the decision-making process. As such, it serves as a valuable tool for objective trend analysis.

2.2.2. Subjective trend analysis

By identifying and classifying trend groups based on the failure rate, which is represented by the slope of the cumulative failures against event time graph, subjective trend analysis offers a helpful complement to more traditional approaches to understanding the behavior of complex systems, see Figure 2.2. A more thorough understanding of a system's failure behavior is possible when using this auxiliary method in conjunction with the main objective analysis approach (piecewise linear regression), which addresses a variety of applications including failure prediction, risk assessment, and system performance evaluation.

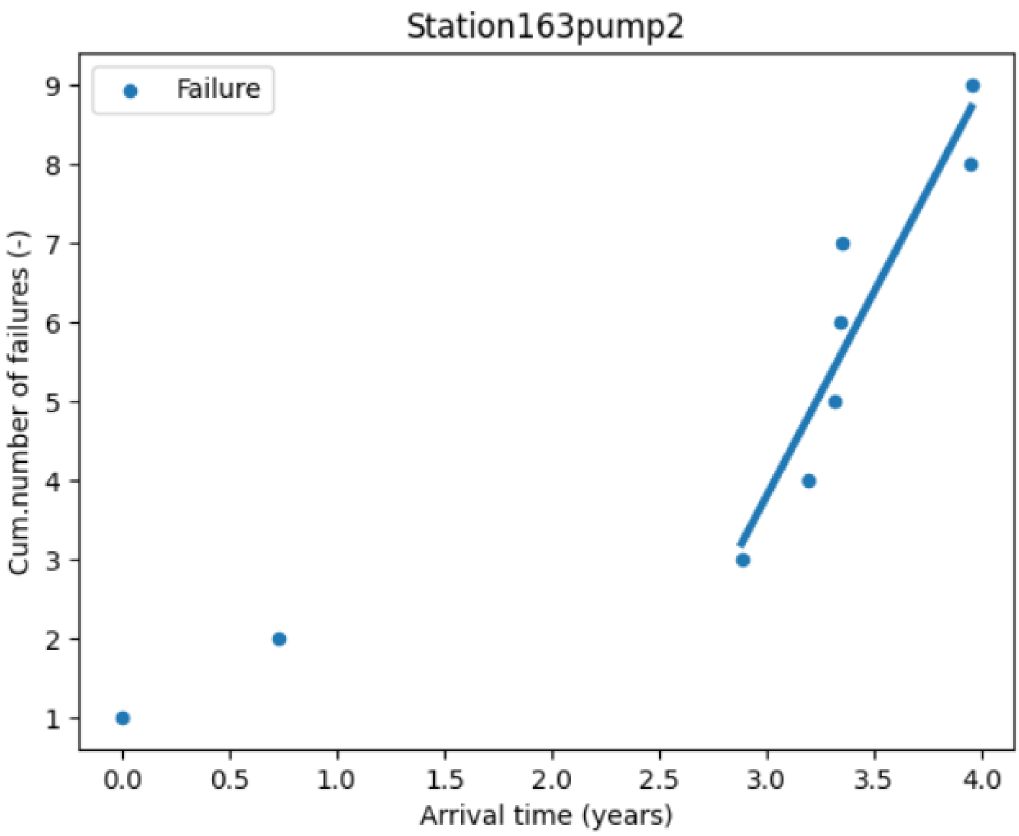


Figure 2.2: Station 163 Pump 2 with Subjective Trend Analysis

By locating breakpoints, improving linear models, and minimizing the sum of squared residuals, piecewise linear regression, the main objective analysis, uses statistical methods and algorithms to split the data into discrete trends. The 'pwlf' library's implementation of the differential evolution method is essential for accomplishing this optimization. Due to the inherent restrictions of mathematical optimization, data noise, or the setting of the algorithm, this method may not always be able to detect all potential patterns in the data. As a result, using subjective trend analysis as an additional technique can help overcome these drawbacks and ensure a more thorough identification of system trends.

Within the framework of custom subjective trend analysis, individuals with specialized knowledge and expertise in a particular domain utilize their skills to visually identify trends and breakpoints within the data. This enhances the results obtained from the objective analysis. The customized methodology provides a comprehensive comprehension of the system's malfunctioning patterns. The utilization of the subjective approach may be vulnerable to partialities and incongruities due to the inherent features of human cognitive processes.

A significant limitation of the gradient-based approach is evident when a data point is situated between two other points, resulting in a steep slope among the three points. The observed steep incline could potentially be attributed to an outlier, however, the approach utilized has the ability to partition the dataset into distinct clusters utilizing this atypical value. The aforementioned limitation underscores the significance of meticulously verifying the subjective patterns identified by professionals.

The subjective approach is recognized for its potential limitations; however, it is utilized as a supplementary method to support the objective analysis. The utilization of subjective trend analysis as a related instrument can enhance the comprehensiveness and dependability of the comprehension of the system's failure conduct by integrating the insights from both techniques. This approach can effectively surmount the partialities and incongruities that may emerge from the utilization of either method in isolation.

The primary rationale behind employing subjective trend analysis is to identify a multitude of trends, thereby compensating for the constraints of the 'pwlf' library in detecting the complete spectrum of potential trends. The library's failure to identify all trends may be ascribed to a multitude of factors, including:

(1)The configuration of an algorithm, specifically the differential evolution algorithm, is a crucial factor that affects its ability to converge to an optimal solution. Parameters such as population size, mutation rate, and crossover rate play a significant role in determining the algorithm's performance. Modifying these parameters has the potential to enhance trend identification, yet striking a harmonious equilibrium between computational efficacy and solution excellence may present a formidable task.

(2)The accuracy of trend and breakpoint identification by algorithms can be impeded by the existence of noise, outliers, or irregular patterns in the data, which collectively constitute data quality issues. The act of preprocessing the data with the intention of mitigating these issues has the potential to enhance the algorithm's performance. However, it is possible that certain limitations may still persist.

(3)The intricacy of the system being studied or the fundamental dynamics that dictate the failure behavior may pose a difficulty in modeling through piecewise linear regression in certain instances. The incorporation of subjective analysis offers an additional vantage point that may aid in the identification of patterns that are not readily discernible through quantitative models.

Subjective trend analysis is a valuable complementary method for identifying trends in complex systems, addressing the limitations of objective analysis and ensuring a more exhaustive detection of trends. By carefully balancing the use of both objective and subjective approaches, it is possible to gain a deeper understanding of the system's failure behavior and make more informed decisions related to maintenance, risk assessment, and other critical aspects of system management.

2.2.3. Application of Trend Analysis in the Proposed Methodology

As mentioned before, purpose of the trend analysis is to investigate pump failure data to identify various trends and patterns, which may indicate changes in the behavior of the pumps over time. This analysis employs a piecewise linear fitting approach to assess the cumulative failure count as a function of time (in years). The identified trends are then used to categorize the data into different groups based on the nature of the pump failure behavior. The following subsections describe the methodology and results of the trend analysis in detail.

(1) Data preparation: The input dataset is modified to include the time in years since the beginning of the observation period. This is computed by taking the difference between each failure's starting date and the initial starting date in the dataset, then converting the resulting timedelta to years. Additionally, a new column is added to store the cumulative failure count, which is computed using the cumulative sum of the failure instances.

(2) Model fitting: A piecewise linear fitting model is applied to the prepared dataset to identify trends in the cumulative failure count over time. The model is fitted using a specified number of segments, set at 4 based on visual inspection of the trend, which determines the granularity of the trend analysis. This model fitting step generates breakpoints that mark the boundaries of each segment.

(3) Segment filtering: The segments obtained from the fitted model are filtered based on the following criteria: a. each segment must contain at least three data points, and b. the maximum distance between adjacent points within the segment should not exceed a predetermined threshold value, set at one out of three of whole life period. This filtering step ensures that the identified trends are reliable and meaningful for further analysis.

(4) Visualization: The filtered segments are plotted along with the original data points. Each segment is displayed using a different color to distinguish between the various trends. The plot's title and filename are generated based on the pump station and pump number information.

(5) Trend categorization: Each valid segment is subjected to a combined statistical test. Based on the test results, the segments are categorized into three groups (i.e. NHPP, HPP, RP) that represent different types of pump failure behaviors. These categories are critical for understanding the underlying causes of pump failures and devising appropriate maintenance strategies.

2.3. Proposed Analysis Framework

The proposed analysis framework for assessing system performance and reliability consists of a four-step methodology, which includes segmenting each pump failure into similar trend groups based on objective and subjective trend analysis, performing multiple trend tests and model selection, see figure 2.3. The primary objective of this framework is to discern the evolution of pump performance over specified durations, such as a 5-year period or particular failure duration. This includes ascertaining whether the present failure patterns correspond with or diverge from historical data. In particular, a comparison is made with findings from Korving et al. from two decades prior,allowing analysts to make informed decisions regarding maintenance, system improvements, and resource allocation.

2.3.1. Step 1 : Segmenting Pump Failures into Similar Trend Groups

The first step involves segmenting each pump failure into similar trend groups based on objective and subjective trend analysis. This step includes data collection and categorizing units based on trend behavior. By combining both objective and subjective trend analyses, this approach offers a comprehensive view of the system's performance and allows for the identification of potential areas of concern.

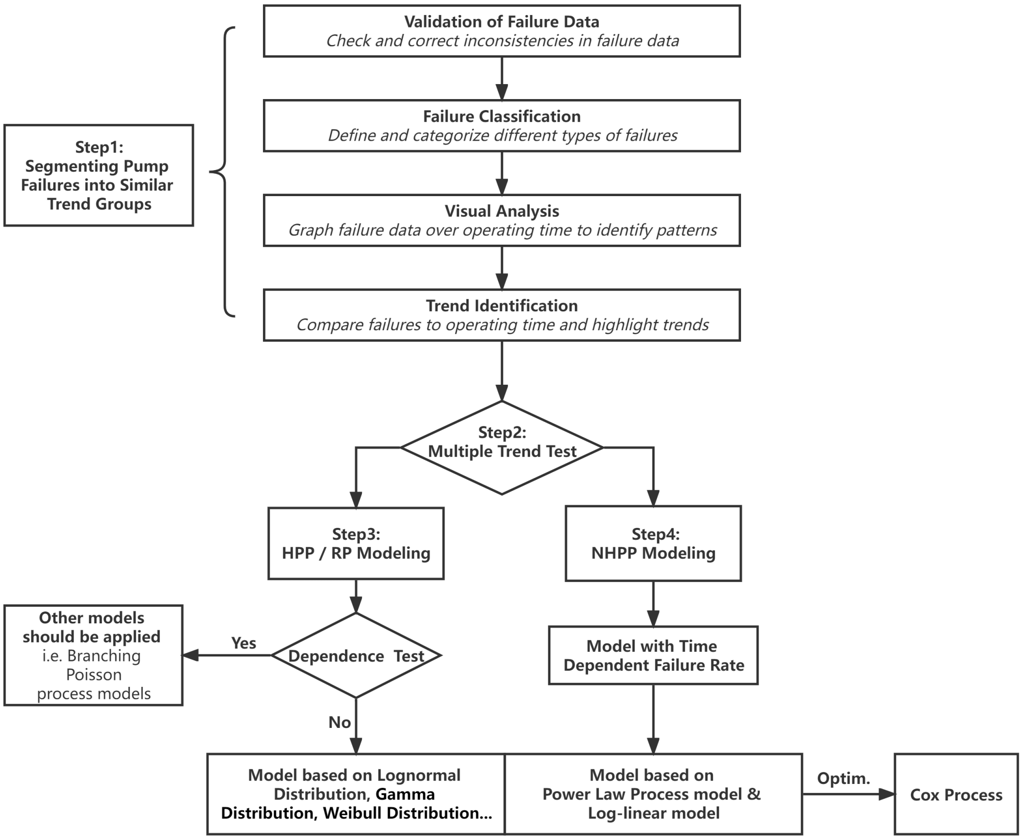


Figure 2.3: Procedure for analysis of failures of repairable systems

2.3.2. Step 2 : Multiple Trend Test

The second step involves performing multiple trend tests, such as the Laplace, Military Handbook, and Mann tests, on the collected failure data. These tests help identify whether the system's performance follows a Homogeneous Poisson Process (HPP), Renewal Process (RP), or Non-Homogeneous Poisson Process (NHPP).

Based on the results of the multiple trend tests, the system behavior is categorized as HPP, RP, or NHPP. The Laplace and Military Handbook test results are combined to reject or accept the null hypothesis of a Homogeneous Poisson Process. If either the Laplace or Military Handbook test result rejects the null hypothesis, the Mann test is performed to determine if the system follows an RP or NHPP. If neither the Laplace nor the Military Handbook test rejects the null hypothesis, the system is classified as an HPP.

The primary aim of Steps 1 and 2 in the analysis framework is to distinguish between trend-free groups and groups with trends. If the units in a group are trend-free, the analysis proceeds to Step 3; otherwise, it moves to Step 4.

2.3.3. Step 3 : HPP / RP Modeling

In Step 3, a dependency test is conducted for trend-free groups to determine if there is a correlation among the group data. A widely used approach for this analysis is to calculate the time between failures (TBF) and apply autocorrelation tests to detect any potential dependencies in the data. One such test is the autocorrelation function (ACF), which measures the correlation between a time series and its lagged values. Upon conducting the autocorrelation test for each renewal process segment, potential dependencies in the failure data can be detected, and appropriate reliability models can be selected. If a dependency is found among the group data, the branching Poisson process is deemed an appropriate reliability model; otherwise, the Homogeneous Poisson Process (HPP) or Renewal Process (RP) models are suggested.

In the analysis framework's second phase of Step 3, the primary goal is to determine the most suitable distribution type for each RP / HPP segment. This is a crucial step for understanding the reliability model governing each segment. The process involves fitting various probability distributions to the data and selecting the one that provides the best fit. To accomplish this, the Akaike Information Criterion (AIC) is utilized as a measure of the relative quality of statistical models for the given dataset.

The procedure begins by preparing the data for analysis. This involves extracting the relevant values from the data and converting them into a suitable format, such as a NumPy array. It is essential to ensure that the data is free of any inconsistencies or missing values that might impact the analysis. Once the data has been properly formatted, a list of candidate probability distributions is compiled, including the Beta, Exponential, Log-normal, Weibull, Weibull_max, and Gamma distributions. These distributions are chosen because they are commonly used in reliability engineering and have been found to adequately model various types of failure data (Ascher & Feingold, 1984).

The initial parameter estimates for the distributions, also known as hyperparameters, play a fundamental role in statistical modelling. Specifically, they serve as the starting points for the optimization algorithms employed to fit the probability distributions to the dataset. These initial estimates are crucial, as the efficiency and success of the convergence of the optimization algorithms are contingent on their appropriate selection. In other words, an ill-chosen initial estimate may culminate in an ill-fitting model or even preclude the algorithm's convergence.

In the current implementation, initial estimates of 1 are provided for the Beta, Log-normal, Weibull minimum, and Weibull maximum distributions. For the Exponential and Gamma distributions, no initial parameter estimates are specified, whereby the scipy library defaults to its inherent initial values.

To evaluate the goodness-of-fit for each distribution, the AIC is calculated for every candidate distribution. The AIC takes into account both the likelihood of the data given the model and the number of parameters in the model, providing a balance between the model's complexity and its ability to explain the data (Burnham & Anderson, 2002). Lower AIC values indicate a better fit, with the minimum AIC value being the preferred choice. This method helps avoid overfitting and ensures that the most parsimonious model is selected.

The calculation of the AIC involves several steps. First, the parameters of the candidate distribution are estimated using the method of maximum likelihood estimation (MLE), which maximizes the likelihood of the observed data given the model. This results in a set of parameter estimates that provide the best fit to the data according to the MLE criterion. The log-likelihood of the data given the estimated parameters is then calculated, providing a measure of how well the model explains the observed data. Finally, the AIC is computed by combining the log-likelihood with the number of parameters in the model, following the formula:

$$AIC = 2 * k - 2 * \text{log\_likelihood} \tag{2.2}$$

where k is the number of parameters.

Upon computation of the Akaike Information Criterion (AIC) for each potential distribution, the distribution exhibiting the minimum AIC value is designated as the optimal fit for the given dataset. Subsequently, the RP / HPP segments are classified into clusters based on the distribution that provides the most optimal fit.

Apart from the Akaike Information Criterion (AIC), it is advantageous to conduct a visual evaluation of the adequacy of fit for the most suitable distribution by means of Quantile-Quantile (Q-Q) plots. The graphical representations presented in this study offer a means of comparing observed data with the expected data derived from the most optimal distribution. Such plots serve as a visual tool for evaluating the model's capacity to accurately capture the underlying patterns in the data, as noted by (Cullen et al., 1999).

2.3.4. Step 4: NHPP Modeling

In the context of analyzing repairable systems, two commonly employed non-homogeneous Poisson process (NHPP) models are examined: one utilizing a power law process and the other utilizing a log-linear function. The utilization of models is implemented to capture the failure patterns and trends present in the data, which facilitates the generation of more precise predictions and maintenance planning.

The power law process model is a prevalent and versatile model utilized for the examination of failure data in repairable systems. The metric is derived from the aggregation of the total count of unsuccessful events over a period of time. The present model employs the Maximum Likelihood Estimation (MLE) technique to estimate the parameters that represent the failure rate and trend. The trend parameter is responsible for capturing the trend present in the data. If the value of the trend parameter surpasses a certain threshold, it indicates the presence of an increasing failure trend. Conversely, if the value of the trend parameter falls below the threshold, it indicates a decreasing trend. Finally, if the value of the trend parameter is equal to the threshold, it signifies a constant failure rate.

The log-linear model is a frequently employed methodology for the purpose of analyzing failures in systems that are capable of being repaired. The underlying assumption is that the total count of failures follows a logarithmic-linear pattern over time, with specific parameters reflecting the observed trend in the data. The method of moments is utilized to estimate these parameters. The trend parameter serves as an indicator of the direction of the trend observed in the data. A positive value denotes an upward trend, while a negative value denotes a downward trend. A value of zero indicates the absence of any discernible trend.

Upon application of the power law process model and the log-linear model for the purpose of scrutinizing the failure data of repairable systems, a more profound comprehension of the discerned patterns is attainable. The acquired data, which is expressed through trend parameters and failure counts, offers comprehensive understanding of the system's performance and facilitates more sophisticated trend evaluation. The models' identification of failure patterns and trends enables the classification of data segments into discrete groups based on whether they exhibit an upward trend, downward trend, or no trend. The classification is derived from an evaluation of the Mean Squared Error (MSE) metrics and patterns in the data. The segmented groups presented in this study serve as a fundamental basis for conducting further investigations and establishing maintenance schedules and system reliability. The study demonstrates the successful implementation of the Non-Homogeneous Poisson Process (NHPP) models in improving our understanding of system failures and facilitating strategic allocation of resources for maintenance operations.

Building on the foundation set by the Non-Homogeneous Poisson Process (NHPP) models, there's a shift to exploring another modeling technique: the Cox process. The impetus for this exploration arises from certain challenges encountered with traditional methods in scenarios with rapid alterations in failure rates.

The Cox process implementation serves to complement the NHPP models. What differentiates the Cox process is its innate flexibility in accounting for variations in failure patterns. Unlike traditional models that may treat failures as independent events spread evenly over time, the Cox process acknowledges the reality that failures can often cluster or occur periodically. By doing this, it gives a model that better captures the changes in system failures, leading to a more precise analysis.

