Joint Research Programme
BTO 2024.060 | September 2024

De impact van interne factoren op storingen

Joint Research Programme

KWR

Bridging Science to Practice

Colophon

De impact van interne factoren op storingen

BTO 2024.060 | September 2024

This research is part of the Joint Research Programme of KWR, the water utilities and Vewin.

Project number
402045/359

Project manager
I (Ina) Vertommen MSc.

Client
BTO - Bedrijfsonderzoek

Author(s)
Dr. M (Mohamad) Zeidan en B (Bram) Hillebrand MSc

Quality Assurance
Dr. ir. E.J.M (Mirjam) Blokker

Sent to
This report is distributed to BTO-participants and is public one year after publication.

Keywords
Impact, interne factoren

Year of publishing
2024

More information
T +31(0)30 6069738
E mohamad.zeidan@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

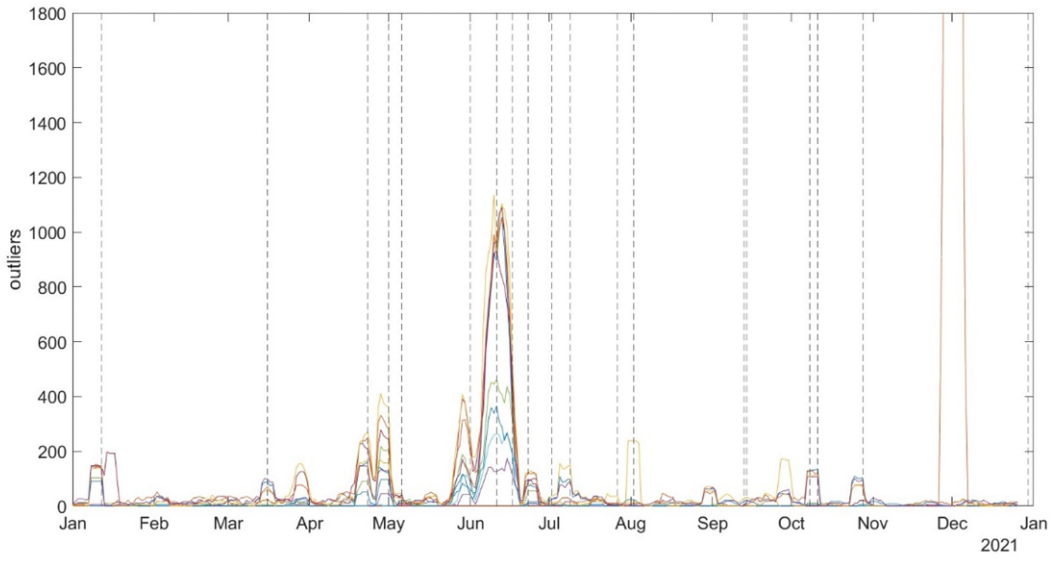
T +31 (0)30 60 69 511
E info@kwrwater.nl
I www.kwrwater.nl

Management samenvatting

Drukschommelingen en drukstoten verminderen zal het aantal storingen mogelijk reduceren

Authors Mohamad Zeidan, Bram Hillebrand.

Voor waterbedrijven is het interessant om te proberen het optreden van drukschommelingen en drukstoten te verminderen. Dit levert mogelijk een reductie op in het aantal storingen. Er zijn namelijk aanwijzingen voor een relatie tussen drukstoten en storingen (casus Evides II) en tussen zogenoemde abnormale drukken en storingen (Casus WML). Dit is het resultaat van onderzoek van vier casussen (twee bij Evides, elk een bij De Watergroep en WML) om te achterhalen welke factoren invloed hebben op storingen en ook beïnvloedbaar zijn door de waterbedrijven. De relatie tussen gebieden met veel drukschommelingen en storingen verdient aandacht in verder onderzoek (Casus de Watergroep).



Figuur 1. De stippellijnen laten de momenten van (individuele) storingen zien, de grafieken de drukanomalieën per sensor (casus WML). Een statistische analyse heeft aangetoond dat de relatie tussen pieken van drukanomalieën en storingen statistisch significant is.

Belang: Grip op interne factoren die storingen veroorzaken

Storingen aan leidingen kunnen zowel externe als intern oorzaken hebben. Een externe oorzaak is bijvoorbeeld directe graafschade, de meeste andere lekken worden gecategoriseerd als interne oorzaak. Een leiding faalt door te veel belasting. Het kan gaan om permanente overbelasting (door grondbelasting of waterdruk) of dynamisch overbelasting (door een drukstoot of verkeersbelasting) zijn. Het is belangrijk te weten welke van deze belastingen een interne factor hebben, omdat dat betekent dat het waterbedrijf er direct of via klanten/aannemers invloed op kan uitoefenen. Daarom is onderzocht in hoeverre storingen en de bijbehorende reparaties andere storingen veroorzaken en in hoeverre verschillende drukeigenschappen storingen veroorzaken.