3 Case Study

The primary focus of this case study is the failure analysis of a five-year dataset (2016-2020) obtained from the Department of Public Works of Rotterdam. The dataset contains detailed records of failure data on an annual basis, incorporating elements such as the identifiers for pump stations, the number of pumps within each station, precise timings for the commencement and conclusion of failures, the underlying causes of these failures, and the corresponding solutions executed to resolve these issues.

Throughout the five years under study, the dataset initially reports an extensive total of 30,851 failures. When analyzed further, 10,095 failures were identified as being of the type 'Storing Algemeen' or 'thermisch', across 1,698 pumps. These types translate to 'general problem' and 'thermal problem', respectively. This subset represents a significant 35% of the total number of reported failures.

After undergoing a rigorous sorting procedure that involved data cleaning, removing duplicate entries, and ensuring that modified failures were reliable, the final number of pump failures deemed suitable for inclusion in the study was 7,619. This adjusted figure represents a reduction from the previously mentioned count of 10,095. It's important to highlight that while the initial dataset comprised data from 1,608 pumps, after a thorough filtration process, the final sample size was trimmed down to 1,493. This reduction was mainly due to the removal of entries containing incomplete or insufficient data.

Among all the pump stations, Station 28 stands out as experiencing the highest number of failures, accounting for 178 in total. The pumps exhibit varying degrees of reliability, with the majority proving to be quite robust. Table 3.1 displays the distribution of the total number of failures per pump.

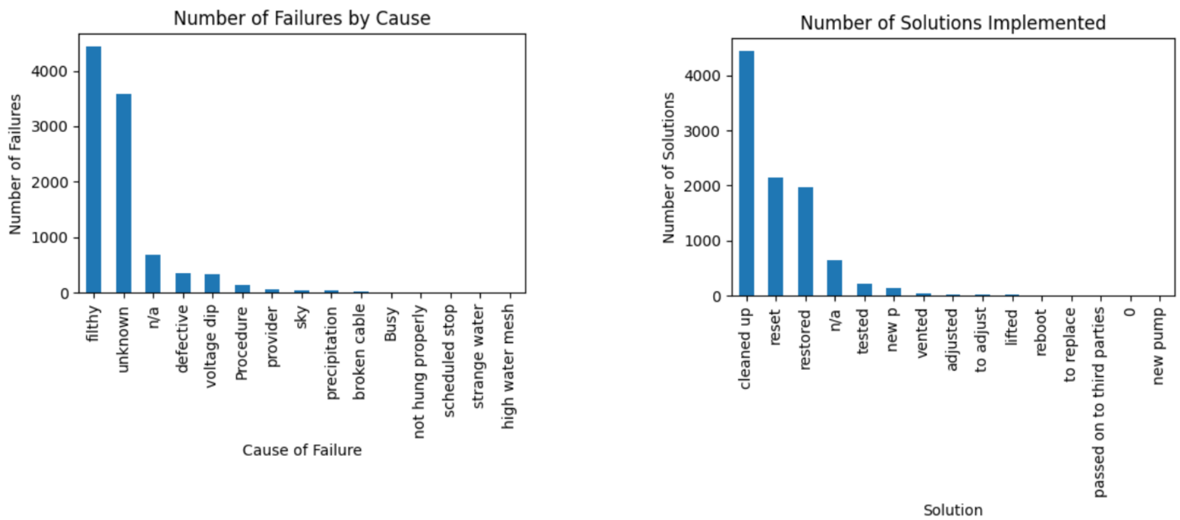
Identifying that a large proportion of failures is attributed to 'general problem' or 'thermal problem' types underscores the need for these areas to be prioritized for targeted maintenance and prevention strategies.

The dataset provides a detailed breakdown of failures based on their causes and the solutions implemented. A significant portion of these failures was attributed to excessive dirt accumulation within the pumps. As a result, the most prevalent corrective action involved a thorough cleaning of the affected pump. However, it's crucial to note that a substantial number of failures were classified as 'unknown' or 'n/a', nearly matching the frequency of those labeled 'filthy'. This considerable segment of unspecified failures cannot be overlooked, as they might represent unidentified issues or data recording discrepancies. To visually represent these findings, I have prepared two bar charts. The 'Number of Failures by Cause' chart showcases the frequency of each identified cause, with 'Filthy' and 'Unknown' being the predominant categories. In parallel, the 'Number of Solutions Implemented' chart demonstrates the commonly undertaken solutions, with 'Clean-Up' standing out as the predominant action. Refer to Figure 3.1 for a comprehensive view of these insights.

In this context, the term "filthy" refers to failures caused by excessive dirt or contaminants within the pumps. Essentially, when a pump is labeled as having a "filthy" failure, it indicates that the malfunction was due to dirt accumulation that hindered its operation. On the other hand, failures labeled as "sky" denote those caused by air intrusion or air pockets within the system. When a pump experiences a "sky" failure, it means that the presence of air disrupted its regular water flow or operation. Additionally, it should be noted that the "n/a" designation is consistent across both the cause and solution graphs, representing instances where the specific cause or solution is not available or not applicable. It's essential to comprehend these terminologies as they provide insights into the root causes of failures and guide the appropriate corrective actions.

Range of Failures	Number of Pumps
0-10	1379
10-50	202
50-100	21
>100	6

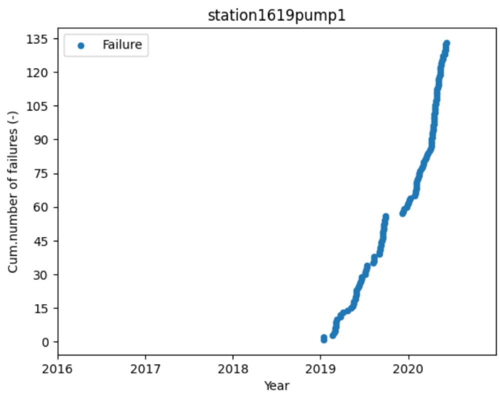
Table 3.1: Distribution of Total Failures Per Pump



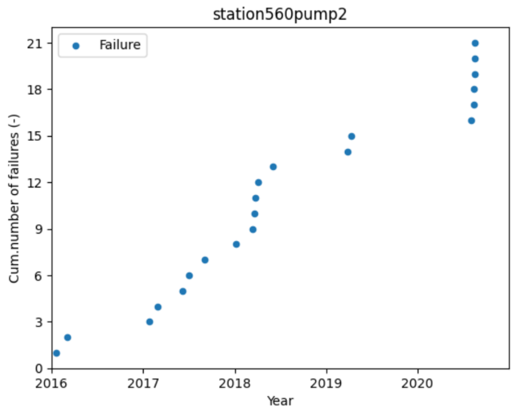
(a) Cause of Failure **(b)** Solution

Figure 3.1: Comparative analysis of the Number of Failures and Solutions.

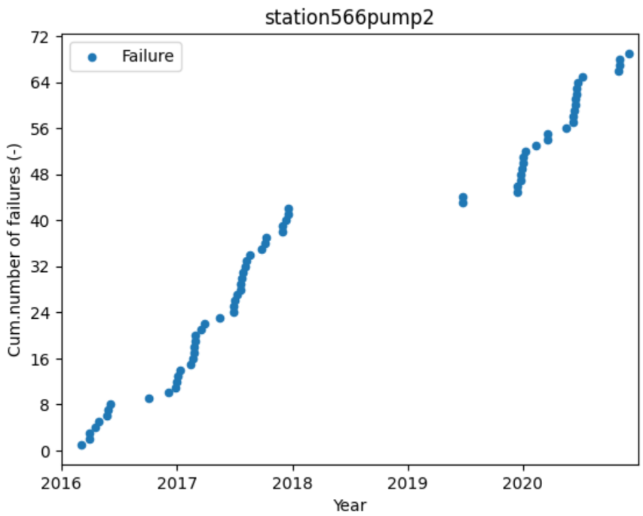
For a more comprehensive examination of the dataset, three pumps, namely Station 1619 pump1, Station 566 pump2, and Station 560 pump2, were singled out due to their clear trend analysis outcomes and their representative nature in model application. Graphs delineating the cumulative failures of these pumps over time were generated, vividly showcasing the correlation between the timing of failure events and their occurrence frequency. These visual tools have been particularly illuminating, offering pronounced insights into the reliability and operational performance of these pumps over the five-year span, as depicted in Figures 3.2. Further details and results pertaining to this analysis will be expanded upon in the ensuing chapters.



(a) Cumulative number of failures against arrival time (years) for station1619pump1



(b) Cumulative number of failures against arrival time (years) for station560pump2



(c) Cumulative number of failures against arrival time (years) for station566pump2

Figure 3.2: Cumulative number of failures against arrival time (years) for various pumps.

4
Result

4.1. Examination of Failure Data

The analysis provided insights into the Rotterdam pump system's reliability and maintenance. The selected 447 pump stations are integral to this system. By studying pump malfunctions and their causes, we can better strategize to maintain pump performance and extend its lifespan.

4.2. Trend of Interarrival Time

This section utilizes a piecewise linear fitting method to assess the cumulative failure count in relation to time (measured in years). The aforementioned approach facilitates the discernment of discrete patterns, which can subsequently be employed to classify the data into diverse clusters contingent upon the characteristics of the pump malfunction.

In order to showcase the efficacy of the piecewise linear regression methodology, we provide three discrete instances, each emphasizing distinct facets of the examination. The graphs presented depict the cumulative sum of failures over time for different pump stations and pumps, serving as illustrative examples. The application of piecewise linear regression facilitates the discernment of discrete trend clusters, furnishing significant perspectives into the failure characteristics of said systems.

In order to enhance the clarity of the trend analysis, the following graphs adjust the time scale. Rather than spanning the entire five years, the x-axis will be rescaled to represent the period during which failures occurred. It will start from the point where data collection began, and end with the last recorded failure. This allows for a more detailed exploration of the malfunction patterns.

The initial instance (station 1619 pump 1) displays a near-constant escalation in the cumulative total of malfunctions, see Figure 4.1. The graphical representation displays four discernible clusters of trends, indicating that the piecewise linear regression has effectively captured the fundamental patterns in the pump's failure behavior. While the graphical representation reveals what might be interpreted as four clusters of trends, it's important to note the limitations of my approach. Particularly in the arrival time between 0.6 to 0.8 years, there's a noticeable deviation between the failure data and trend analysis data, which struggles to adapt to swift changes in failure rates.

The graph in the second instance (station 566 pump 2) exhibits a depiction of solely three segments

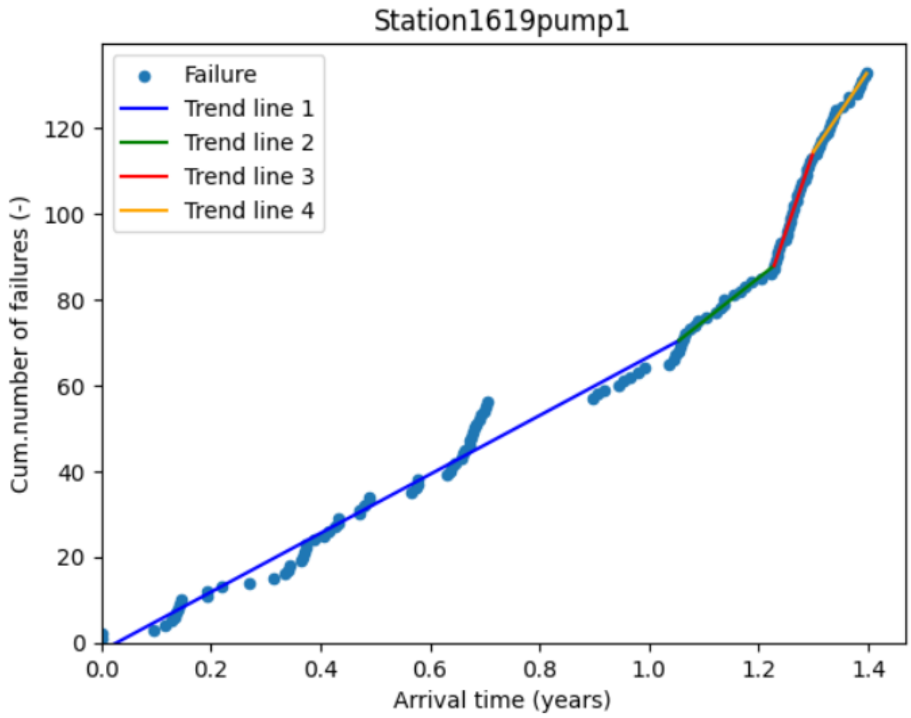


Figure 4.1: Station 1619, Pump 1 - Four Trend Groups

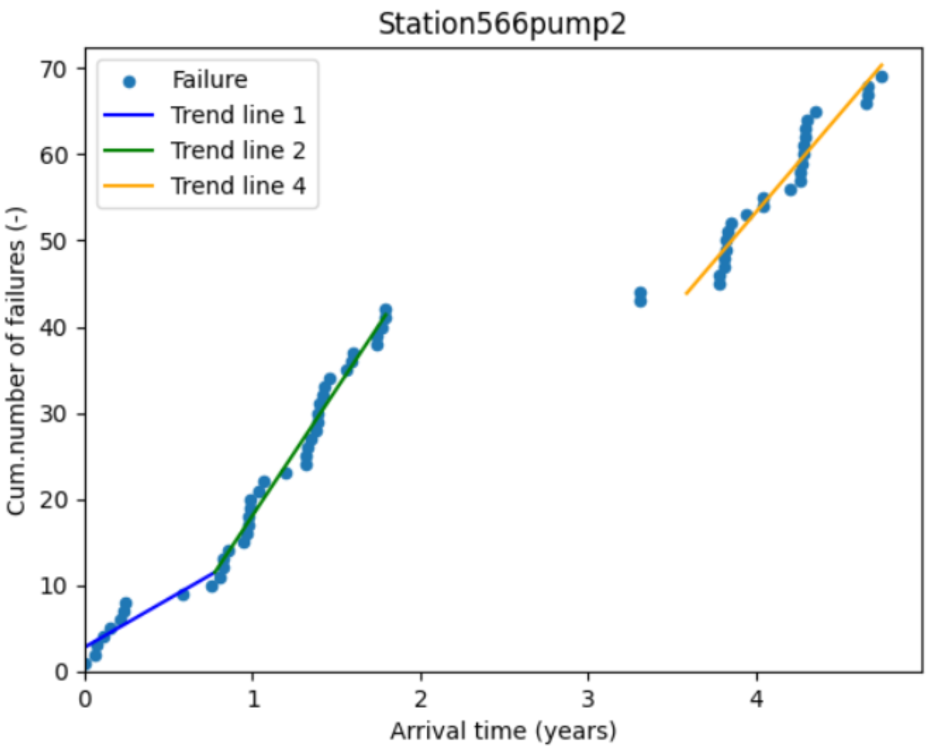


Figure 4.2: Station 566, Pump 2 - Segment 3 Excluded

(see Figure 4.2), thereby implying the exclusion of the third segment. The aforementioned exclusion is a result of the temporal distance between consecutive occurrences within the given segment exceeding the pre-established threshold, which has been defined as one-third of the total lifespan of the pump. Through the exclusion of this particular segment, the analysis guarantees the dependability and significance of the identified patterns, thereby reducing the impact of false or temporary variations in the failure data.

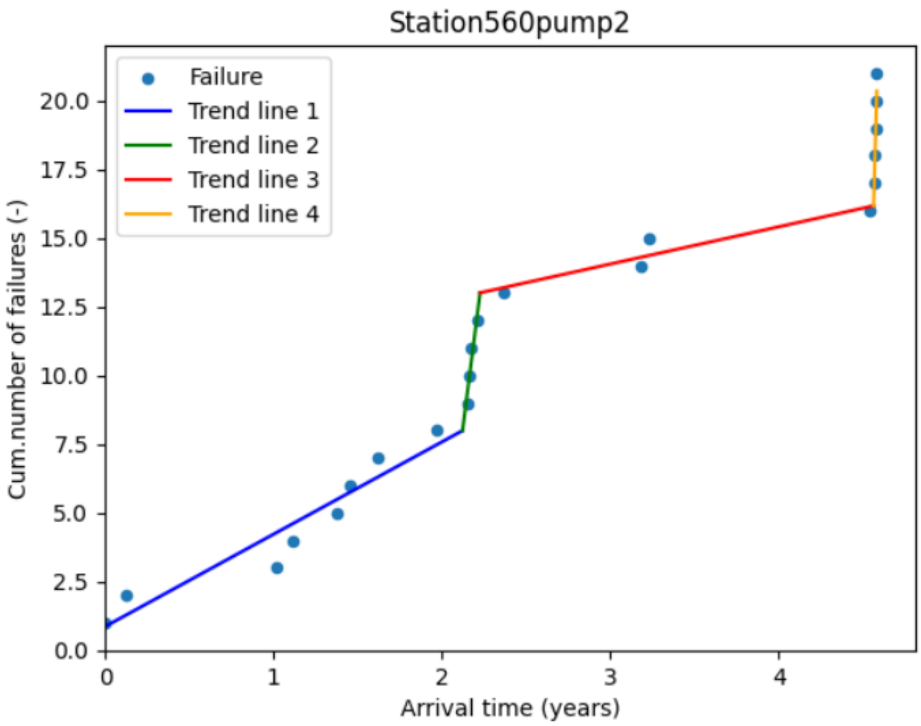


Figure 4.3: Station 560, Pump 2 - Four Trend Groups: Two Stable and Two Increasing

In the next example (station 560 pump 2), the graphical representation of the data shows four distinct trend groups, out of which two exhibit a stable pattern and the other two demonstrate a sharp increase (see Figure 4.3). This variation indicates that the pump experienced periods of both relative stability and increased failure, reflecting a complex reliability profile. The consistent pattern in Groups 2 and 3, where no discernible trend is observed, offers insight into the pump's periods of operational stability. Recognizing this stability, or the absence of a trend, is essential and should be considered in subsequent RP/HPP modeling steps.

Figure 4.4 presents station 214 pump 1, which exhibits an increase in failures during a specific timeframe, followed by a prolonged period with few failures. The graph underscores the significance of taking into account the temporal dynamics of pump failure behavior. The marked surge in failures can be ascribed to particular operational or environmental conditions during that timeframe.

To summarize, the aforementioned four instances exemplify the varied spectrum of failure patterns that can be detected via piecewise linear regression analysis. Through the analysis of the cumulative sum of failures over time, this methodology facilitates the identification of discrete trend cohorts, thereby affording significant elucidation regarding the fundamental origins of pump malfunctions. Upon performing integrated trend analyses, the ensuing results were acquired:

The analysis indicates that there were no Homogeneous Poisson Processes (HPP) segments identified, which suggests that there were no occurrences where the rate of failure remained consistent throughout the duration of the observation period.

The study identified 121 segments as Renewal Processes (RP), indicating that the inter-arrival time exhibited a constant intensity without any discernible trend.