Aanpak: Statistische analyse van relaties in vier casussen

Verschillende statistische methodes zijn ingezet om relaties te testen tussen eerdere storingen en nieuwe storingen, druk (maximaal, minimaal en gemiddeld) en storingen, een drukgolf (bijvoorbeeld als gevolg van pompbedieningen) en storingen en een druk “vingerafdruk” en storingen. Dit is gedaan voor vier verschillende casussen: twee bij Evides en een elk bij WML en De Watergroep.

Resultaten: Relatie tussen druk en storingen is aantoonbaar, tussen storingen en storingen niet

De verschillende casussen laten zien dat storingen niet vaker dicht bij elkaar plaatsvinden. Het is dus niet aan te tonen dat storingen andere storingen veroorzaken. Uit een casus van Evides (Evides II) valt wel af te leiden dat er kort na een drukgolf zo veel storingen plaatsvinden dat toeval onwaarschijnlijk is. Het is dus statistisch aannemelijk dat deze storingen zijn veroorzaakt door de drukgolf. Ook de casus van WML laat zien dat er een statistische relatie is tussen druk en storingen (in tijd). De casus van De Watergroep laat zien dat dezelfde relatie in ruimte niet evident is, maar wel meer onderzoek verdient waarvoor vooral meer data nodig zijn.

Toepassing: Maatregelen nemen om drukschommelingen te verminderen

De resultaten laten zien dat het voor waterbedrijven interessant is om te proberen het optreden van drukschommelingen en drukstoten te verminderen. Dit levert zeer waarschijnlijk een reductie op in het aantal storingen. De relatie tussen gebieden met veel drukschommelingen en storingen verdient aandacht in verder onderzoek (Casus de Watergroep).

Rapport

Dit onderzoek is beschreven in het rapport *De impact van interne factoren op storingen* (BTO 2024.060).

More information
T +31(0)30 6069738
E mohamad.zeidan@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

.

More information
T +31(0)30 6069738
E mohamad.zeidan@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

Contents

1 Introduction

Water distribution networks form the backbone of modern urban water infrastructure, facilitating the vital supply of clean drinking water to communities. The reliable and uninterrupted functioning of these networks is paramount to ensuring public health and daily life. However, these systems are not immune to challenges, with burst events posing a significant concern. Burst incidents lead to water losses, service disruptions, and costly repairs, making them a critical focus for water utilities and network operators. Most Dutch drinking water companies use a uniform way of registering bursts called USTORE (Beuken & Moerman, 2022). Burst incidents can have several causes and are often the result of several factors of which one ends up being fatal and causing the bursts. Among potential factors that were studied before are climate (Wols, van Summeren, Mesman, & Raterman, 2016) and traffic (Moerman & Wols, 2015). Recently in the BTO an effort was made to try to quantify/model effects of water hammer or pressure differentiations on the condition of pipes (Dash, van Laarhoven, & Wols, 2023).

In this report, we delve into a comprehensive analysis of burst events in distinct water distribution networks in the Netherlands and Flemish region. Burst events are a multifaceted phenomenon influenced by an array of factors, including network's composition (i.e., age and material), environmental conditions, and operational practices. By scrutinizing data spanning a five-year period, we aim to identify patterns, dependencies, and potential correlations that can offer insights into burst behaviors and pave the way for improved network management.

This analysis encompasses various aspects, including pipe material compositions, network age distribution, and an extensive focus on pressure analysis. Additionally, we utilize a range of statistical methods, to gain deeper insights into the spatial and temporal aspects of burst events. We meticulously investigate abnormal pressures and their relationship with burst events in various networks. The pressure analysis sheds light on the intricate dynamics between abnormal pressure events and burst occurrences.

The aim of the research is to provide water utilities, network operators, and policymakers with valuable data-driven information to enhance the understanding of burst events in water distribution networks, leading to more effective maintenance and improved planning for the future.

1.1 Reading Guide

In chapter 2, the different statistical tests and other methods employed to assess the validity of our hypotheses are detailed. Chapter 3 presents the results for each case study: two case studies from Evides (Evides I and Evides II), one from WML, and one from the Watergroep. The hypothesis regarding a potential relationship between bursts and various pressure attributes and bursts was tested for the Evides I case study. Additionally, for the Evides II case study, the same two hypotheses were examined, alongside an investigation into whether a single transient event triggered bursts. In the WML case study, the temporal correlation between pressure and bursts was explored using more intricate pressure data. Lastly, the Watergroep case study examined the spatial correlation between pressure and bursts. Chapter 4 presents conclusive remarks and recommendations.

2 Methodology

2.1 Data Collection and Descriptive Statistics

Comprehensive datasets were collected for four case studies from Evides (2), WML (1) and de Watergroep(1). The datasets included information on the entire networks' pipe materials, installation dates, geographical locations, and operational pressures at the pumping stations, as well as pressure data obtained from sensors distributed throughout the networks. The bursts within the datasets were categorized into two primary groups: spontaneous bursts and external bursts. Subsequently, within the spontaneous burst category, a further subdivision was made between asbestos cement (AC) pipes, polyvinyl chloride (PVC) pipes and for the case study for de Watergroep also steel (ST) pipes. This categorization allowed for a more detailed analysis of the distribution system's condition and structural integrity. For three of the case studies, the percentage of AC pipes and their respective ages were calculated to gain a comprehensive understanding of the system's composition and the potential influence of pipe age on bursts; while the percentage and age of steel pipes were calculated for the Watergroep.

2.2 Return Period Analysis

Return Period Analysis is a fundamental technique used to understand the recurrence and probability of specific events in a dataset. A return period analysis was conducted to assess the recurrence and probability of specific burst events in the dataset. This technique aids in understanding the long-term frequency of bursts. The analysis is typically conducted using the following steps:

Sorting the Dataset: The first step involves sorting the dataset in either ascending or descending order, depending on the variable under analysis. For instance, if you are studying the maximum daily temperatures, sort the dataset in descending order based on the temperature values.

Determining the Rank: Next, you need to determine the rank (M) of the particular value you are interested in within the sorted dataset. This rank corresponds to the position of that value within the dataset.

Calculating the Return Period: The return period (T) is calculated using the following equation:

$$P(x \geq x_m) = \frac{m}{N+1} \tag{1}$$

$$T = \frac{1}{P(x \geq x_m)} = \frac{N+1}{m} \tag{2}$$

Here, *T* represents the return period, which signifies the average time period between events of similar or greater magnitude. *N* represents the total number of data points or observations in the dataset, and *m* is the rank of the specific value of interest.

The derived return period value (T) offers insights into the average time span between events of similar or greater magnitude. For example, if T is calculated as 5 months, it implies that, on average, an event of the same magnitude or greater can be anticipated once every 5 months.

Return Period Analysis is a valuable tool for assessing the long-term frequency and probability of events, aiding in the development of strategies to manage and mitigate risks associated with extreme occurrences. It allows decision-makers to plan for and respond to events based on their expected recurrence and probability.

2.3 Autocorrelation Function

The Autocorrelation Function (*ACF*) is a statistical technique used to measure the linear relationship between a time series and its lagged values. In the context of burst data analysis, the ACF is applied to investigate the presence of temporal patterns or dependencies within the burst occurrences.

The ACF is calculated using the following equation:

$$ACF(k) = \frac{[\Sigma\ (X_t - \bar{X})(X_{t+k} - \bar{X})]}{[\Sigma\ (X_t - \bar{X})^2]} \quad (4)$$

Where $ACF(k)$ is the autocorrelation at lag k. X_t represents the data at time t. $\bar{X}$ is the mean (average) of the entire dataset. k is the lag or time interval for which the correlation is calculated.

This way, the correlation between the values at time t and the values at time t+k is calculated. The ACF function is computed for various lags (k) to examine the correlations at different time intervals. A set of autocorrelation coefficients are generated.

2.4 Pearson Correlation Analysis

Pearson Correlation Analysis is a statistical technique used to measure the linear association between two continuous variables. In burst data analysis, Pearson correlation is employed to investigate the relationship between burst occurrences and other variables, such as pressure, temperature, or system parameters.

The Pearson correlation coefficient (r), often denoted as Pearson's r, is calculated using the following equation:

$$r\ = \frac{\sum(X\ -\ \bar{X})(Y\ -\ \bar{Y})}{\sqrt{\sum(X\ -\ \bar{X})^2 * \sum(Y\ -\ \bar{Y})^2}} \quad (5)$$

Where X and Y are the variables for which you want to calculate the correlation. $\bar{X}$ (X-bar) and $\bar{Y}$ (Y-bar) represent the means of the X and Y variables, respectively. Σ denotes the sum of the values, so $\Sigma(X - \bar{X})^2$ represents the sum of the squared differences between each X value and the mean of X, and $\Sigma(Y - \bar{Y})^2$ represents the sum of the squared differences between each Y value and the mean of Y.

The Pearson correlation coefficient (r) is a measure of the linear relationship between two variables. It quantifies the strength and direction of the linear association. The value of r can range from -1 to 1, where: r = 1 indicates a perfect positive linear relationship. r = -1 indicates a perfect negative linear relationship. r = 0 indicates no linear relationship between the two variables.

2.5 Pressure Analysis

Pressure abnormalities, or outliers, are defined as instances deviating by three standard deviations from the yearly average. The annual calculation of these outliers is conducted independently, ensuring avoidance of normalization issues arising from diverse operational regimes and protocols. The identification of outliers serves as a critical indicator of unstable hydraulic conditions that have the potential to compromise the integrity of the water distribution system. While some outliers may be associated with pressure transients, it is imperative to note that the

available time step resolution may be insufficient for capturing such transients comprehensively. Nonetheless, the identified outliers effectively highlight hydraulic instability and potential issues within the system.