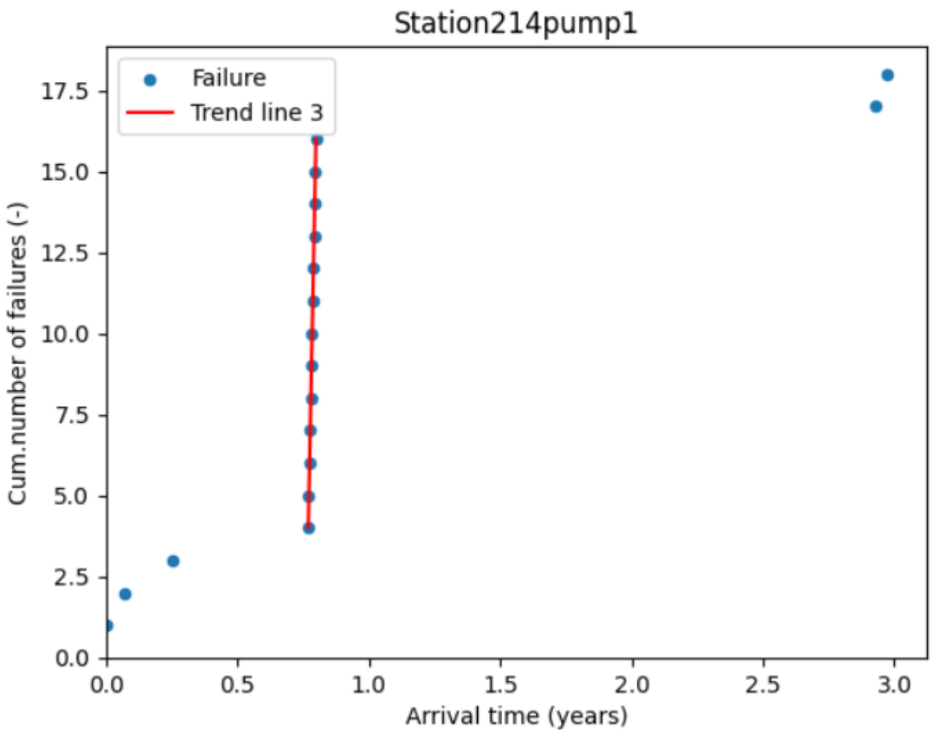


Figure 4.4: Station 214, Pump 1 - Growth and Stability

The study identified a total of 459 segments that were categorized as Non-Homogeneous Poisson Processes (NHPP). This classification indicates that the failure rate of the pumps in these segments exhibits a non-linear pattern of change over time.

The aforementioned results provide perspectives on the malfunctioning mechanisms of the 447 pumps that were chosen for analysis. The NHPP behavior was observed in the majority of the pumps, indicating that their failure rates display non-linear fluctuations over time. Approximately 50% of the pumps (218 pumps) did not exhibit any discernible pattern according to the initial piecewise regression analysis.

4.3. Model Application

4.3.1. NHPP Modelling

The utilization of NHPP models has produced valuable findings in the pump failure data examination. In the context of NHPP modeling, a total of 351 segments were subjected to log-linear modeling, while 106 segments were modeled using the PLP model, see Table 4.1. Additionally, 415 segments were observed to demonstrate an upward trend, while 42 segments exhibited a downward trend. The findings indicate that in certain cases, the Cox log-linear model was the only suitable fit, as three segments were found to be non-compliant with Crow's model.

The Cox's log-linear model was employed to assess the adequacy of Station 1619 Pump1 Segment1 and Station 566 Pump2 Segment4. The results indicated that both stations exhibited a satisfactory fit, with corresponding mean squared error (MSE) values of 12.79 and 9.63. It is noteworthy that the model exhibited substantial deviation from observations for Station 1619 Pump1 Segment1 within the time frame of 0.6-0.8 years, see Figure 4.5. Similarly, for Station 566 Pump2 Segment4, the model demon-

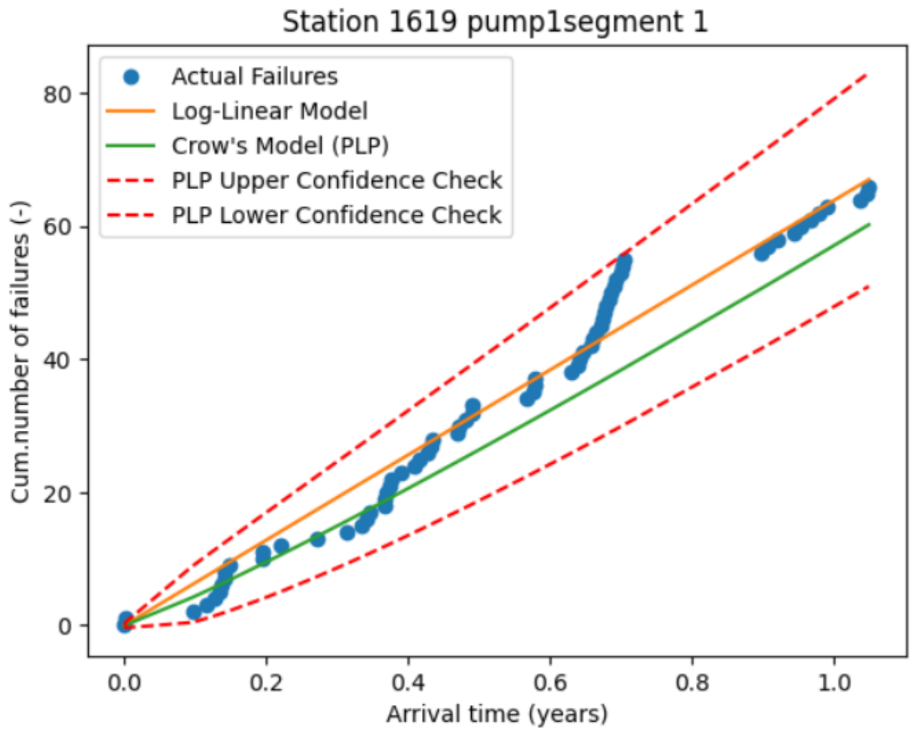


Figure 4.5: Comparison of Crow's and Cox-Lewis' model with observed failures of Station 1619 Pump 1 Segment 1

Model/Process	Number of Segments
Log-Linear Model	351
PLP Model	106
Upward Trend	415
Downward Trend	42

Table 4.1: Summary of NHPP Modelling Results

strated considerable deviation from observations throughout the entire duration, particularly when the frequency of failures increased significantly, see Figure 4.6. Despite the aforementioned deviations, a limited number of observations were found to lie beyond the 95% confidence interval for both pumps.

Adding to this, in the context of Station 560 Pump 2 Segment 1, both the Cox's log-linear model and Crow's PLP model were applied with fine results, albeit with minor differences. The Crow's PLP model proved to be slightly more congruent with a mean squared error (MSE) of 0.74, as opposed to the Cox's model with an MSE of 1.15, as reflected in the analysis.

4.3.2. RP Modelling

As mentioned before, Station 560 Pump2 Segment 2 was employed as an illustrative instance for the purpose of RP modeling. The majority of RP datasets were composed of four distinct data points, wherein two lag values were chosen for analytical purposes. The results of the dependency test revealed that the Time Between Failures (TBF) data exhibited no correlation, as illustrated in the figure representing the dependency test with 5% significance limits.

Given the absence of correlations in the failure data, the RP technique was deemed suitable for modeling the reliability of repairable units. For instance, Station 560 Pump2 Segment2 was best fitted with a Weibull distribution (AIC = -55.18), with the Q-Q plot revealing a good fit for the four data points. In

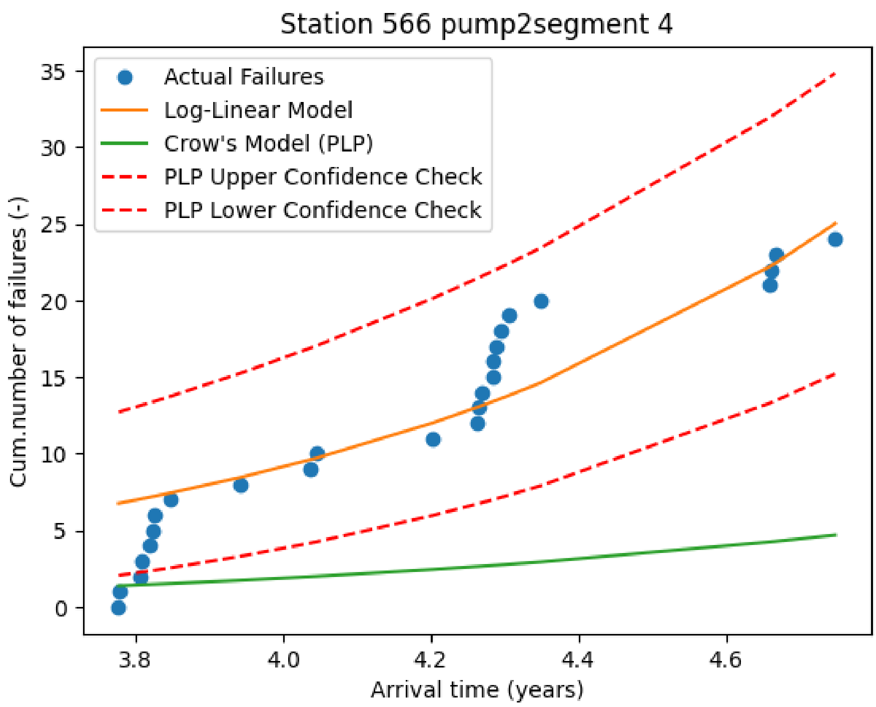


Figure 4.6: Comparison of Crow's and Cox-Lewis' model with observed failures of Station 566 Pump 2 Segment 4

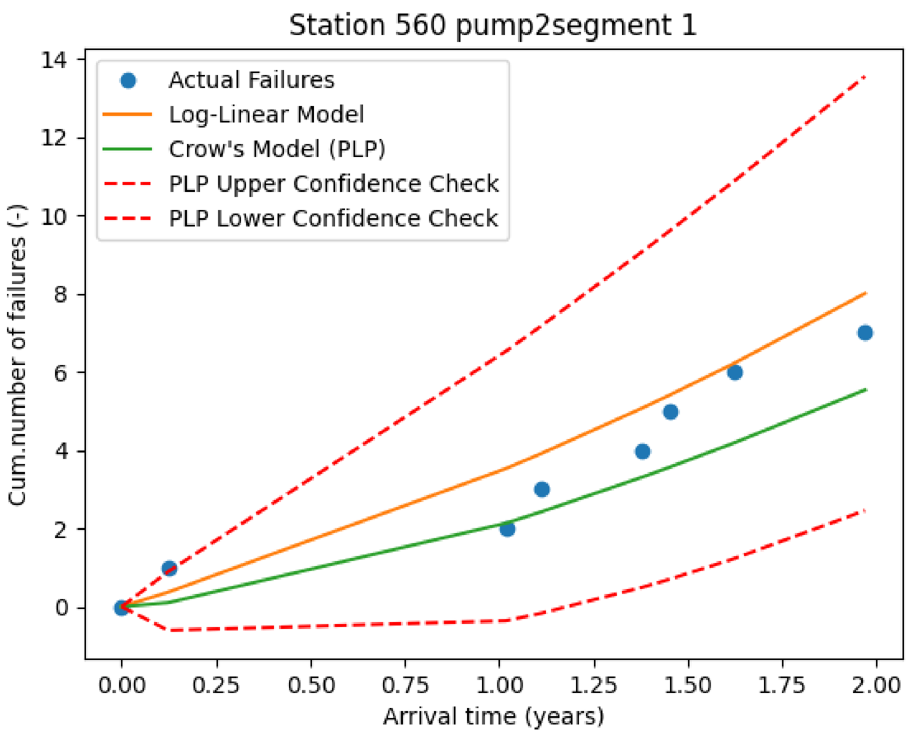


Figure 4.7: Comparison of Crow's and Cox-Lewis' model with observed failures of Station 560 Pump 2 Segment 1

summary, the final fitting results were as follows: 11 segments conformed to a gamma distribution, 14 segments to a lognormal distribution, 17 segments to a Weibull Minimum Extreme Value Distribution, 48 segments to a Weibull Maximum Extreme Value Distribution, and 31 segments to a beta distribution.

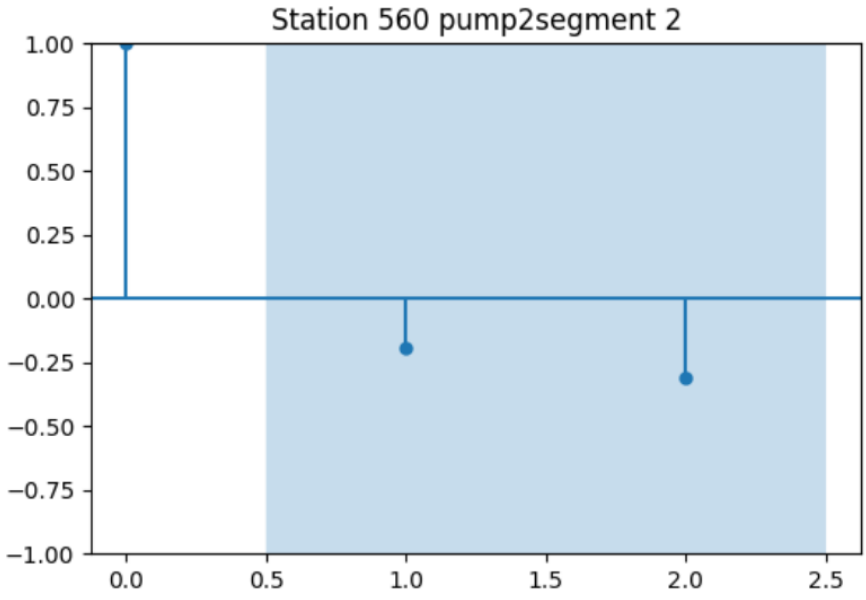


Figure 4.8: Dependency test of TBF with 5% significance limits of Station 560 Pump 2 segment 2

Distribution Type	Number of Segments
Gamma Distribution	11
Lognormal Distribution	14
Weibull Minimum Extreme Value Distribution	17
Weibull Maximum Extreme Value Distribution	48
Beta Distribution	31

Table 4.2: Summary of RP Modelling Results

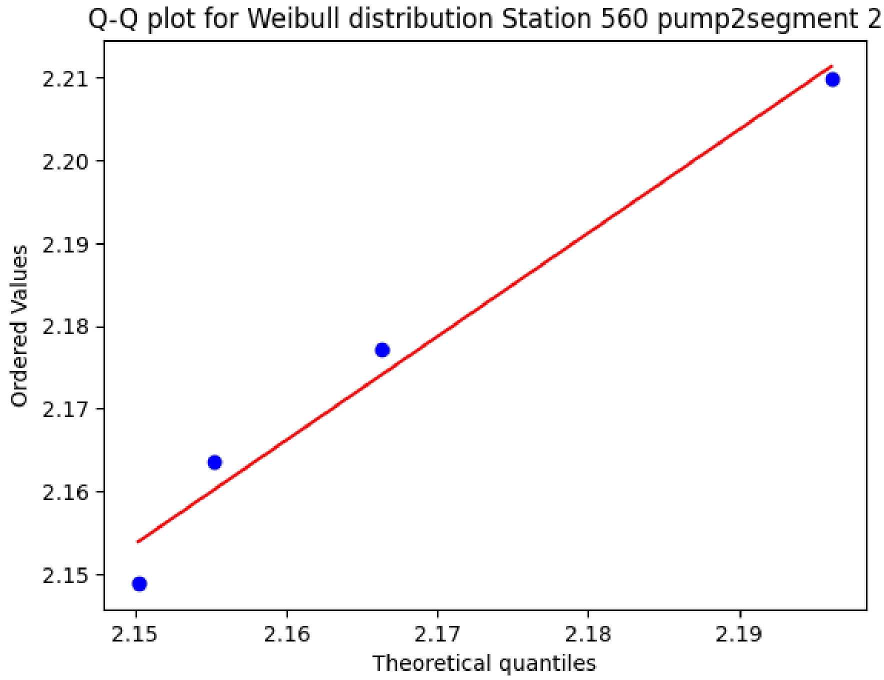


Figure 4.9: Q-Q Plot for Assessing Goodness-of-Fit of Selected Distribution of Station 560 Pump 2 segment 2

These results demonstrate the effectiveness and utility of both NHPP and RP modeling techniques for analyzing pump failure data. By identifying the best-fitting models for each segment, reliability engineers can better understand the underlying failure patterns and improve maintenance strategies accordingly. Despite the valuable insights gleaned from this analysis, there are undoubtedly areas for improvement and further research, particularly concerning our modeling methodologies and pump failure segmentation. These key points and potential enhancements, which could further refine our understanding of pump reliability, will be thoroughly discussed in the following section.

In addition to the effectiveness and utility of NHPP and RP modeling techniques for analyzing pump failure data, it is important to note that segments or groups are merely auxiliary tools used to analyze the pump failure behavior. Different kinds of trend segments may exhibit inconsistent failure patterns, while consistent failure patterns can also emerge within specific segments. These variations and patterns will be further explored and discussed in Section 6.7, where an examination of these aspects and potential enhancements that can refine the understanding of pump reliability will be conducted.

5

Exploring the Process: Methodology and Result Enhancement

5.1. Trend analysis Optimization

The results of investigation suggest that the utilization of piecewise linear regression was not efficacious for roughly 50% (218 pumps) of the scrutinized pumps. In order to tackle this issue, the methodology was enhanced by integrating the Bayesian Information Criterion (BIC) to automatically determine the optimal number of breakpoints for the piecewise linear regression model. Formerly, an indiscriminate initial value, such as 4, was utilized, whereas presently, the employment of the BIC has been adopted. The Bayesian Information Criterion (BIC) serves as a criterion for model selection by balancing the model's complexity with its ability to fit the data, thereby avoiding the issues of overfitting or underfitting. The aforementioned abilities are indicative of this enhancement.

(1) The BIC calculation function: This function calculates the BIC for a given number of segments and a piecewise linear regression object. The number of parameters and the number of data points are determined based on the input. The residual sum of squares (rss) is computed as the sum of the squared differences between the observed and predicted values. The BIC is then calculated using the appropriate formula:

$$BIC = n_{\text{data points}} * \log\left(\frac{rss}{n_{\text{data points}}} + 10^{-10}\right) + n_{\text{params}} * \log\left(n_{\text{data points}} + 10^{-10}\right)$$

To ensure numerical stability, a small positive constant (10^{-10}) is added to the denominators in the logarithmic terms.

(2) The segment optimization function: This function aims to optimize the number of segments for the piecewise linear regression model by minimizing the BIC value. The employed approach utilizes an optimization technique to explore the optimal integer value of segments within predetermined boundaries, ranging from 2 to the default upper limit of 5. The process of optimization entails the minimization of the Bayesian Information Criterion (BIC) value, which is computed by the BIC calculation function. Subsequent to optimization, the outcome is conveyed as an integer.

Furthermore, a strict convergence criterion is established by imposing an absolute tolerance of 1 on the optimization process, thereby facilitating a proficient and accurate identification of the optimal quantity of segments.

The successful detection of trends in 19 pumps has been attributed to the implementation of the enhanced methodology. The graph depicted in Figure 1, pertaining to station 1621 pump 1, illustrates that prior to 0.1 years, there is a noticeable pattern in the occurrence of approximately 8 cumulative failures. The utilization of Bayesian Information Criterion (BIC) for optimizing the quantity of segments in the piecewise linear regression model has been established as a significant improvement, facilitating precise and resilient trend identification in the examination of pump malfunctions.

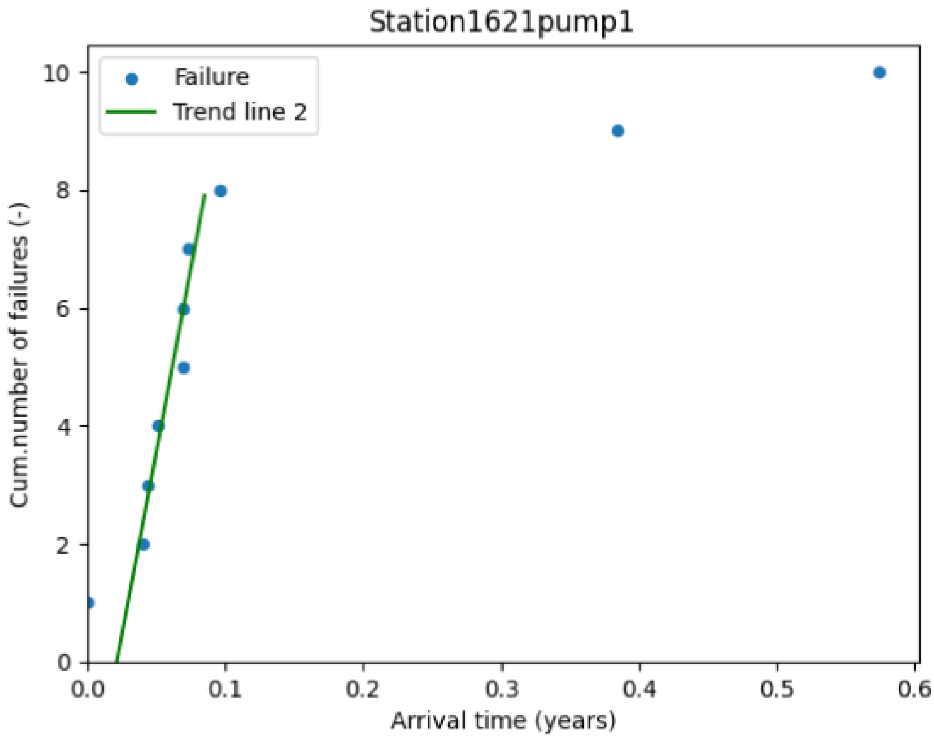


Figure 5.1: Cumulative Failures and Optimized Piecewise Linear Regression Segments of Station 1621 Pump 1

The methodology incorporates a secondary innovative approach that aims to identify potential trends among pumps that have not demonstrated any noticeable trends through the primary method. Given that 199 pumps did not exhibit a discernible trend following the implementation of the primary method, it became imperative to investigate supplementary techniques in order to maximize the amount of data that could be extracted from the aforementioned pumps. The methodology known as subjective trend analysis, as outlined in Section 3.2.2, endeavors to accomplish this objective through the utilization of a more permissive approach to the identification of trends. By setting the threshold for failure rate to infinity, the approach facilitates the detection of a maximum number of patterns among the pumps that are still operational. In order to attain the desired outcome, a series of interrelated procedures are carried out, encompassing data pre-processing, group creation, trend identification and classification, linear regression, and visualization.

During the process of data preprocessing, the cumulative count of failures and the time intervals between consecutive failure dates are computed. Additionally, an internal threshold is determined based on the overall time duration of the dataset. The dataset is subjected to iteration, wherein groups are formed to represent potential trend segments based on year differences and failure rates. Each group necessitates a minimum of three data points for further analysis.

Subsequently, the groups are subjected to scrutiny for plausible patterns through employment of the amalgamated trend examination. The segments of trends that are detected are then categorized as either NHPP, RP, or HPP. A linear regression model is applied to each legitimate group, and the resultant line of best fit is graphed on a scatter plot that depicts cumulative failures over time. This approach effectively illustrates the identified patterns for each pump.

Through this interconnected approach, the secondary gradient-based methodology aims to enhance the trend detection process by revealing additional trends among the remaining pumps. One example is Station 412 Pump 2. After applying the secondary gradient-based approach, the fitted line demonstrates an excellent fit to the data, indicating the success of the methodology in identifying trends for this particular pump.

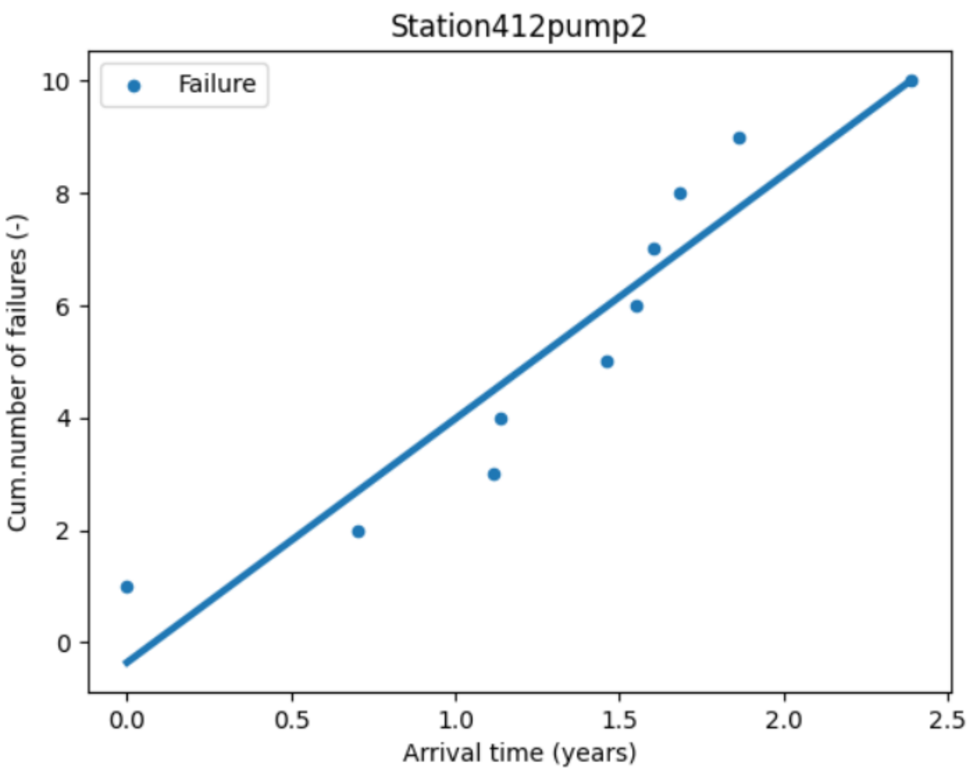


Figure 5.2: Subjective Trend Analysis for Station 412 Pump 2

To sum up, the outcomes derived from implementing the primary and secondary novel methodologies demonstrate a noteworthy enhancement in identifying trends. Following the implementation of the secondary methodology, 110 pumps out of the original 199 pumps are left unclassified, despite the lack of any discernible trend in the initial dataset. This underscores the efficacy of the implemented improvements in tackling the difficulty of discerning patterns in pump malfunction data.

The ultimate categorization of the recognized trend segments is presented as follows:

The number of segments in the Non-Homogeneous Poisson Process (NHPP) is 513. There are a total of 172 segments in the Renewal Process (RP). The current state of the Homogeneous Poisson Process (HPP) is that there are no segments.

In general, the utilization of Bayesian Information Criterion (BIC) optimization for segment determination in piecewise linear regression, coupled with the subjective trend analysis approach for scrutinizing failure rates, has effectively augmented the capacity to identify and categorize trends in pump failure data.

Subsequently, the results Ire observed by following the NHPP and RP modeling steps as outlined here. In the context of NHPP modeling, it was observed that 388 segments Ire subjected to fitting with a log-linear model, while 122 segments Ire subjected to fitting using the PLP model. Additionally, 459 segments exhibited an upward trend, whereas 51 segments displayed a downward trend, see Table 5.1. It is noteworthy that in three instances, certain segments did not adhere to Crow's model, thereby indicating that the Cox log-linear model was the sole appropriate fit.

Model/Process	Number of Segments
Log-Linear Model	388
PLP Model	122
Upward Trend	459
Downward Trend	51
Not Adhering to Crow's Model	3

Table 5.1: Summary of Revised NHPP Modeling Results

With regards to RP modeling, the ultimate fitting outcomes Ire as follows: 17 segments Ire found to conform to a gamma distribution, 20 segments to a lognormal distribution, 25 segments to a Weibull Minimum Extreme Value Distribution, 61 segments to a Weibull Maximum Extreme Value Distribution, and 49 segments to a beta distribution, see Table 5.2, an additional example is introduced.

This example involves Station 1613 Pump 1 Segment 3, a segment identified using a novel methodology that incorporates subjective trend analysis. Figure 5.3 presents the dependency test of time between failures (TBF) with 5% significance limits for this segment. Additionally, Figure 5.4 offers a Quantile-Quantile (Q-Q) plot for the same segment.

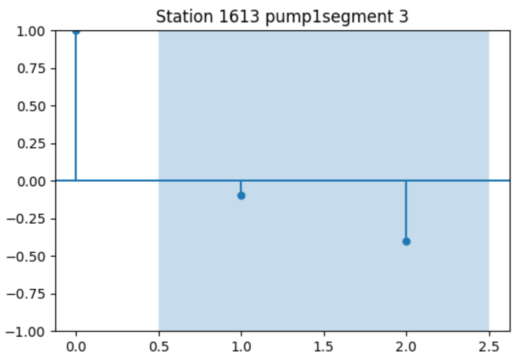


Figure 5.3: Dependency test of TBF with 5% significance limits of Station 1613 Pump 1 segment 3

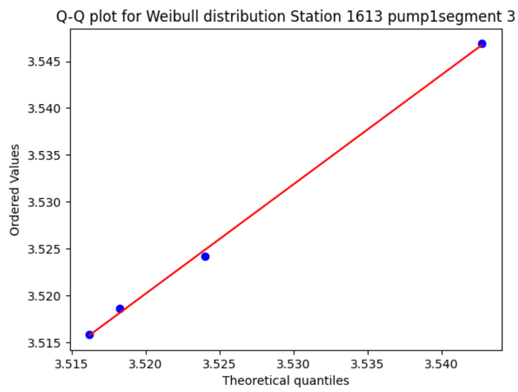


Figure 5.4: Q-Q Plot for Assessing Goodness-of-Fit of Selected Distribution of Station 1613 Pump 1 segment 3

The inclusion of this example demonstrates the practical application and effectiveness of the new methodological approach, combining statistical modeling with subjective trend analysis, in the identification and characterization of additional segments in the system..

The results indicate that the refinement of trend analysis is primarily effective in identifying the trend category for RP modeling, underscoring the significance of the improvements implemented in this investigation.

Although the findings exhibit noteworthy enhancements in trend detection, NHPP modeling still encounters certain challenges. As an illustration, certain models exhibit substantial deviations from the empirical data, and broad intervals in specific instances may suggest the existence of clustering. The

Distribution	Number of Segments
Gamma	17
Lognormal	20
Weibull Minimum Extreme Value	25
Weibull Maximum Extreme Value	61
Beta	49

Table 5.2: Summary of Revised RP Modeling Results

applicability of conventional growth models, such as Crow's and Cox-Lewis' models, in characterizing this phenomenon may be limited due to their constrained nature stemming from their underlying assumption of a monotonically increasing or decreasing failure rate over time. In instances where there is a significant increase in the number of failures, the initial NHPP model may not effectively replicate the instantaneous failure rate.

The aforementioned constraint underscores the necessity for alternative techniques of modeling, such as the Cox process. The Cox process, which is also referred to as the doubly stochastic Poisson process, is a highly adaptable and potent technique for representing point processes that exhibit fluctuating intensity. The utilization of a stochastic intensity function enables enhanced adaptability in capturing non-monotonic patterns and accommodating the dynamic nature of failure rates (Bartlett, 1963).

5.2. Cox Process

The Cox process implementation in this study is designed to address the limitations of traditional methods when dealing with rapid changes in failure rates. By allowing for the assumption that failures occur either clustered or periodically, the Cox process provides a more appropriate solution in such scenarios.

5.2.1. Gaussian Process

Prior to fitting a Gaussian process to the data, it is necessary to generate input data that can be used for fitting the Gaussian process model. The data generation process involves preparing the input time data (X) and generating corresponding random Poisson-distributed values (Y). The input time data, X, is derived from the original dataset, which contains the interarrival times of failure events. In addition to the input time data, a set of corresponding Poisson-distributed random values, Y, must be generated as an array of Poisson-distributed random values, with a size equal to length of dataset (Williams et al., 2006).

Then, a Gaussian process is applied to fit to the data using GPy, which is a flexible and efficient Python library for Gaussian process modeling. The GPy library provides a user-friendly interface for constructing, training, and making predictions with GP models. To fit the GP to the data, a radial basis function (RBF) kernel is used as the covariance function. The RBF kernel is a widely employed kernel in GP modeling due to its smoothness and ability to capture complex, non-linear relationships in the data. It is defined by two parameters: the input dimension, which is set to 1 for univariate time series data, and the lengthscale, which determines the smoothness of the function.

The RBF kernel is initialized with a variance of 1.0 and a lengthscale of 0.1, which are both hyperparameters that influence the model's flexibility and ability to capture the data's structure. These hyperparameters can be tuned to optimize the model's performance. The GP model is constructed using the GPy.models.GPRegression class, with the input time data and the corresponding Poisson-distributed random values provided as inputs, along with the RBF kernel.

To find the optimal hyperparameters for the GP model, the $model.optimize_restarts()$ function is employed, which performs optimization using the selected kernel and restarts the optimization process multiple times to avoid getting stuck in local minima. The number of restarts is set to 10 in this case, ensuring a thorough exploration of the hyperparameter space while keeping computational costs manageable.

5.2.2. Prediction of Intensity function

Upon successful fitting of the Gaussian process model to the data, the intensity function - a fundamental component of the Cox process - can be predicted. The process of prediction entails the generation of a collection of novel input values, which will serve as the basis for estimating the intensity function. The creation of New X involves the generation of a linearly spaced array comprising 100 values that span the range between the minimum and maximum values of the initial time data. After acquiring the latest input parameters, the model utilizes its predict function to estimate the corresponding mean and variance values.

Following this, the intensity function's gradient is calculated by finding the difference between sequential average values, then dividing it by the difference between the related input values. This procedure yields an array made up of intensity gradients, which are then used to determine the average gradient. The mean slope provides insight into the general pattern observed in the intensity function. The categorization of the data into one of three groups is determined by the polarity of the mean slope. In the event that the mean slope exhibits positivity, it can be inferred that the intensity function is on an upward trend, thereby indicating a reduction in the interarrival times. Conversely, in the event that the mean slope is negative, the intensity function exhibits a declining trend, thereby indicating a corresponding increase in the interarrival durations. When the average slope tends towards zero, it can be inferred that the intensity function exhibits a state of near-constancy. This observation may be indicative of interarrival times that are also relatively uniform.

5.2.3. Cox Process Simulation

After acquiring the anticipated intensity function, the next step involves utilizing it to simulate a Cox process. The present simulation exercise relies on the implementation of the thinning method, a technique that allows for the generation of event times that exhibit a high degree of similarity with the anticipated intensity function's pattern.

The aim of the simulation procedure is to produce a novel sequence of event times that closely approximates the pattern demonstrated by the initial dataset. The aforementioned procedure can be regarded as a computational depiction of a stochastic trial. The simulation accurately replicates the chronological order of events, while strictly adhering to the guidelines set forth by the anticipated intensity function.

The simulation continues until the total duration of the simulated event times aligns with the total time span encompassed by the original dataset. This is a crucial step to ensure that the simulated data and the original data are on the same temporal scale.

Following the generation of the Cox process, a series of computational steps are performed to analyze the quality of the simulated data in relation to the original dataset.

Initially, a computation is executed to calculate the cumulative number of failures for both the original and simulated data. These cumulative failures are subsequently plotted against time to visually compare the trends of the simulated and original datasets. This graphical representation provides an intuitive comparison betIen the performance of the Cox process and the original data.

In order to enhance the precision of the simulation, 95% confidence interval lines have been computed and incorporated into the graph. The aforementioned confidence intervals furnish a range that encompasses the expected location of 95% of the data points. A broader confidence interval is indicative of an increased level of uncertainty regarding the veritable value. On the contrary, a confidence interval that is limited in scope indicates an elevated level of accuracy in the estimation.

After that, the real and simulated data's Empirical Cumulative Distribution Functions (ECDFs) are calculated. The ECDF is a technique for describing how a set of numbers, in this case the failure times, are distributed. A statistical test is then run to determine the goodness-of-fit using the ECDFs of the original and simulated data.

In particular, the non-parametric Kolmogorov-Smirnov (K-S) test is used to compare the distributions of two sample sizes. The idea that the two samples Ire taken from the same distribution is tested as the null hypothesis. An impartial gauge of the goodness-of-fit is provided by the test statistic and p-value from the K-S test.

The null hypothesis cannot be rejected if the p-value is larger than a predetermined significance level, such as 0.05. This indicates that the Cox process and the initial data fit each other Ill. In contrast, a p-value below the level of significance would indicate a poor fit. This thorough study makes it easier to evaluate the Cox process' effectiveness in simulating the initial failure data.

5.2.4. Optimisation of Cox Process

The successful application of the Cox process in the simulation must be carefully contextualized within the framework of the uncertainties inherent in Gaussian Process Regression. A Radial Basis Function (RBF) kernel was utilized, a common choice due to its wide-ranging use and robust theoretical underpinnings. Nevertheless, the RBF kernel predicates on smooth, non-linear data correlations. In the event that the underlying process deviates from these criteria, exhibiting a non-smooth nature or different non-linear characteristics, the model's fit could be compromised, leading to an increased level of uncertainty.

Further, the function $model.optimize_restarts$ seeks optimal kernel parameters that minimize a certain cost function, a process often reliant on numerical optimization techniques. These techniques, while effective, are not infallible and can occasionally lead to convergence at local rather than global minima. The model mitigates this risk by initiating the optimization process from ten distinct starting points ($num_restarts$=10), which aims to enhance the possibility of locating the global optimal solution. However, there still persists a degree of uncertainty regarding the achievement of the optimal solution. Moreover, the simulation of the Cox process requires the generation of pseudorandom numbers following various distributions (Poisson, exponential, uniform), and while contemporary random number generators perform reliably, they fall short of generating truly random numbers, which introduces another layer of uncertainty.

An important progression in the revised code is the integration of a Monte Carlo methodology via multiple simulations (*n_simulations*=1000), enhancing the scope of the exploration of the stochastic attributes intrinsic to the Cox process. This technique facilitates the acquisition of a comprehensive understanding of potential variability, hence enhancing the robustness of the simulation analysis.

In this amended script, the most representative simulation is identified based on the p-value from the Kolmogorov-Smirnov (K-S) test and the mean squared error (MSE) between the simulated and actual cumulative failure data. This progression furnishes a quantifiable measure for choosing the most accurate simulation, thereby guaranteeing a more robust congruence with the observed empirical data.