Burst Event Probability Test

This test assesses the correlation between burst events and pressure anomalies using three key metrics. These metrics are designed to evaluate the likelihood of detecting burst events compared to non-burst periods and random occurrences throughout the year. By mapping the cumulative outliers, we seek to identify patterns that may contribute to a proactive understanding of system behavior, ultimately enhancing the predictive capabilities for identifying vulnerable conditions within the water distribution network.

The first test evaluates the detection of burst events relative to the total number of recorded events. This metric provides insight into the overall effectiveness of burst event detection. In the second test, burst events are removed along with the five days preceding each event to establish a non-burst sampling group. The objective is to assess the number of events detected during non-burst periods and quantify the percentage of falsely detected events. These falsely detected events are indicative of false positives. The third test focuses on determining the probability of detecting a burst event on a random day throughout the year. This serves as a baseline comparison for assessing the significance of burst event detection in the first test. The following criteria are then examined:

- Real Positives (%): Represents the percentage of burst events detected relative to the total number of recorded events. Higher percentages indicate a greater correlation between burst events and pressure anomalies.
- False Positives (%): Indicates the percentage of falsely detected events during non-burst periods. Lower percentages suggest a more accurate burst event detection system with fewer false positives.
- Random (%): Reflects the probability of detecting a burst event on a random day throughout the year. Comparing this percentage with the real positives provides insights into the significance of burst event detection beyond random chance.

Root Mean Square Error Analysis

To gauge the deviation of the current pressure situation from the yearly average, we employed the Root Mean Square Error (RMSE) with respect to the Yearly Pressure Average. This metric provides an indication of how "unusual" the current pressure conditions are in relation to the established yearly average. A higher RMSE signifies a greater deviation from the norm. Our analysis focused specifically on the pressure readings corresponding to the dates of bursts occurring within a five-day window. By calculating RMSE in this context, we aimed to identify instances where the pressure conditions significantly diverged from the expected yearly average in the days leading up to a burst event.

2.6 Area division

For the Watergroep case study we divided the study area into separate regions around individual pressure sensors. The intention being to identify areas which have a different pressure “fingerprint” (see next section). For this study we started with the simple division in Voronoi polygons. This means that each point within that polygon (and therefore each burst) is closest in straight distance to the pressure sensor of that polygon. This was done for all pressure sensors of which the location was known. It is important to underline that this closest distance was determined as the crow flies.

2.7 Definition of the pressure “fingerprint”

The polygons mentioned above all contain one pressure sensor which yielded detailed information. With this pressure information we tried to identify “busy” and less “busy” areas (pressure “fingerprints”). Meaning we examined the amount in which pressure changed within an area in terms of amplitude and number of changes. For this we analysed the delta P using pressure information per second. An example is give in Figure 2.

In an effort to capture such a figure into a single number (so that we can perform a correlation test) we used the 75% quantile. Meaning that 75% of all delta P’s fall below that value and 25% above. A 75 percentile provides insight into whether an area often experiences significant pressure fluctuations; a higher percentile (e.g. 95)would provide insight into outlier events, which are not what we are interested in for now.

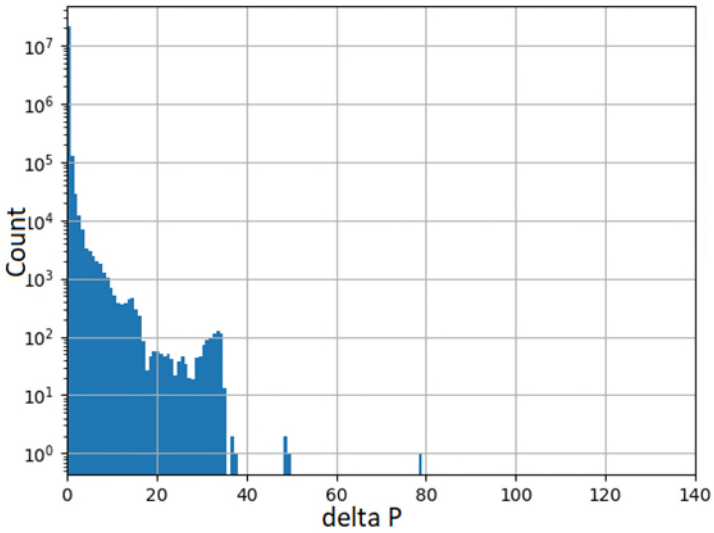


Figure 2. Example of a pressure "fingerprint".

3 Results

3.1 Case Study 1: Evides I

3.1.1 Network pipe composition

The water distribution network of this area consists of pipes made from various materials that have been installed over an extended period. This diversity implies that the network contains pipes of different ages, materials, and conditions. From the classification of pipe materials, it is evident that AC pipes make up a substantial portion, representing 40%, with PVC pipes comprising the majority at 50%. The remaining 10% comprises pipes made from other various materials.

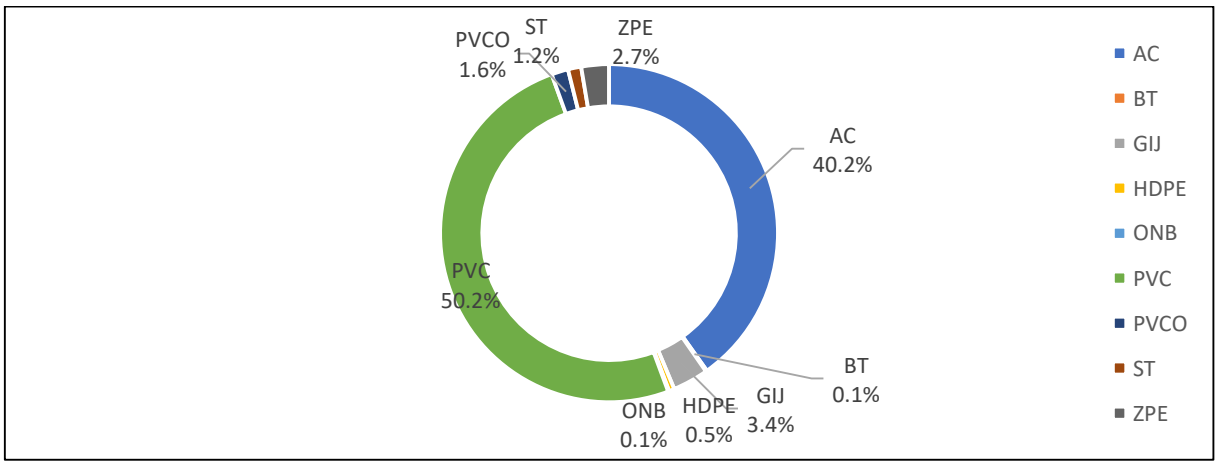


Figure 3. Pipe length material composition for Evides I case study network.

By analyzing the age of the different pipe material groups, it is evident that the age of AC pipes is notably higher compared to other pipe materials. AC pipes exhibit an age range spanning from 37 to 122 years, with an average age of 70 years. However, the age of 122 years may be considered an outlier due to its improbability. In contrast, PVC pipes have an average age of 11 years, with an age range from 1 to 18 years (Figure 4).

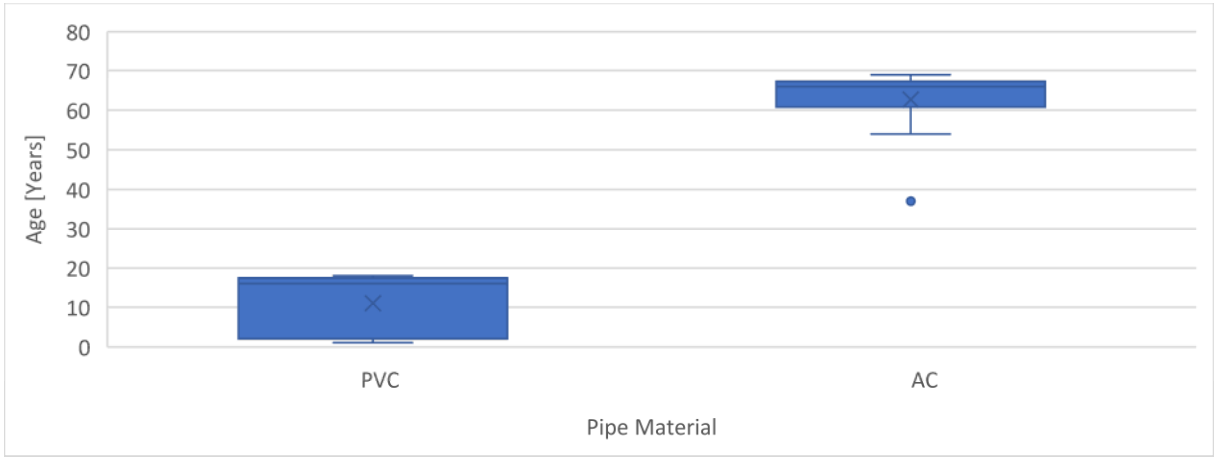


Figure 4. The age spectrum for pipe material categories in Evides I case study network

In the water distribution network for the Evides I case study area, a dataset concerning pipe bursts covering the years 2017 to 2022 was provided to understand the temporal pattern of bursts. This data provides valuable insights into the frequency and seasonality of burst events within the network. During the period spanning 2017 to 2021, a total

of 765 bursts were recorded in this network. These bursts were categorized into two main types: spontaneous bursts (408) and external bursts (357), hereafter referred to as "derden" bursts. Figure 5 provides a breakdown of the material composition responsible for spontaneous bursts.