Expanding this to the three practical examples presented in Section 4.3.1, I see the versatility of this

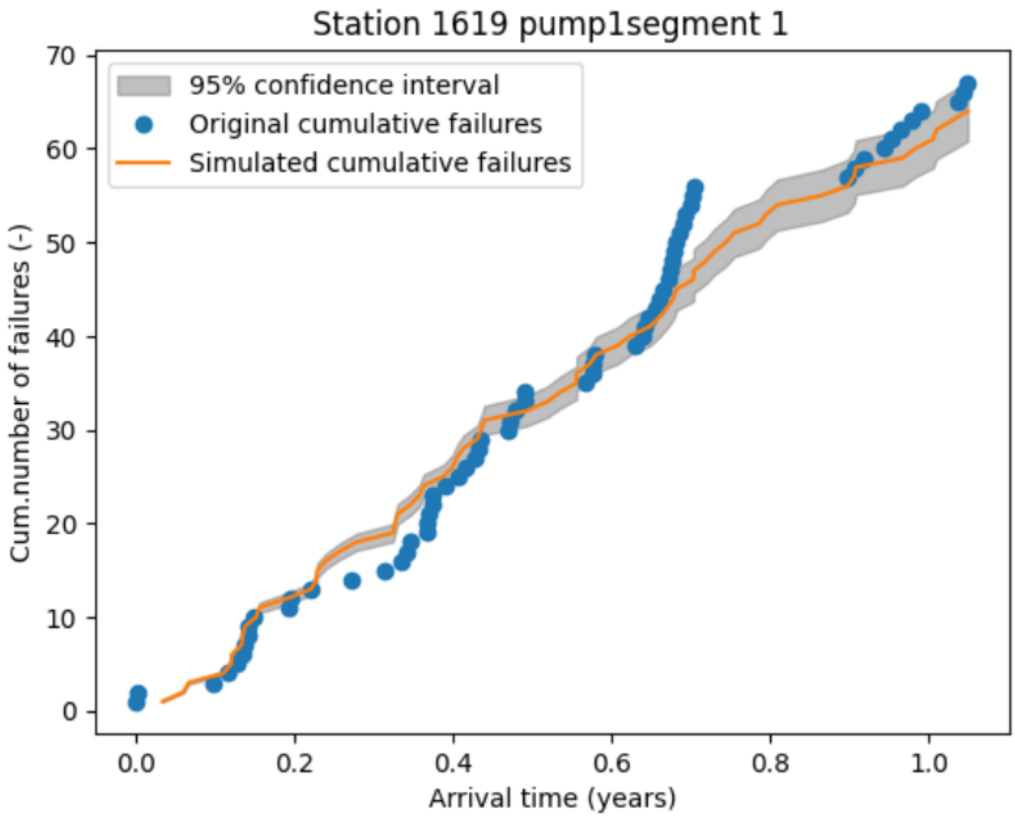


Figure 5.5: Cox Process simulation for Station 1619 Pump 1 Segment 1

approach. The first example is Station 1619 Pump 1 Segment 1, as can be seen in Figure 5.5. This particular example exhibits a fit that is superior to that achieved by Cox's log-linear model.

The second example is Station 566 Pump 2 Segment 4, which is detailed in Figure 5.6. The increase in the number of failures here is quite dramatic, and this rapid escalation is not captured accurately by the modeling process. However, even with this shortcoming, the fit is significantly better than the Cox's log-linear model, demonstrating the flexibility and adaptability of the modeling process.

However, considering the third example (Station 560 Pump 2 Segment 1), it's evident that the accuracy of the Crow's model has already been confirmed in the preceding section. Hence, in this particular scenario, there's no necessity to incorporate the Cox process. The superior performance of the Crow's PLP model in accurately capturing the situation attests to its self-sufficiency, negating the need for supplemental processes like the Cox model.

Model/Process	Number of Segments
Upward Trend	275
Downward Trend	238

Table 5.3: Summary of Final NHPP Modelling Results

Following the previously discussed examples, another case comes into focus, providing a clear illustration of the value derived from optimizing the Cox process. In the case of Station 1619 Pump 1 Segment 3, both the log-linear and the Crow's Power Law Process (PLP) models fall short in their fitting accuracy. The Mean Squared Error (MSE) scores - 39.08 for the log-linear model and a considerably higher 183.05 for the Crow's PLP model - further underline the lack of optimal fit provided by these models. The relatively large MSE values signify the discrepancy betlen the observed data and the values pre-

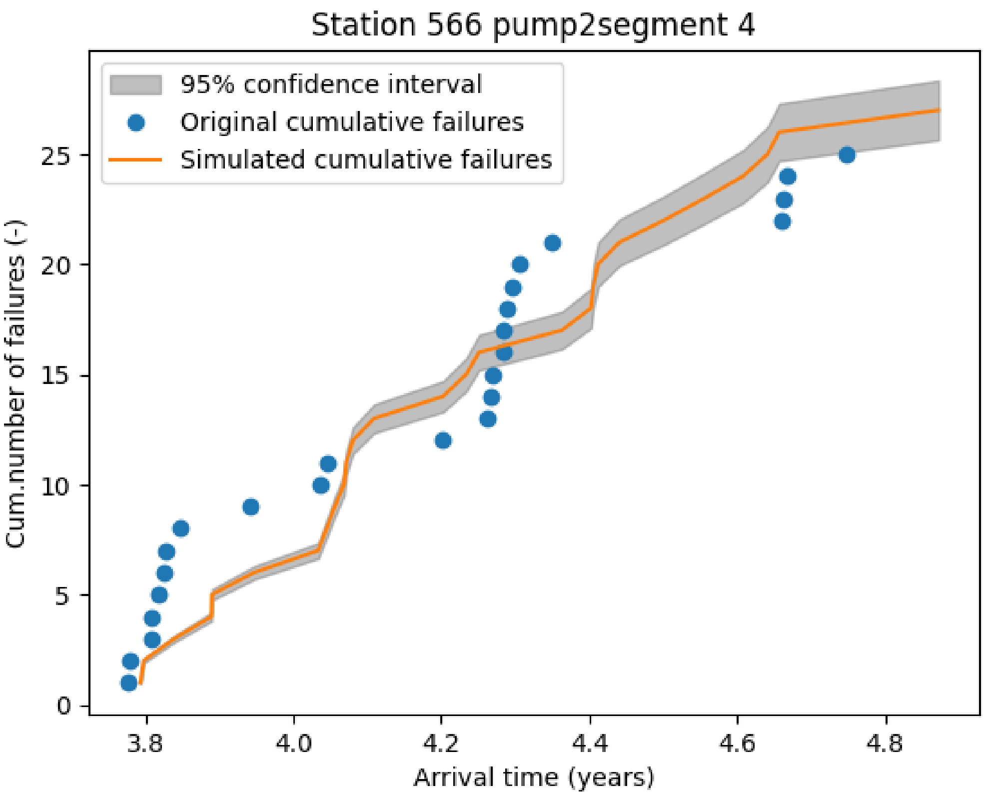


Figure 5.6: Cox Process simulation for Station 566 Pump 2 Segment 4

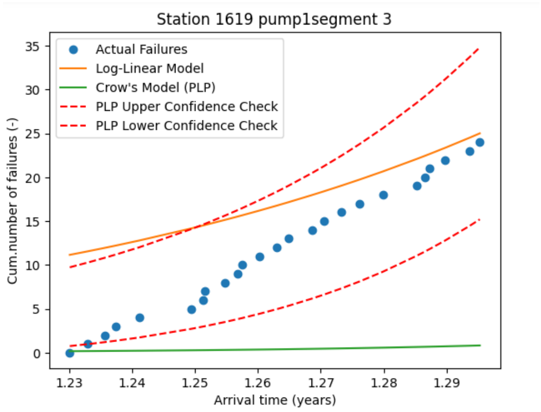


Figure 5.7: Cox Process simulation for Station 1619 Pump 1 Segment 3

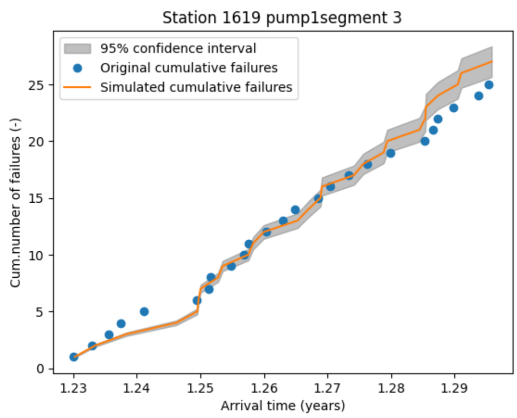


Figure 5.8: Comparison of Crow's and Cox-Lewis' model with observed failures of Station 1619 Pump 1 Segment 3

dicted by these models, highlighting their limitations in this particular instance. Despite the original models offering a certain degree of precision, their fit could not be described as optimal, see Figure 5.7. However, the implementation of the Cox process presents a marked improvement. The fit achieved by the Cox process aligns much more closely with the observed data, underscoring the utility and value of integrating the Cox process in such contexts. The significantly better fitting result achieved through the Cox process optimization is made evident in Figure 5.8.

This added example further emphasizes the adaptable nature of the modeling strategy. It showcases that, even though the Crow's Power Law Process (PLP) and Cox's log-linear model possess the ability to perform effectively as a standalone tool, there are instances where coupling it with an optimized Cox process leads to a more precise fit with empirical data. Therefore, the approach can adapt to a variety of conditions, applying different methodologies as necessary to secure the most accurate outcomes. However, the Cox process has been a successful tool in many instances, but it also comes with its fair share of constraints.

Following a thorough analysis, it was observed that 275 segments showed an upward trend, while 238 segments displayed a downward trend, as detailed in Table 5.3. This contrasts with the results derived from both lognormal and power law Non-Homogeneous Poisson Process (NHPP) modelling, which respectively identified 459 segments with an increasing trend and 51 segments with a decreasing trend.

To further elucidate these results, specific trends across various segments and their corresponding fitting models are outlined. For Station 1619, Pump 1, four distinct segments were examined, see Table B.1 in Appendix B. Segment 1 is characterized by a decreasing trend and was effectively modeled by the Cox process. Segment 2, in contrast, reveals an increasing trend and is also accurately modeled by the Cox process. Similarly, Segments 3 and 4 both exhibit an increasing trend and are fittingly modeled using the Cox process.

In the case of Station 560, Pump 2, four segments were also analyzed, see Table B.2 in Appendix B. Segment 1 and Segment 4 display an increasing trend. The former was successfully modeled using the log-linear model, while the latter fits well with the Crow's model. For Segment 2 and Segment 3, RP modeling was employed, resulting in a fitting Weibull distribution for Segment 2 and a gamma distribution for Segment 3.

Lastly, for Station 566, Pump 2, three segments were inspected, see Table B.3 in Appendix B. Segment 1 exhibits a decreasing trend and was accurately modeled by the log-linear model. On the other hand, both Segment 2 and Segment 4 reveal an increasing trend, and both were successfully modeled using the Cox process.

In summary, the models chosen for each segment align well with their respective trends, demonstrating the practicality and versatility of employing different models depending on the observed trend. This section thus offers a comparative perspective on trend detection and the suitability of different modeling approaches for each segment.

6
Discussion

6.1. Discussion on Trend Analysis Optimization

The optimization of trend analysis was a crucial component of this study. My initial application of piecewise linear regression proved insufficient for a significant portion of the pumps under examination. Specifically, this method failed to effectively analyze 50% (or 218 pumps) of the total. This finding necessitated a revision of my methodology.

Consequently, I incorporated the Bayesian Information Criterion (BIC) into my approach. The BIC provided a more nuanced balance between model complexity and its fit to the data, thereby mitigating issues of overfitting or underfitting. This strategic inclusion of BIC, along with its associated calculation and segment optimization functions, was instrumental in enhancing the accuracy of my trend identification process.

The revised methodology yielded reliable trend analysis for 19 pumps. In particular, the failure analysis of station 1621 pump 1 demonstrated an evident trend with approximately 8 cumulative failures occurring within 0.1 years. This outcome attests to the effectiveness of my improved methodology.

Nevertheless, my approach faced a significant challenge when 199 pumps from the initial dataset failed to exhibit any discernible trends, even after the application of the optimized methodology. This issue prompted the development of a more innovative secondary method, termed subjective trend analysis.

This novel method prioritized the identification of potential trends among pumps initially showing no evident trends. The application of this secondary method facilitated the discovery of additional patterns and trends that were not detectable using the primary method.

The introduction of these methodological advancements significantly improved the categorization of trends. Although there were 199 pumps that initially showed no discernible trend, the application of the secondary methodology reduced the number of unclassified pumps to 110. This considerable reduction illustrates the efficacy of the implemented methodological improvements

6.2. Discussion on Small Data Set Reliability Analysis

Analyzing small datasets in reliability studies poses its own set of challenges. Units with fewer than four failures are categorized as having a small dataset. This straightforward definition underscores the intricacies involved in extracting meaningful insights from such limited data.

A frequently encountered issue in reliability data analysis is the dearth of substantial data. As highlighted by Bendell (Bendell, 1998), conventional statistical methods may not be ideally suited for small datasets due to their restricted information content. This often results in outcomes that are less than satisfactory in terms of reliability and accuracy.

In light of these challenges, the Bayesian method emerges as a promising alternative. This technique hinges on the utilization of prior information, which can range from historical data and system estimates to expert insights. Such information forms the basis of a prior distribution. When infused with new data, this distribution evolves, resulting in a posterior distribution. The strength of the Bayesian approach lies in its ability to harmonize historical data with new findings, presenting a comprehensive view of reliability (Garmabaki et al.,2016).

Although the Bayesian method wasn't applied in this context, its potential merits in navigating the challenges posed by small datasets cannot be overlooked. For instance, units with scanty failure data, either due to data collection anomalies or inherent high reliability, could be illuminated more effectively through the Bayesian lens.

However, every method, including the Bayesian approach, has limitations. Units with minimal failure data can still be challenging to analyze. Refining the Bayesian method or considering other approaches may be beneficial in future studies.

In short, small datasets pose analytical challenges, but methods like the Bayesian approach offer potential for deriving solid insights, blending both historical and recent data.

6.3. Comparative Assessment of Initial and Optimized Methodologies in Pump Failure Analysis

A thorough comparison of the primary and optimized applications of the Non-Homogeneous Poisson Process (NHPP) and Renewal Process (RP) models elucidates their respective efficacies in the exploration of pump failure data. Initially, the NHPP modeling facilitated the categorization and examination of 351 segments through the log-linear model, while the Power Law Process (PLP) model was applied to 106 segments. These models disclosed upward trends in 415 segments and downward trends in 42 segments, thus underscoring the multifaceted nature of failure dynamics across heterogeneous pump segments.

The analysis of these trends revealed the necessity of precise model selection. Particular instances were identified where the Cox log-linear model was the only fit, suggesting the presence of unique operational dynamics in these segments. Such specificity was particularly evident in three segments, which failed to conform to Crow's model.

The RP model, due to its suitability in scenarios where Time Between Failures (TBF) data demonstrated no correlation, further emphasized the importance of adaptability and flexibility in model selection.

Upon the implementation of the optimized trend analysis methodology, the modeling process was significantly enhanced. The revised NHPP modeling allowed the log-linear model to fit 388 segments, and the PLP model was used for 122 segments. Notably, an increase in the number of segments demonstrating both upward and downward trends was observed, thus reflecting the effectiveness of the optimized approach.

The refined methodology continued to highlight the requirement for diverse models to accurately capture various failure dynamics across segments. A substantial increment in the number of segments adhering to different distributions in RP modeling was identified, which served to reaffirm the complexity of failure dynamics.

A key element of the refined methodology was the inclusion of subjective trend analysis, which facilitated the effective identification and characterization of additional segments. Specifically, in the case of Station 1613 Pump 1 Segment 3, this technique confirmed its practicality and demonstrated its potential in advancing reliability analysis.

However, despite the noted improvements, some challenges were identified within the NHPP modeling realm. Certain models continued to display significant deviations from empirical data, with broad intervals in some instances suggesting the existence of clustering. This raises queries about the applicability of traditional growth models, such as Crow's and Cox-Lewis', primarily due to their underlying assumption of a monotonically increasing or decreasing failure rate over time.

The study also identified that, in situations of significant failure surges, the initial NHPP model might not effectively emulate the instantaneous failure rate. This indicates a demand for more versatile modeling techniques, such as the Cox process. With its utilization of a stochastic intensity function, the Cox process potentially provides enhanced adaptability in encapsulating non-monotonic patterns and accommodating dynamic failure rates.

In summary, both the initial and optimized implementations of NHPP and RP models have proven instrumental in pump failure data analysis. A comparison between these two approaches emphasizes the essential need for continuous refinement and innovation in modeling methodologies for an accurate representation and interpretation of the dynamics of pump failures.

6.4. Balancing Between Traditional Models and the Cox Process

The utilization of the Cox process in the current study provides an innovative methodology to address challenges associated with fluctuations in failure rates, offering a compelling alternative to traditional methods. This novel approach leverages the power of Gaussian Processes (GP) to capture complex, non-linear correlations in data, paving the way for more accurate prediction of the intensity function.

GP fitting proves particularly advantageous for time series data, especially those exhibiting variable intensity, demonstrating a substantial upgrade over previous approaches. This aptitude resonates with the purpose of the Cox process, augmenting its relevance to this study. The versatility of GPs, as well as their capacity to model uncertainty, contribute to a smoother and more efficient data analysis process.

However, the use of GPs and the Cox process does not come without its challenges. Although the GP model was applied successfully with the GPy library and optimized using the model.optimizer restarts() function, the choice of kernel, in this case, the RBF kernel, raises questions about the model's universal applicability. The RBF kernel assumes smooth, non-linear correlations which may not always hold true for all types of data. As a result, instances where the underlying data exhibits different characteristics could potentially compromise the model's accuracy.

Furthermore, the process of optimization, despite being enhanced by multiple restarts, does not guarantee the achievement of global optima, given the inherent uncertainty associated with numerical optimization methods. This uncertainty is further compounded by the simulation of the Cox process, which involves the generation of pseudorandom numbers following various distributions.

Nevertheless, the implementation of a Monte Carlo methodology introduces a level of robustness to the process. By conducting multiple simulations, this study explores a broad range of stochastic attributes intrinsic to the Cox process. This extensive exploration not only mitigates the uncertainties related to the Gaussian Process Regression but also provides a comprehensive understanding of the potential variability in the model's performance.

The model's adaptability is further demonstrated by the variations in the quality of fit achieved across different datasets. While the model outperforms the Cox's log-linear model in certain cases such as Station 1619 Pump 1 Segment 1 and Station 1619 Pump 2 Segment 4, it falls short of the Crow's model for Station 560 Pump 2 Segment 1. This variation underscores the necessity of tailoring the approach to the unique characteristics of each dataset, further reinforcing the versatility of the Cox process.

However, the implementation of the Cox process doesn't always yield superior results. In cases such as Station 1619 Pump 1 Segment 3, traditional models like the log-linear and Crow's PLP models, despite their limitations, provided more accurate fits. This reinforces the fact that while the Cox process and GP modeling add substantial value, they should not completely replace traditional methods.

In conclusion, the Cox process represents an exciting addition to the toolbox of techniques for dealing with rapidly changing failure rates. Although it comes with its own set of limitations and uncertainties, when utilized effectively and in combination with other approaches, it has the potential to provide improved insights and more accurate predictions. This study underscores the importance of adopting a flexible, multi-method approach to data analysis in order to better capture the complexities of failure events.