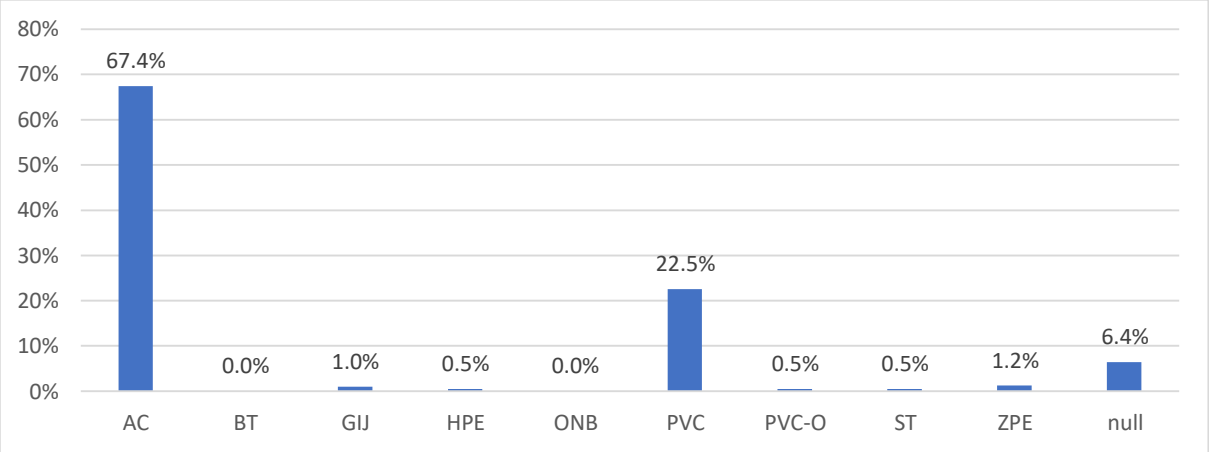


Figure 5. Percentage of pipe bursts per material, null refers to pipe bursts with no associated pipe material in the dataset.

It is evident that AC pipes are accountable for a substantial 67% of the total spontaneous bursts during this period, making them the primary contributor to burst events. In contrast, PVC pipes, the second most prevalent material, accounted for 22.5% of the spontaneous bursts. Some bursts in the dataset were not associated with a specific material and are represented here as "null". These findings emphasize the critical role of pipe material, especially the presence of AC pipes, in influencing the incidence of bursts in this network.

3.1.2 Spontaneous Bursts Temporal Distribution

A temporal distribution analysis was conducted to explore patterns, such as uniformity or recurring yearly and monthly patterns. The spontaneous bursts data are examined for the two primary pipe material groups (AC and PVC), as well as the total bursts. The data exhibited in this visual representation underscores the dynamic and evolving nature of burst occurrences within the network. Notably, there is a substantial deviation observed from one month to the next, and this irregularity suggests that no clear and predictable pattern can be confidently concluded from the available data. Figure 6 presents a comprehensive depiction of how the number of burst incidents fluctuates over the recorded time period.

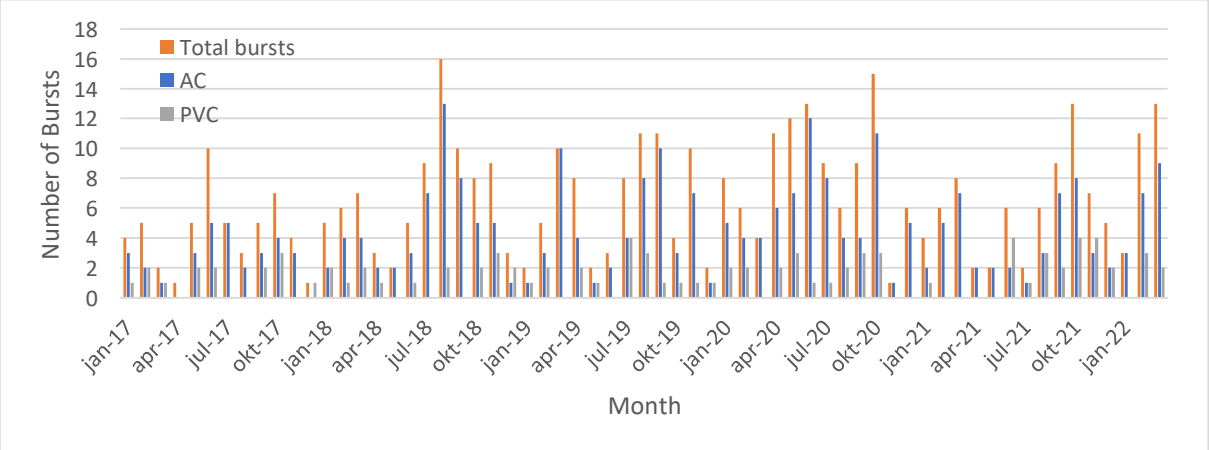


Figure 6. The temporal distribution of spontaneous pipe bursts over the recorded period

The absence of a discernible pattern could be attributed to the complex interplay of various factors that influence burst events. Nevertheless, amid this variance, a prominent trend emerges – the significant role played by AC pipes in contributing to the burst incidents. To provide a more comprehensive visual representation of burst distribution,

a surface graph has been generated. Within this graph, the highest number of bursts are distinctly represented in green, and an intriguing pattern unfolds.