6.5. Discussion on Trend Detection and Model Suitability

The analysis performed in this study demonstrates the value of selecting fitting models that align with observed trends in the data, particularly the interarrival times of failure events. For each station and pump, the trends observed in the segments, and the models selected to represent them, provide insights into the nature of the underlying processes and highlight the efficacy of the various modeling techniques employed.

For Station 1619, Pump 1, all four segments showed variations in interarrival time trends. Segments 1, 2, 3, and 4 all found the Cox process to be the most suitable model. This successful alignment might be attributed to the Cox process's ability to model non-homogenous Poisson processes, adeptly capturing the changes in interarrival times. The Cox process, with its ability to represent a range of intensity functions, is particularly effective for data sets with considerable variability, as seen in these segments.

Station 560, Pump 2, presented a different scenario. Here, Segments 1 and 4 displayed an increasing trend in interarrival times, suggesting longer intervals between failures. The log-linear model fit Segment 1 well, possibly because this model effectively represents systems improving over time, which aligns with the increase in interarrival times. Segment 4 was best represented by the Crow's model, an extension of the log-linear model that accounts for early failures and aging, indicating a more complex underlying failure process.

For Segment 2 and Segment 3, RP modeling was employed, fitting a Weibull distribution and a Gamma distribution respectively. This suggests that for these segments, the interarrival times followed these specific distributions. The Weibull distribution, known for its flexibility, can model various failure rates, which could explain its successful application in Segment 2. The gamma distribution, conversely, could have been suitable for Segment 3 due to its ability to model a variety of skewed data, potentially capturing unique characteristics in the interarrival times.

Finally, at Station 566, Pump 2, Segment 1 displayed a decreasing trend in interarrival times, which was accurately modeled by the log-linear model. This could suggest that the system experienced a "learning period," where improvements led to a reduction in failure frequency over time. Segments 2 and 4, on the other hand, revealed an increasing trend in interarrival times, and both were successfully modeled using the Cox process. This once again underscores the versatility of the Cox process, particularly for modeling non-stationary Poisson processes, where the intensity function changes over time.

Furthermore, delving deeper into specific instances, such as Station 1619, Pump 1, yields valuable, actionable insights. A cross-tabulation analysis was performed to understand the causes of failure and their corresponding solutions across all segments for this pump, see Figure 6.1 and Figure 6.2. The analysis unveiled a consistent pattern across all segments of the pump. 'Dirt' emerged as the primary cause of failure in every instance. Concurrently, 'cleaning' was identified as the suggested solution, indicating a strong correlation between the failure cause and its solution. This uniformity of failure cause and solution across the pump's four segments is compelling, especially considering the differing interarrival times and the corresponding models applied to each segment. It suggests that the cyclical trends observed in the pump's performance might be heavily influenced by the intervals or effectiveness of the cleaning activities. This consistency across varying interarrival times and models points to the significant impact of regular maintenance activities on the pump's performance.

6.6. Comparative Analysis of Pump 1 and Pump 2 at Station 560

A noteworthy aspect of our study is the observed similarities in the failure trends between Pump 1 and Pump 2 at Station 560. The analytical procedure employed scatter plots to depict the cumulative number of failures against time for each pump, effectively illustrating their respective failure trajectories over time, see Figure 6.1.

One important insight derived from this investigation is the apparent congruence in cumulative failure trends between the two pumps. Despite being separate units with potentially different operational histories, both pumps manifest similar failure patterns over time, suggesting a commonality in the underlying causes or contributing factors influencing their respective failure rates.

An interesting divergence, however, is seen in the initial spike in failures observed in Pump 1. This could be construed as a "burn-in" phase, where the pump is subject to an initially high failure rate that subsequently diminishes over time. This is a characteristic often noted in newly commissioned or significantly refurbished systems where initial glitches or issues are rectified during the early stages of operation.

The congruent failure patterns in both pumps, with the exception of the initial surge in Pump 1, might imply that the pumps are exposed to similar operational conditions and stresses. This could also indicate that the maintenance activities, including cleaning as indicated in the previous analysis, are executed with comparable frequency and effectiveness across both pumps.

While the cumulative failure analysis for the pumps at Station 560 demonstrated certain similarities, it is crucial to acknowledge the contrasting scenarios presented by the interarrival time trends.

For Pump 1, the patterns in interarrival times suggest a multi-phased trajectory of pump performance. The initial segment indicates an early stage characterized by a decreasing trend in interarrival times, a dynamic commonly associated with the 'burn-in' period. This is followed by segments indicating increasing trends, suggestive of a gradual escalation in failure rates. This distinction in trends across segments underscores the complex and multi-faceted nature of pump performance over time, see Table

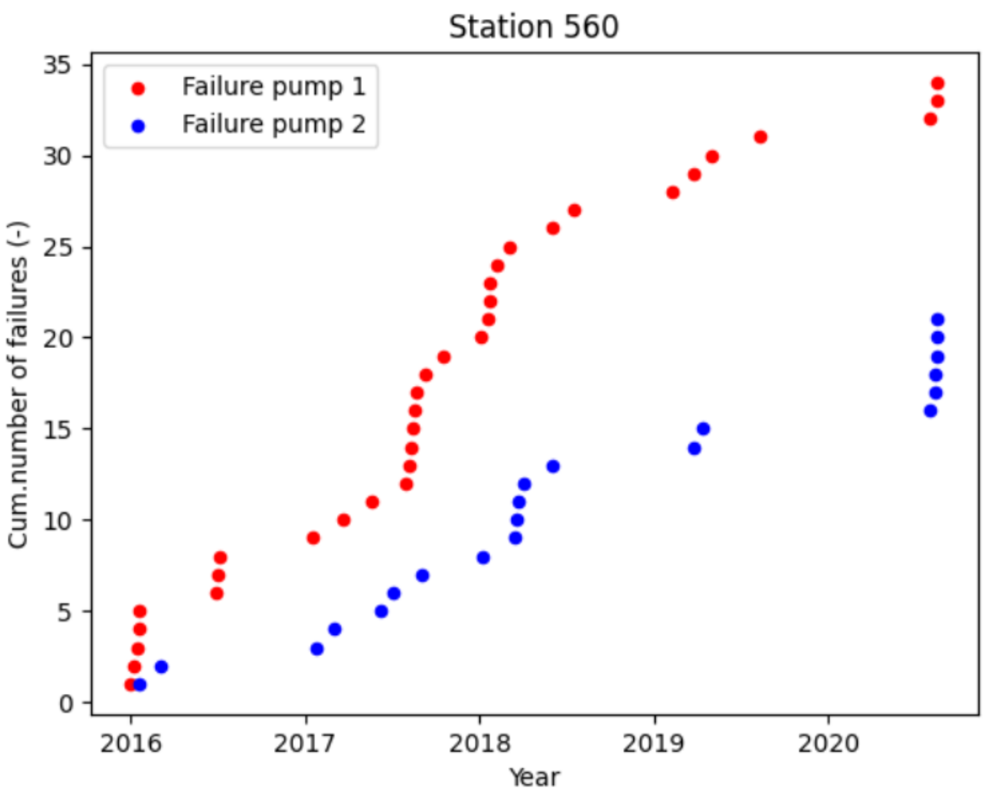


Figure 6.1: Cumulative Failures of Pumps at Station 560

B.5 in Appendix B .

Conversely, the interarrival time trends of Pump 2 unfold a different narrative. Both the first and fourth segments exhibit an increasing trend, suggesting extended intervals between failures. However, the models that best fit these trends are different, indicating unique underlying processes influencing each segment's failure pattern. The second and third segments further complicate the picture, with their interarrival times best described by two different distributions within the Renewal Process (RP) model.

6.7. Comparative Analysis of Pump 1 and Pump 2 at Station 566

The dynamics of pump performance at Station 566 were also subjected to rigorous analysis, and the assessment revealed a similar degree of complexity as observed at Station 560, albeit with a few distinct patterns, see Figure 6.2.

Pump 1's trend analysis suggests a relatively simplified scenario. The Renewal Process (RP) model fitted with a Beta distribution was deemed most appropriate, indicating that the interarrival times for this pump could be best described by this specific distribution.

Conversely, Pump 2's performance dynamics at Station 566 presented a more complex scenario. Across three segments, two distinct trends were identified. The first segment indicated a decreasing trend in interarrival times, which was aptly captured by the log-linear model. This pattern could be interpreted as a learning or "burn-in" period for the pump, leading to a decrease in failure frequency over time. Conversely, both the second and fourth segments revealed an increasing trend, accurately modeled by the Cox process, suggesting a gradual rise in failure rates over time.

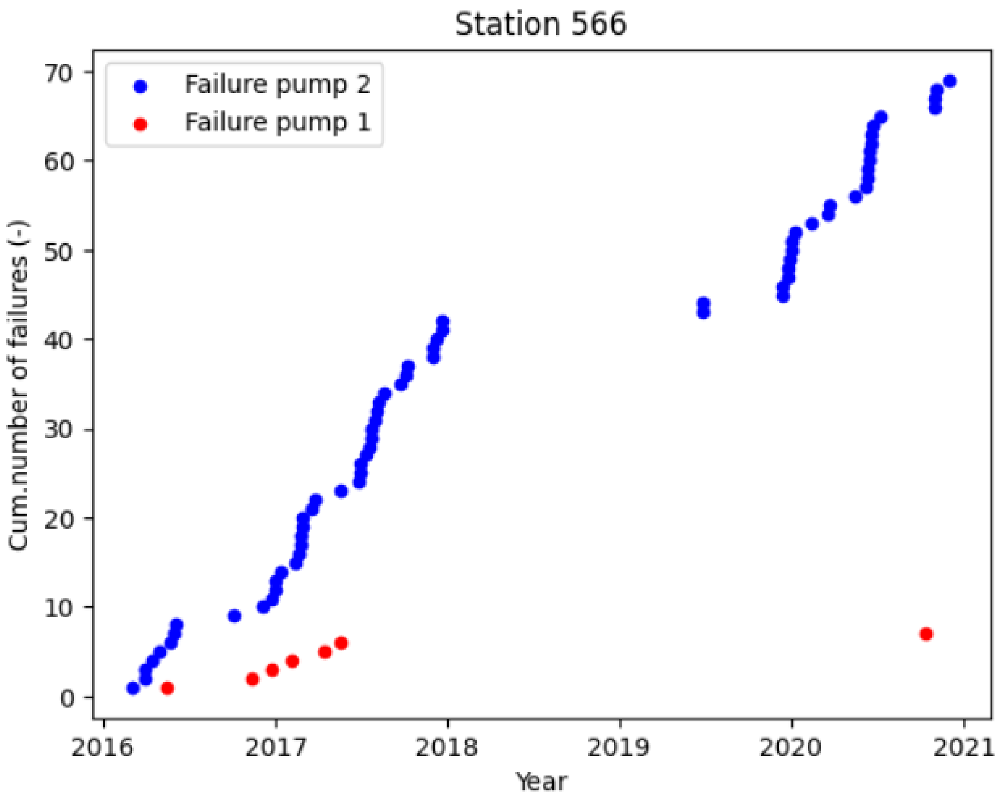


Figure 6.2: Cumulative Failures of Pumps at Station 566

This contrasting representation of trend detection between the two pumps at Station 566 emphasizes the unique characteristics and performance dynamics of each pump. It also underscores the versatility and applicability of different modeling processes in capturing these complex dynamics. Ultimately, these insights could guide more targeted and effective maintenance and management strategies for each pump.

6.8. Statistical Analysis of Five-Year Pump Performance Patterns

The detailed analysis of pump performance across different segments provides essential insights into the dynamics of pump failure patterns. However, these findings must be contextualized within the typical life cycle of pump systems and their maintenance schedules.

Pumps and their electrical equipment are often subject to renovation cycles spanning 15 years, with service intervals as frequent as once per year. Given this context, it becomes crucial to consider pump failure dynamics over a longer and more consistent timeframe.

One significant question that arises from this perspective is whether pumps improve, stay the same, or deteriorate over time. While the detailed segment analysis yields microscopic insights, a broader view helps to discern macroscopic trends over a longer timescale, such as five years. This larger view aids in identifying consistent failure patterns over this timeframe.

From an asset management standpoint, consistent failure patterns are generally ideal as they offer predictability, facilitating streamlined planning for maintenance and renovation activities. On the other hand, different failure patterns, especially those that fluctuate, increase complexity. As demonstrated and discussed previously, these diverse patterns introduce uncertainty, making them more challenging to manage. They often necessitate more detailed studies to decode the underlying causes of such variations and to devise appropriate strategies for managing them.

An analysis was undertaken to examine the patterns of pump failure over a span of five years, specifically targeting segments that displayed consistent failure trends. In this context, 'consistent trends' refer to the instances where all the identified segments demonstrated the same trend over time. The analytical approach yielded the following results:

Overall, the total duration of operation across all pumps was 100,354 hours, with an average duration of 395 hours. Within this duration, 254 pumps were analyzed, with a total of 3,211 pump failures observed.

In terms of individual pump performance, Pump 1 at Station 512 displayed the maximum number of failures, with a total of 123 failures. On the other hand, Pump 2 at Station 511 showed the minimum number of failures, totaling 4 failures. The average number of failures across all pumps was approximately 12, see table 6.2.

Pump at Station	No. of Failures	Total Duration (hours)
512, Pump: 1	123 (maximum)	3993 (longest)
511, Pump: 2	4 (minimum)	-
All	12 (average)	395 (average)
17, Pump: 1	-	18 (shortest)

Table 6.1: Overview of pump failures and operating durations

When comparing pump operation durations, Pump 1 at Station 512 also had the longest total duration of 3,993 hours. In contrast, Pump 1 at Station 17 had the shortest total duration of operation, with a mere 18 hours.

A specific subset of pumps, those modeled with the Renewal Process (RP), were the focus of this analysis. The RP modeling was applied to 98 pumps, reflecting those pumps that exhibited no trend with interarrival time. For these RP-modeled pumps, Pump 1 at Station 223 demonstrated the maximum number of failures (8), while Pump 2 at Station 511 again showed the minimum number of failures (4). In total, these pumps experienced 304 failures. On average, the RP-modeled pumps had around 3 failures based on the total length of operation. The pump with the longest total duration under this category was Pump 1 at Station 1633, operating for 2,011 hours. Conversely, Pump 1 at Station 17 had the shortest duration of operation (18 hours). The average duration per pump for the RP-modeled pumps was 159 hours, see Table 6.3.

Pump at Station	No. of Failures	Total Duration (hours)
223, Pump: 1	8 (maximum)	-
511, Pump: 2	4 (minimum)	-
All	3 (average)	159 (average)
1633, Pump: 1	-	2011 (longest)
17, Pump: 1	-	18 (shortest)

Table 6.2: Details of RP modelled pumps

Pump at Station	No. of Failures	Total Duration (hours)
512, Pump: 1	123 (maximum)	3993 (longest)
1469, Pump: 1	5 (minimum)	-
All	18 (average)	584 (average)
26, Pump: 3	-	40 (shortest)

Table 6.3: Details of NHPP modelled pumps

Non-Homogeneous Poisson Process (NHPP) models provided valuable insights for pumps with increasing and decreasing trends. Among the pumps modelled with NHPP, 146 exhibited an increasing trend in interarrival times, suggesting an improvement in pump performance over time. In contrast, only 10 pumps showed a decreasing trend, indicating a potential deterioration of pump performance. Pump 1 at Station 512 reported the highest number of failures (123), while Pump 1 at Station 1469 reported the least (5). The average number of failures per pump, based on the total length of operation, was approximately 18. In terms of operation duration, Pump 1 at Station 512 exhibited the longest duration of 3,993 hours, while Pump 3 at Station 26 had the shortest duration of 40 hours, see Table 6.4. This preponderance of pumps with increasing interarrival times is beneficial from a system perspective, indicating a trend towards improved performance and potentially reduced maintenance needs.

Following the in-depth analysis of pump performance, attention is turned specifically to Pump 1 at Station 512, identified as the pump experiencing the most problems. Detailed inspection revealed four distinctive segments of operation for this pump. The specifics of these segments, including the number of failures, the time period, and the average failure duration, are elaborated in Table 6.5 and Table B.4.

Segment	No. of Failures	Period	Average Failure Duration (hours)
1	19	2016 - 2016-07-12	34
2	32	2016-07-12 - 2018-09-21	35
3	59	2018-09-21 - 2020-03-20	31
4	13	2020-03-20 - 2020-12-27	31

Table 6.4: Failure Characteristics of Pump 1 at Station 512

The results from the detailed five-year timeframe analysis indicate consistent failure patterns across the pumps at different stations. These findings hold significant value for asset management, as they help identify both the pumps and stations that require additional attention and those that demonstrate reliable performance. Further, the identified trends provide a better understanding of pump failure dynamics, guiding more effective and targeted maintenance and decision-making processes.

6.9. Comparison with Prior Studies

In terms of consistent failure patterns, a comparative study of the present investigation with the data collected by Korving et al. (2006) twenty years ago unveils some intriguing trends in pump failure dynamics. In the current analysis, the average operational duration preceding a pump failure was observed to be 395 hours, substantially less than the 642 hours documented by Korving et al. (2006). This alteration indicates a decrease in the mean time to failure over the past two decades.

Moreover, this study recorded a marked improvement in the average number of failures per pump. The data showed an average of 12 failures per pump, significantly fewer than the 62 failures per pump reported in the earlier study.

In considering the extreme scenarios, the pump with the longest failure duration in this study was recorded at 3,993 hours, which is significantly shorter than the 6,465 hours noted in Korving's research. Likewise, the pump with the highest frequency of failures in this investigation encountered issues 123 times, a count considerably lower than the 364 failures documented in the previous study.