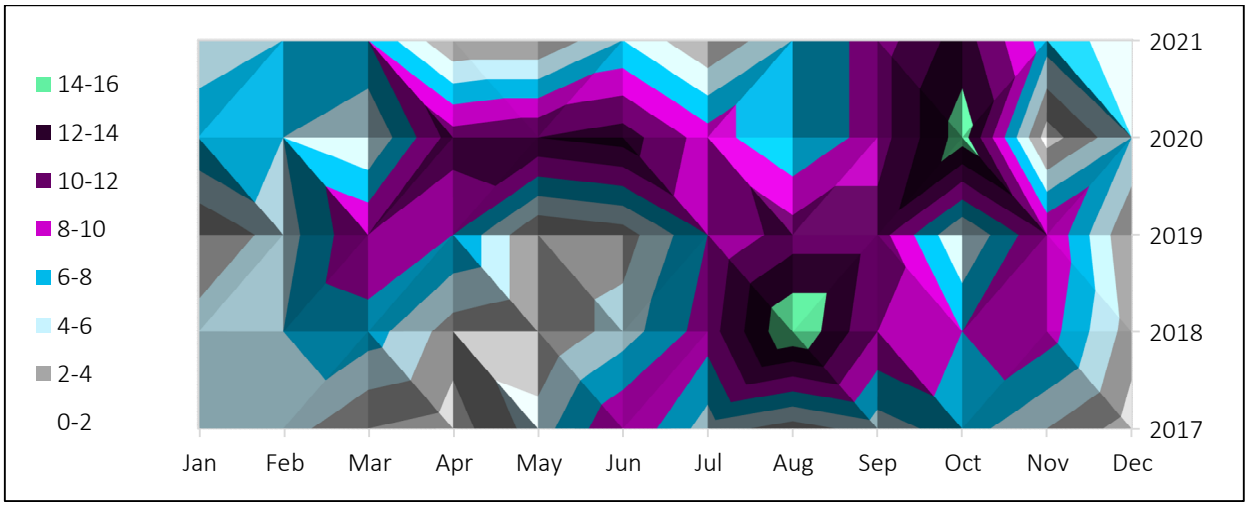


Figure 7. Surface graph for the number of spontaneous bursts per month over the recorded period.

Notably, the zenith is reached in August 2018, with an astonishing 16 bursts recorded within this single month. The surface graph serves as an invaluable tool for assessing burst distribution across the temporal spectrum. It offers a dynamic portrayal of how burst incidents have manifested over time. While individual months may exhibit sporadic fluctuations, this graphical representation allows us to discern more profound trends and shifts within the dataset. One discernible observation is that, beginning in 2019, there is a notable uptick in bursts from March to June. This suggests a potentially evolving dynamic in burst behavior during these months. Conversely, the period from July to October consistently witnesses a substantial volume of burst incidents, suggesting a recurring pattern of heightened burst activity during these months. This detailed analysis, facilitated by the surface graph, provides valuable insights into the temporal variations of burst occurrences.

The data set, spanning five years, demonstrates variations in the number of spontaneous bursts observed in different months and years. For instance, in 2020, there were 100 spontaneous bursts in total, while 2018 saw 83 spontaneous bursts, illustrating differences in annual burst patterns. Moreover, the distribution of bursts across months exhibits seasonality, with certain months experiencing a higher number of burst events compared to others.

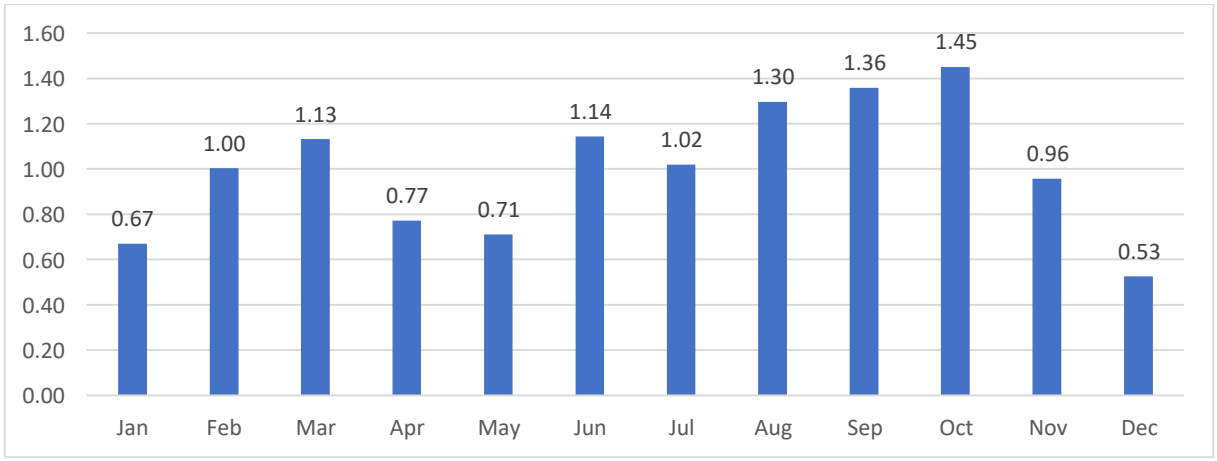


Figure 8. Monthly factor deviation from yearly average for spontaneous bursts.