A comparative examination was made between a pump specifically highlighted in Korving's analysis and the corresponding pump in the present study, namely Pump 1 at Station Rm5. The data from both analyses are presented below:

Exploring the failure characteristics of Pump 1 at Station Rm5 in both studies, several fascinating insights are revealed. Significantly, an improvement in the pump's reliability over the two periods between the studies is indicated by the decrease in the average annual failure from 8.4 in Korving's study to 4.4

Table 6.5: Korving et al.: Failure Characteristics of Pump 1, Station Rm5

	Failures (1/year)	Failure Duration (h)	Operation Duration (h/year)
1998	2.0	19.2	38.4
1999	15.0	14.6	218.7
2000	7.0	18.5	129.8
2001	11.0	7.0	77.3
2002	7.0	13.6	95.1
Average	8.4	14.4	111.9

Table 6.6: Present Study: Failure Characteristics of Pump 1, Station Rm5

	Failures (1/year)	Failure Duration (h)	Operation Duration (h/year)
2016	6.0	16	96
2017	1.0	18	18
2018	7.0	32	224
2019	7.0	54	378
2020	1.0	23	23
Average	4.4	28.6	147.8

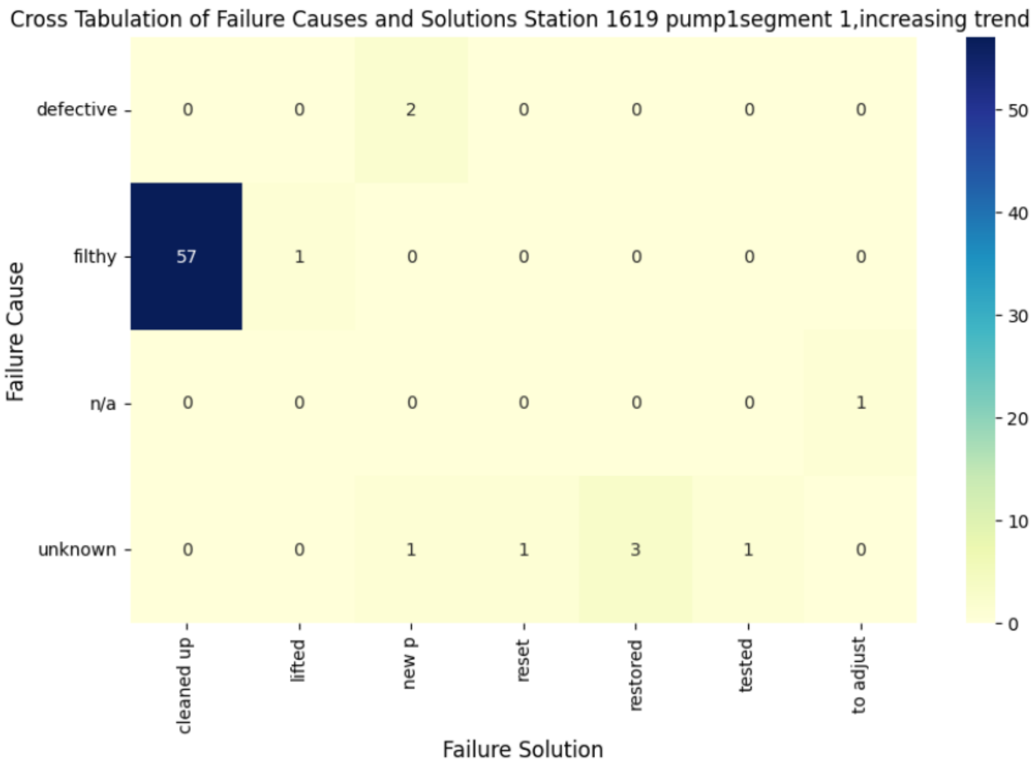
in the current analysis. This demonstrates the effectiveness of the present study's approach.

Moreover, the trend analysis and model application in the current study identified three segments with increasing trends, signifying fewer failures per year over the five-year period. This further substantiates the notion of enhanced pump reliability over time

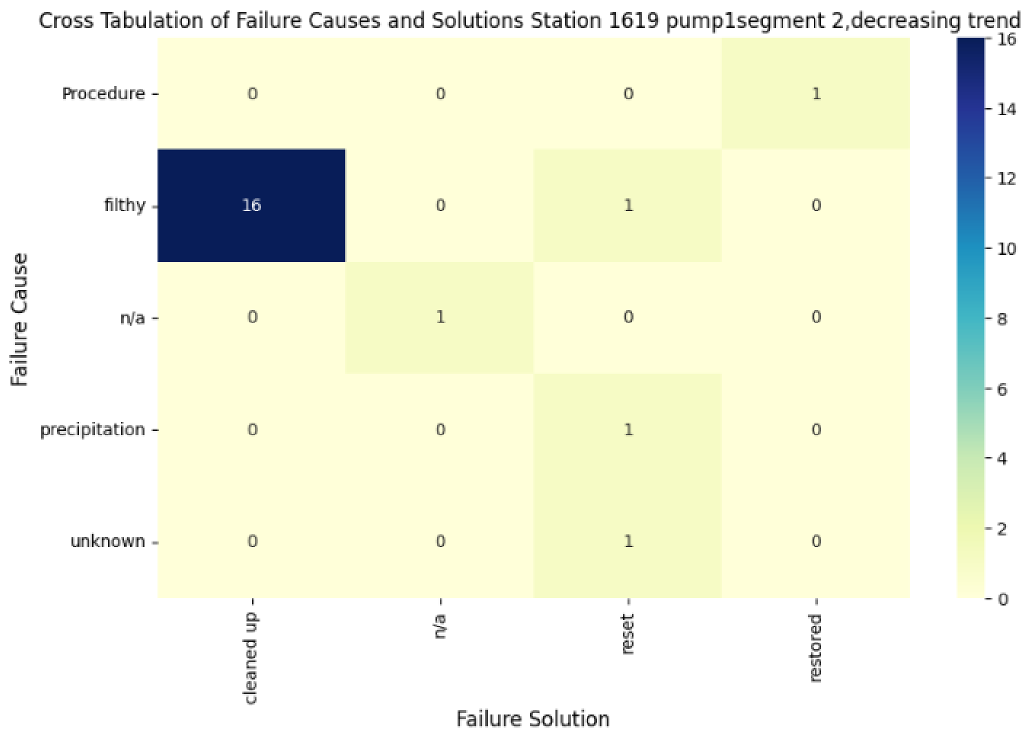
However, the average failure duration displays an opposing trend. In the current study, the average failure duration for the pump escalated from 14.4 hours as reported by Korving et al. (2006), to 28.6 hours. Although the pump's reliability appears to have improved, with less frequent failures, the prolonged average duration of each failure incident could negate the benefits achieved from the lower failure rate. This suggests that, while failures may be less common, they take more time to address when they do occur. This increase might be a result of more complex failure modes encountered or alterations in maintenance protocols over the years.

Additionally, the total operational duration per year for the pump in the current study significantly surpasses that reported by Korving et al. (2006). While the latter's research showed an average of 111.9 hours/year, the current analysis indicates an annual duration of 147.8 hours. This discrepancy may be due to potential inaccuracies in recording the failure end times. If these times are erroneously noted as later than the actual, it would invariably overestimate the duration of the failure and the total operational time. Furthermore, this could indicate that the relevant authorities did not perform maintenance in a timely manner, or the methods of maintenance require further enhancement. Therefore, there is a need to intensify supervision to ensure accurate recording and prompt, effective maintenance.

Given these insights, it appears that the improvements in pump reliability may be somewhat counterbalanced by an increase in failure duration and total operational time. This finding underlines the importance of comprehensive asset management strategies, which take into account not just the frequency of failures but also their duration and the total operational time. Further studies could investigate the root causes of the increased failure duration, which would be beneficial for refining maintenance strategies and optimizing asset management practices.

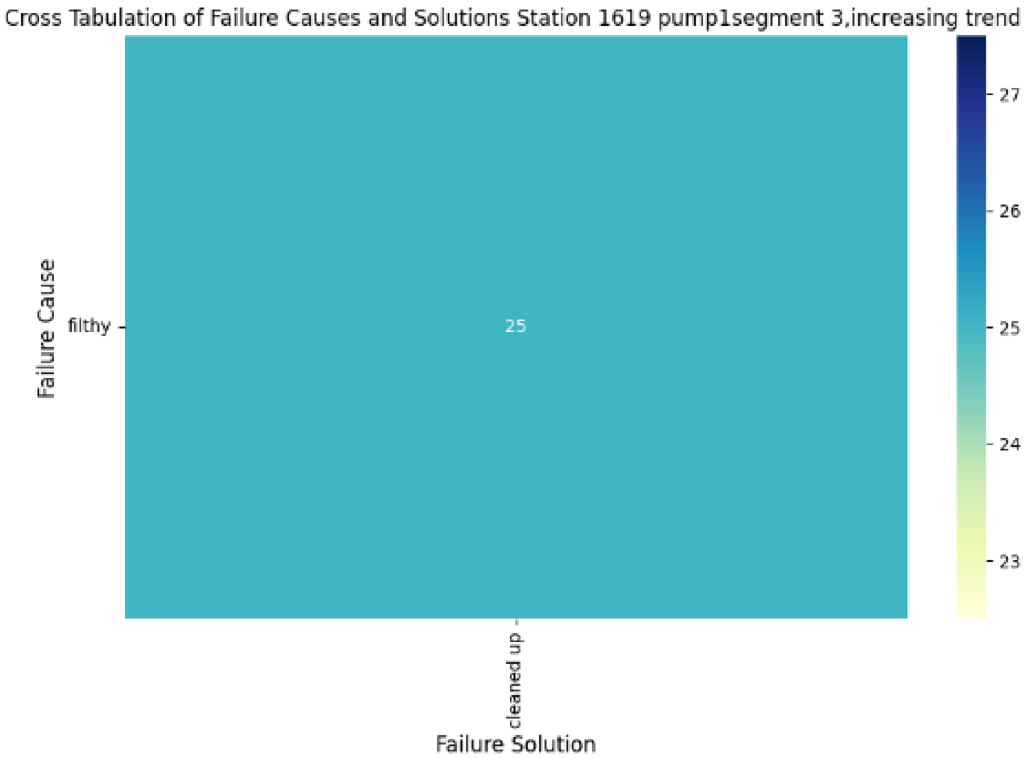


(a) Segment 1

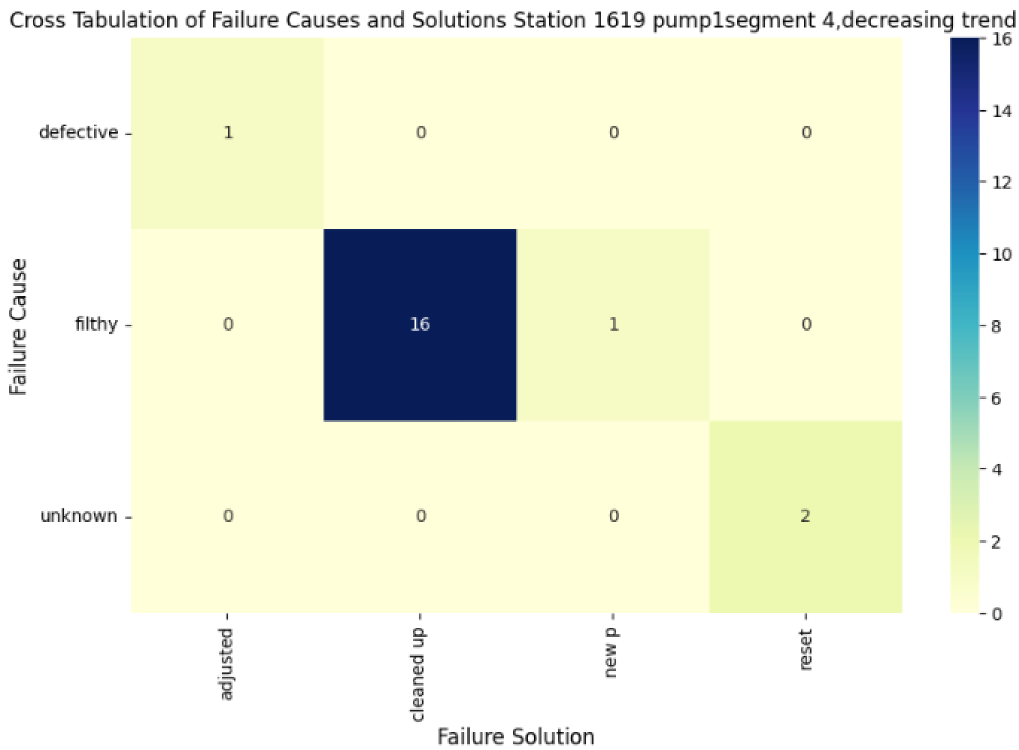


(b) Segment 2

Figure 6.3: Cross Tabulation of Failure Causes and Solutions for Station 1619 Pump 1 (Part 1)



(a) Segment 3



(b) Segment 4

Figure 6.4: Cross Tabulation of Failure Causes and Solutions for Station 1619 Pump 1 (Part 2)

7 Conclusions

The present research endeavor aims to investigate a range of interrelated queries, encompassing the analysis of discrepancies in failure data, the delineation and categorization of failures, the identification of alterations in failure trends across time, and the selection of the most appropriate statistical framework to depict the data. The primary objective was to provide insightful viewpoints regarding the effectiveness and reliability of pump systems.

An integral part of this investigation is the development of a comprehensive decision-making framework to evaluate the performance and efficiency of sewage pumping stations. The framework, as presented in this study, integrates objective and subjective trend analyses, multiple trend tests, and various models such as the Homogeneous Poisson Process (HPP), Renewal Process (RP), and Non-Homogeneous Poisson Process (NHPP). This framework demonstrates its effectiveness in identifying alterations in system reliability, offering valuable insights for decision-making regarding system enhancements.

Furthermore, this study made a contribution to the comprehension and improvement of maintenance procedures in sewage pumping stations, particularly in relation to the Department of Public Works located in Rotterdam. By employing statistical methodologies for the analysis of pump failure data, extensive knowledge has been obtained about the trends and patterns of such failures.

Analyses underscored that the original method, reliant on piecewise linear regression, was insufficient for about half of the pumps. Consequently, the study proposed a revised approach integrating the Bayesian Information Criterion (BIC) into methodology to enhance the model selection procedure and address overfitting and underfitting concerns. Furthermore, a secondary gradient-based approach is implemented to detect patterns in an additional 108 pumps that were not discernible through the initial method.

An examination of interarrival time trends within system failure data revealed two distinct patterns – namely, increasing and decreasing trends in the intervals between failures across diverse pump segments. Over the five-year research period, it became evident that failure trends for individual pumps could vary considerably. Case studies of three distinct pumps demonstrated the capacity for failure patterns to alternate between increasing, decreasing, and steady states over different periods. This indicates that a pump's failure behavior is dynamic, susceptible to shifts influenced by an array of factors including changing operating conditions, usage patterns, and maintenance regimes.

Furthermore, these insights hold importance for effective asset management, given that consistent failure patterns offer predictability while variable ones necessitate a comprehensive understanding and strategic management approach. Findings include the need for increased attention to Pump 1 at Station 512 due to its high failure rate. In contrast, improved reliability is observed in a subset of pumps modeled with the Renewal Process, exhibiting no trend with interarrival time. Insightful comparisons with data collected two decades ago illuminate intriguing shifts in pump failure dynamics. Despite a decrease in the mean time to failure - suggesting an increase in pump failure incidents - an overall reduction in the average number of failures per pump points towards enhanced pump reliability. However, the seeming improvement in pump reliability is contrasted with an increase in the average failure duration. This underlines the necessity for comprehensive asset management strategies.

Moreover, when comparing different pumps within the same station, both similar and different failure patterns were observed. This finding signifies that while certain failure trends may be common due to shared operating conditions or design parameters, individual differences between pumps, such as age, model, usage intensity, and maintenance history, can lead to unique failure patterns.

These observations add another layer of complexity to the issue and highlight the need for a dynamic and adaptable maintenance strategy. While substantial progress has been made in understanding the mechanics of pump failure and establishing a robust analytical framework. However, it is recognized that this research has certain limitations that could be addressed in future work.

For instance, when considering repairable systems, caution must be taken in defining 'failure'. It can range from minor issues requiring only a swift online adjustment to more serious defects that necessitate comprehensive investigations and extensive repair procedures. Importantly, the causes of failures in repairable systems are multifaceted, extending beyond mere design deficiencies, material weaknesses, manufacturing flaws, and regular wear and tear. Factors such as fluctuating environmental conditions, the interplay of different components, and human interventions by operating and maintenance personnel can also contribute to system failure.

In the context of this study, intricacies of parameter estimation posed significant challenges. Manual configuration was used for setting kernel parameters, specifically variance and lengthscale. It is important to underscore that unsuitable assumptions for these parameters can compromise the precision of predictive models. To address this, future research might consider adopting more advanced techniques for parameter estimation or tuning, such as Bayesian optimization or grid search. These strategies could further enhance the robustness of the models and ultimately contribute to the improvement of the effectiveness and reliability of pump systems.

While the modeling methodologies employed yielded results, their specific nature might have restricted the ability to capture certain subtle patterns. Future endeavors could benefit from a more comprehensive data collection, enabling the application of advanced models like Artificial Neural Networks (ANN).

The challenges of analyzing small datasets in reliability studies are undeniable. When dealing with units reporting fewer than four failures, extracting meaningful insights becomes tricky. As Bendell (1998) highlighted, traditional statistical methods often fall short with limited data, leading to potentially unreliable results. The Bayesian method offers a promising solution by merging prior information, like historical data or expert opinions, with new data. Though this study didn't utilize the Bayesian approach, its potential to address data scarcities is evident. Yet, no method is foolproof. Future research might look into refining the Bayesian technique or exploring other innovative solutions for handling sparse datasets.

However, a pertinent question arises: Do water utilities currently have enough data to facilitate these advanced models, or is there a need to amass and store more? It's important for utilities to invest in data infrastructure, not just for research, but also for real-time monitoring and decision-making.

For future research and practical application, a two-pronged recommendation emerges:

Data Collection: Water utilities should prioritize a more granular data collection strategy, encompassing various pump parameters, usage patterns, and maintenance logs. This would ensure a richer dataset to feed into sophisticated models.

Methodology Evolution: It's imperative to continuously evaluate and adapt modeling methodologies. Exploring avenues like neural networks, while also integrating traditional statistical methods, can provide a more holistic understanding of pump behaviors and failure trends."

In conclusion, the present investigation has made a contribution to the existing body of literature concerning pump malfunction and upkeep methodologies. Through the integration of statistical and machine learning methodologies, a detailed comprehension of pump failure trends and patterns has been achieved. This has resulted in the creation of a comprehensive analysis framework that is reliant on data. Even though there are signs of enhanced pump reliability, this study emphasizes the necessity for ongoing exploration and refinement of maintenance procedures. This statement highlights the dynamic nature of failure rates over time, emphasising the importance of segment analysis in gaining a comprehensive understanding of the fluctuating patterns of pump failures. This work lays a solid groundwork for future studies, aspiring towards heightened operational efficiency and refined asset management strategies. The findings imply that additional research and development could significantly improve the efficiency and reliability of pump systems, thereby strengthing the sustainability of urban water infrastructure.

References

Ascher, H., & Feingold, H. (1984). Repairable systems reliability: Modeling, inference, misconceptions and their causes.

Bartlett, M. S. (1963). The spectral analysis of point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, *25*(2), 264–281.

Bennis, S., Bengassem, J., & Lamarre, P. (2003). Hydraulic performance index of a sewer network. *Journal of hydraulic engineering*, *129*(7), 504–510.

Berardi, L., Giustolisi, O., Kapelan, Z., & Savic, D. (2008). Development of pipe deterioration models for water distribution systems using epr. *Journal of Hydroinformatics*, *10*(2), 113–126.

Burnham, K. P., & Anderson, D. R. (2002). Edition 2. model selection and multimodel inference: A practical information-theoretic approach.

Cox, D. R., & Lewis, P. A. (1966). The statistical analysis of series of events.

Crow, L. H. (1975). *Reliability analysis for complex, repairable systems* (tech. rep.). Army Materiel Systems Analysis Activityaberdeen Proving Ground Md.

Cullen, A. C., Frey, H. C., & Frey, C. H. (1999). *Probabilistic techniques in exposure assessment: A handbook for dealing with variability and uncertainty in models and inputs*. Springer Science & Business.

Ermolin, Y. A., Zats, L. I., & Kajisa, T. (2002). Hydraulic reliability index for sewage pumping stations. *Urban Water*, *4*(3), 301–306.

Garmabaki, A., Ahmadi, A., Block, J., Pham, H., & Kumar, U. (2016). A reliability decision framework for multiple repairable units. *Reliability Engineering System Safety*, *150*, 78–88. https://doi.org/https://doi.org/10.1016/j.ress.2016.01.020

Giustolisi, O., & Savic, D. A. (2006). A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics*, *8*(3), 207–222.

Glover, B., & Bhatt, H. (2006). *Rfid essentials*. " O'Reilly Media, Inc.".

Hamed, K. H., & Rao, A. R. (1998). A modified mann-kendall trend test for autocorrelated data. *Journal of hydrology*, *204*(1-4), 182–196.

Jekel, C. F., & Venter, G. (2019). Pwlf: A python library for fitting 1d continuous piecewise linear functions. *URL: https://github. com/cjekel/piecewise_linear_fit_py*.

Jin, Y., & Mukherjee, A. (2010). Modeling blockage failures in sewer systems to support maintenance decision making. *Journal of performance of Constructed Facilities*, *24*(6), 622–633.

Korving, H., Clemens, F. H., & van Noortwijk, J. M. (2006). Statistical modeling of the serviceability of sewage pumps. *Journal of Hydraulic Engineering*, *132*(10), 1076–1085.

Kvaløy, J. T., & Lindqvist, B. H. (1998). Ttt-based tests for trend in repairable systems data. *Reliability Engineering & System Safety*, *60*(1), 13–28.

Le Gat, Y., & Eisenbeis, P. (2000). Using maintenance records to forecast failures in water networks. *Urban water*, *2*(3), 173–181.

Lee, J.-Y., Won, J.-H., & Hahn, J.-S. (2006). Evaluation of hydrogeologic conditions for groundwater heat pumps: Analysis with data from national groundwater monitoring stations. *Geosciences Journal*, *10*(1), 91.

Lobanoff, V. S., & Ross, R. R. (2013). *Centrifugal pumps: Design and application*. Elsevier.

Malek Mohammadi, M., Najafi, M., Kaushal, V., Serajiantehrani, R., Salehabadi, N., & Ashoori, T. (2019). Sewer pipes condition prediction models: A state-of-the-art review. *Infrastructures*, *4*(4), 64.

Marquez, A. C., & Gupta, J. N. (2006). Contemporary maintenance management: Process, framework and supporting pillars. *Omega*, *34*(3), 313–326.

Martins, A., Leitão, J. P., & Amado, C. (2013). Comparative study of three stochastic models for prediction of pipe failures in water supply systems. *Journal of infrastructure systems*, *19*(4), 442–450.

Price, K. V. (2013). Differential evolution. *Handbook of Optimization: From Classical to Modern Approach*, 187–214.

Rausand, M., & Hoyland, A. (2003). *System reliability theory: Models, statistical methods, and applications* (Vol. 396). John Wiley & Sons.

Savic, D., Giustolisi, O., Berardi, L., Shepherd, W., Djordjevic, S., & Saul, A. (2006). Modelling sewer failure by evolutionary computing. *Proceedings of the Institution of Civil Engineers-Water Management*, *159*(2), 111–118.

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, *85*, 1–16.

Tuzson, J. (2000). *Centrifugal pump design*. John Wiley & Sons.

Xie, M., Dai, Y.-S., & Poh, K.-L. (2004). *Computing system reliability: Models and analysis*. Springer Science & Business Media.

Zacks, S. (2012). *Introduction to reliability analysis: Probability models and statistical methods*. Springer Science & Business Media.

A

Source Code Example

Listing A.1: Objective Trend analysis

```

"""
1 / 3

The function analyse_modified(df, n) in the provided Python code performs a trend analysis on failure data over years, using piecewise linear fitting to identify and plot trends. The function not only classifies each segment of the trend based on the Homogeneous Poisson Process (HPP), Non-Homogeneous Poisson Process (NHPP), or Renewal Process (RP), but it also accounts for potential outlier segments in the data with a built-in validity check based on data density and dispersion.
"""

#Objective Trend analysis logic
def analyse_modified(df, n):
    df = df.copy()
    df.loc[:, 'years'] = (df['Begin Date'] - df['Begin Date'].iloc[0]).dt.total_seconds() / (60 * 60 * 24 * 365)
    df.loc[:, 'cumulative_failure_count'] = df['Failure'].cumsum()

    x = df['years'].values
    y = df['cumulative_failure_count'].values
    max_distance = (max(x) - min(x)) / 3

    my_pwlf = pwlf.PiecewiseLinFit(x, y)
    res = my_pwlf.fit(n)
    colors = ['blue', 'green', 'red', 'orange', 'purple', 'brown', 'cyan', 'magenta']

    # Plot the fitted lines
    breakpoints = my_pwlf.fit_breaks

    #print(my_pwlf.n_segments)
    valid_segments = []

    # Add a counter to track the number of valid segments plotted for each pump
    valid_segment_counter = 0

    # Create a scatter plot with the specified labels and title
    ax = df.plot(kind='scatter', label="Failure", x='years', y='cumulative_failure_count', title=f'Station{df.iloc[0, 0]}pump{df.iloc[0, 1]}')
    ax.set_xlabel('Arrival time (years)')
    ax.set_ylabel('Cum.number of failures (-)')

    for i in range(my_pwlf.n_segments):
        x_min = breakpoints[i]
```

```
        x_max = breakpoints[i + 1]
        x_segment = np.linspace(x_min, x_max, num=1000)
        y_segment = my_pwlf.predict(x_segment)

        # Check if there are at least 3 data points in the segment
        segment_data = df[(df['years'] >= x_min) & (df['years'] <= x_max)]

        if len(segment_data) > 3:
            # Check if there are any adjacent points with distance larger than specific_distance
            segment_data_years = segment_data['years'].values
            adjacent_point_distances = np.diff(segment_data_years)
            max_adjacent_distance = np.max(adjacent_point_distances)

            # If the max_adjacent_distance is larger than specific_distance, skip plotting the segment
            if max_adjacent_distance > max_distance:
                continue
            plt.xlim(left=0)
            plt.ylim(bottom=0)
            ax.plot(x_segment, y_segment, '-', color=colors[i], label=f"Trend line {i + 1}")
            # Save the valid segment data as a DataFrame in the list
            segment_data = pd.DataFrame(segment_data)
            segment_data['order'] = f'Station {segment_data.iloc[0, 0]} pump{segment_data.iloc[0, 1]}segment {i + 1}'

            #display(segment_data)
            valid_segments.append(segment_data)
            valid_segment_counter += 1  # Increment the counter when a valid segment is plotted

    # If no valid segments are plotted for a pump, append the pump identifier to the failed_pump list
    if valid_segment_counter == 0:
        failed_pump.append(df)

    for i in range(len(valid_segments)):
        #display(valid_segments[i])
        years = valid_segments[i]['years'].values
        cumulative_failure_count = valid_segments[i]['cumulative_failure_count'].values

        # Perform the combined test
        result = categorize_units(years)

        # Categorize the results
        if result == 'NHPP':
            NHPP_segments.append(valid_segments[i])
        elif result == 'RP':
            RP_segments.append(valid_segments[i])
        else:  # 'HPP'
            HPP_segments.append(valid_segments[i])

    # Show the legend to display the labels for the trend lines
    ax.legend()

    plt.show()
```

Listing A.2: Subjective Trend analysis

```

"""
2 / 3

This Python code is designed to analyze the failure rate in a dataset, taking into account the dates of failures and the count of failures. It implements a failure rate threshold and a critical value to segment the data and uses a linear regression model to fit a trend line for each segment, thus allowing for an insightful and comprehensive trend analysis.
"""

#Subjective Trend analysis logic
```

```python
def analyze_failure_rate(df, failure_date_col, failure_count_col, failure_rate_threshold, critical_value):
    # Calculate the cumulative failure count
    df['cumulative_failure_count'] = df[failure_count_col].cumsum()

    # Calculate the difference between adjacent failure dates
    df['years'] = (df[failure_date_col] - df[failure_date_col].iloc[0]).dt.total_seconds() / (60 * 60 * 24 * 365)
    internal_threshold = (df[failure_date_col].iloc[-1] - df[failure_date_col].iloc[0]).total_seconds() / (60 * 60 * 24 * 365)
    df['year_diff'] = df['years'].diff().fillna(0)
    # Initialize the groups list
    groups = []
    group = []

    # Initialize the counter for the number of fitted lines
    num_lines = 0

    # Initialize the first failure rate
    first_failure_rate = None

      # Initialize the first failure rate
    first_failure_rate = None

    # Get the index of the dataframe
    index = df.index

    # Get the number of rows in the dataframe
    n = df.shape[0]

     # Add a counter for the number of groups
    num_groups = 0

    # Add a counter to track the number of valid segments plotted for each pump
    valid_segment_counter = 0

    # Loop through the rows of the dataframe
    for i in range(n):
        if i == 0:
            group.append(df.iloc[i, :])
            if df.iloc[i, :]['year_diff'] == 0:
                first_failure_rate = 0
            else:
                first_failure_rate = (df.iloc[i, :]['cumulative_failure_count']) / df.iloc[i, :]['year_diff']
        else:
            date_diff = df.iloc[i, :]['year_diff']
            if date_diff > (internal_threshold / 3) or date_diff > critical_value:
                if len(group) >= 3:
                    groups.append(group)
                group = [df.iloc[i, :]]
                first_failure_rate = df.iloc[i, :]['cumulative_failure_count'] / df.iloc[i, :]['year_diff']
            else:
                if df.iloc[i, :]['year_diff'] != 0:
                    failure_rate = df.iloc[i, :]['cumulative_failure_count'] / (df.iloc[i, :]['year_diff'])
                else:
                    failure_rate = 0

                if first_failure_rate != 0:
                    failure_rate_percent_diff = abs((failure_rate / (first_failure_rate)) - 1)
                else:
                    failure_rate_percent_diff = 0

                if failure_rate_percent_diff < failure_rate_threshold:

                    group.append(df.iloc[i, :])
                #elif failure_rate_percent_diff > critical_value:
                    #group.append(df.iloc[i, :])
```

```
                else:
                    if len(group) > 3:
                        groups.append(group)
                    group = [df.iloc[i, :]]
                    first_failure_rate = failure_rate

    # Add the last group
    groups.append(group)
    #print(len(groups))
    # Initialize the result dictionary
    #result = {'increasing': [], 'constant': [], 'decreasing': [], 'sudden': []}
    df.plot(kind = 'scatter', x='years',y='cumulative_failure_count', label="Failure")
    plt.xlabel('Arrival time (years)')
    plt.ylabel('Cum.number of failures (-)')

    # Loop through the groups
    for i, group in enumerate(groups):
        if len(group) > 3:
            df_group = pd.DataFrame(group)
            df_group['order'] = f'Station {df_group.iloc[0, 0]} pump{df_group.iloc[0, 1]} segment {i + 1}'

            #display(df_group)
            #print(group)
            x = df_group['years'].values.reshape(-1, 1)
            y = df_group['cumulative_failure_count'].values.reshape(-1, 1)
            # Perform the combined test
            result = categorize_units(x)

            # Categorize the results
            if result == 'NHPP':
                NHPP_segments.append(df_group)
            elif result == 'RP':
                RP_segments.append(df_group)
            else:  # 'HPP'
                HPP_segments.append(df_group)
            # Fit a linear regression model
            model = LinearRegression().fit(x, y)
            slope = model.coef_[0][0]
            intercept = model.intercept_[0]
            y_fit = model.predict(x)
            #print(x)D:\\x\\data1\\
            #print(y_fit)
            plt.plot(x, y_fit, label='Fitted Line', linewidth=3)
            plt.title(f"Station{df.iloc[0, 0]}pump{df.iloc[0, 1]}")
            #filename = str(f'Station{df.iloc[0, 0]}pump{df.iloc[0, 1]}')
            #plt.savefig('D:/x/goat/{}.png'.format(filename))
            valid_segment_counter += 1



    df['num_lines'] = num_lines
    if valid_segment_counter == 0:
        no_fitted_line_pumps.append(df)
```

Listing A.3: Modified Cox Process

```

"""
3 / 3

This Python function simulates a Cox process, a type of stochastic process often used to model random event occurrences over time, such as equipment failures. The function uses Gaussian process regression for the intensity function, then applies the thinning method for the simulation, evaluates the fit through the Kolmogorov-Smirnov test and mean squared error, and finally visualizes the original and simulated cumulative failures, providing an intuitive tool for understanding and forecasting system behavior.
"""

#Modified Cox Process logic
def cox_process_simulation(cox_time_data, n_simulations=1000):
```

```
    if len(cox_time_data) == 0:
        return 'empty'
    your_time_data = cox_time_data['years'].values
    your_time_data = np.array([your_time_data]).reshape(-1, 1)

    # Calculate the total number of events
    total_events = len(your_time_data)

    # Calculate the total time interval
    total_time = your_time_data[-1] - your_time_data[0]

    # Calculate lambda
    lambda_value = total_events / total_time

    # Generate data for fitting Gaussian Process
    n_samples = len(your_time_data)
    X = your_time_data
    Y = np.random.poisson(lambda_value, size=n_samples)

    # Reshape Y to have shape (n_samples, 1)
    Y = Y.reshape(-1, 1)

    # Fit a Gaussian process to the data
    kernel = GPy.kern.RBF(input_dim=1, variance=1.0, lengthscale=0.1)
    model = GPy.models.GPRegression(X, Y, kernel)
    model.optimize_restarts(num_restarts=10, verbose=False)

    # Compute the cumulative number of failures for the original data
    cumulative_failure = np.arange(1, len(your_time_data) + 1)

    # Initialize variables for storing the best simulation
    best_p_value = -np.inf
    best_mse = np.inf
    best_event_times = None
    best_cumulative_failures = None

    # Simulate the Cox process multiple times
    for _ in range(n_simulations):
        # Predict the intensity function
        X_new = np.linspace(X.min(), X.max(), 100).reshape(-1, 1)
        mean, variance = model.predict(X_new)

        # Assess the intensity function
        intensity_slope = np.diff(mean.flatten()) / np.diff(X_new.flatten())
        average_slope = np.mean(intensity_slope)
        if average_slope > 0:
            #print("Interarrival times are decreasing (intensity function is increasing)")
            increasing_segments_simulation.append(cox_time_data)
        elif average_slope < 0:
            #print("Interarrival times are increasing (intensity function is decreasing)")
            decreasing_segments_simulation.append(cox_time_data)
        else:
            #print("Intensity function is approximately constant")
            constant_segments_simulation.append(cox_time_data)

        # Thinning method for simulating Cox process
        def simulate_cox_process(mean_intensity, X_new, T):
            max_intensity = np.max(mean_intensity)
            t = 0
            event_times = []
            while t < T:
                s = np.random.exponential(1 / max_intensity)
                t += s
                u = np.random.uniform(0, max_intensity)
                intensity_t = np.interp(t, X_new.flatten(), mean_intensity.flatten())
                if u <= intensity_t:
                    event_times.append(t)
            return np.array(event_times)

        event_times = simulate_cox_process(mean, X_new, float(total_time)) + your_time_data
            [0]
```

```python
        cumulative_failures = np.arange(1, len(event_times) + 1)

        def ecdf(data):
            x = np.sort(data)
            n = x.size
            y = np.arange(1, n + 1) / n
            return x, y
        # Perform the K-S test
        original_x, original_y = ecdf(your_time_data.flatten())
        simulated_x, simulated_y = ecdf(event_times)

        if len(original_y) == 0 or len(simulated_y) == 0:
            print("Either original or simulated data is empty")
            return 'not_fit', None, None
        else:
            statistic, p_value = ks_2samp(original_y, simulated_y)


        # Compute the MSE for the common length of the series
        common_length = min(len(cumulative_failure), len(cumulative_failures))
        mse = mean_squared_error(cumulative_failure[:common_length], cumulative_failures[:
            common_length])


        # Update the best simulation if necessary
        if p_value > best_p_value or (p_value == best_p_value and mse < best_mse):
            best_p_value = p_value
            best_mse = mse
            best_event_times = event_times
            best_cumulative_failures = cumulative_failures

    # Check the goodness-of-fit
    alpha = 0.05
    if best_p_value > alpha:
        print(f"{cox_time_data.iloc[0, -1]}: The Cox process fits the original data well (
            fail to reject H0)")
        return 'fit', best_event_times, best_cumulative_failures
    else:
        print(f"{cox_time_data.iloc[0, -1]}: The Cox process does not fit the original data
            well (reject H0)")
        return 'not_fit', best_event_times, best_cumulative_failures

empty_segments_simulation = []
not_fit_segments_simulation = []
simulated_segments = []

for i in range(len(NHPP_segments)):
    segment = NHPP_segments[i]
    result, best_event_times, best_cumulative_failures = cox_process_simulation(segment)
    if result == 'empty':
        empty_segments_simulation.append(i)
    elif result == 'not_fit':
        print(i)
        not_fit_segments_simulation.append(i)
    else:
        simulated_segments.append((segment, best_event_times, best_cumulative_failures))


for segment, event_times, cumulative_failures in simulated_segments:
    your_time_data = segment['years'].values
    your_time_data = np.array([your_time_data]).reshape(-1, 1)
    cumulative_failure = np.arange(1, len(your_time_data) + 1)
    lower_bound = 0.95 * cumulative_failures
    upper_bound = 1.05 * cumulative_failures
    plt.fill_between(event_times, lower_bound, upper_bound, color='gray', alpha=0.5, label="
        95% confidence interval")
    plt.plot(your_time_data, cumulative_failure, 'o', label="Original cumulative failures")
    plt.plot(event_times, cumulative_failures, '-', label="Simulated cumulative failures")
    plt.xlabel('Arrival time (years)')
    plt.ylabel('Cum.number of failures (-)')
    plt.title(f"{segment.iloc[0, -1]}")
```

```
    plt.legend()
    plt.show()
```

B

Summary of Pump Station Data

Table B.1: Summary for Station 1619 Pump 1

Parameter	Value
Pumpstation ID	1619
PUMP ID	1
No. of Failures (5 years)	133
Total Failure Duration	3751
No. of Segments	4
Trends	Decreasing, Increasing, Increasing, Increasing
Models	Cox Process, Cox Process, Cox Process, Cox Process

Table B.2: Summary for Station 560 Pump 2

Parameter	Value
Pumpstation ID	560
PUMP ID	2
No. of Failures (5 years)	21
Total Failure Duration	773
No. of Segments	4
Trends	Increasing, -, -, Increasing
Models	Log-linear, RP Modelling (Weibull), RP Modelling (Gamma), Crow's Model

Table B.3: Summary for Station 566 Pump 2

Parameter	Value
Pumpstation ID	566
PUMP ID	2
No. of Failures (5 years)	69
Total Failure Duration	2069
No. of Segments	3
Trends	Decreasing, Increasing, Increasing
Models	Log-linear, Cox Process, Cox Process

Table B.4: Summary for Station 512 Pump 1

Parameter	**Value**
Pumpstation ID	512
PUMP ID	1
No. of Failures (5 years)	123
Total Failure Duration	3993
No. of Segments	4
Trends	Increasing, Increasing, Increasing, Increasing
Models	Log-linear, Cox Process, Log-linear, Cox Process

Table B.5: Summary for Station 560 Pump 1

Parameter	**Value**
Pumpstation ID	560
PUMP ID	1
No. of Failures (5 years)	34
Total Failure Duration	1375
No. of Segments	3
Trends	Decreasing, Increasing, Increasing
Models	Cox Process, Cox Process, Log-linear