Figure 8 illustrates the dynamic behavior of deviation factor from yearly average over five years. On average, spontaneous bursts frequencies vary from month to month, with the highest monthly average recorded in October,

reaching 9.4 bursts or 1.45 deviation from the yearly average. Some months clearly exhibit higher burst activity than others, which may be influenced by a combination of factors, including potential weather changes, shifts in community behavior, fluctuations in tourist activity, or even scheduled maintenance and development work within the network area. While the precise causes remain speculative due to the absence of specific data, the observed monthly variations emphasize the need for vigilant network management and planning. Effective resource allocation is crucial to mitigate potential disruptions and ensure network reliability, especially during periods of heightened burst activity.

3.1.3 Analysis of The Average Distance Between Clusters

The analysis of the average distance between clusters of burst incidents occurring within 7 days provides valuable insights into the spatial and temporal behavior of these incidents. This analysis encompasses both spontaneous (spontaan) and external (derden) bursts, offering a comprehensive understanding of the network's burst patterns. To investigate these burst patterns, the distance between each burst and all (other) bursts within a 7 days period are calculated. This approach allows for an important examination of the distribution of burst events, which is crucial for network management and maintenance decisions. This analysis serves as a foundation for understanding the spatial and temporal dynamics of burst events within the network. It helps in characterizing how bursts affect each other within a 7 days period. The analysis revealed that only 3.1% of the consecutive bursts (within 7 days) happen within less than 0.5 km. and only 5.3% happen within less than a kilometer. The histogram is presented in Figure 9.

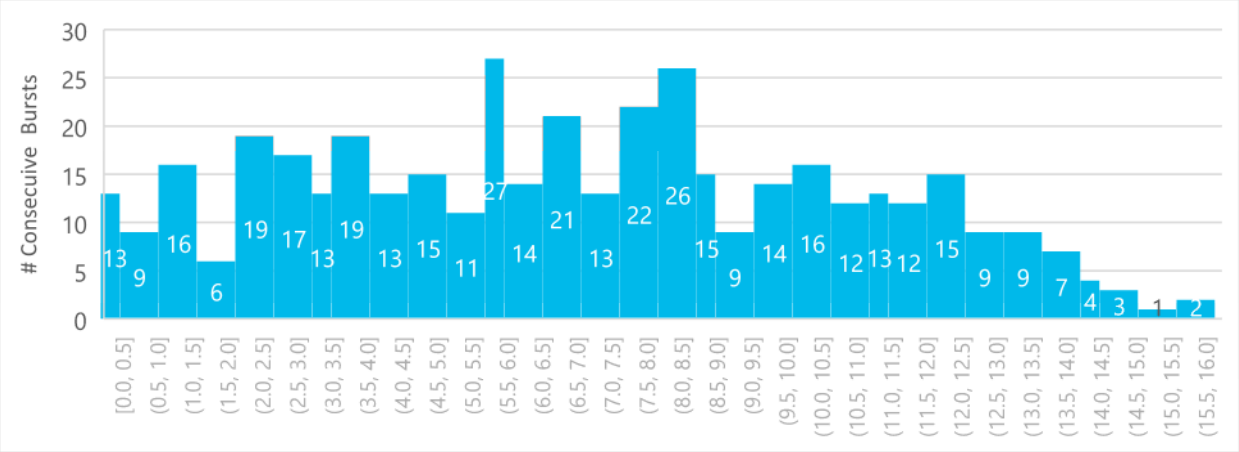


Figure 9. Histogram of the spatial distribution of burst within 7 days.

The histogram graphically illustrates the frequency of distance values, allowing for a clear visualization of the data distribution. The shape of the histogram and the concentration of values within specific distance ranges provide additional information about the spatial patterns of burst clusters. The distribution, displaying a predominantly random pattern, implies an absence of observable correlation between bursts.

3.1.4 Burst Causing Burst Analysis

The Autocorrelation Function (ACF) analysis aims to explore any temporal correlations or dependencies between burst events caused by different factors. Therefore, an ACF correlation test was performed on the burst events with a weekly resolution for 250 weeks. That is to examine the correlation between the bursts, and to check for a seasonality correlation aspect observed in the data. The results, as depicted in Figure 10 indicate that there is no substantial correlation between these events on a weekly basis. Moreover, it shows that there is no detected seasonality pattern.

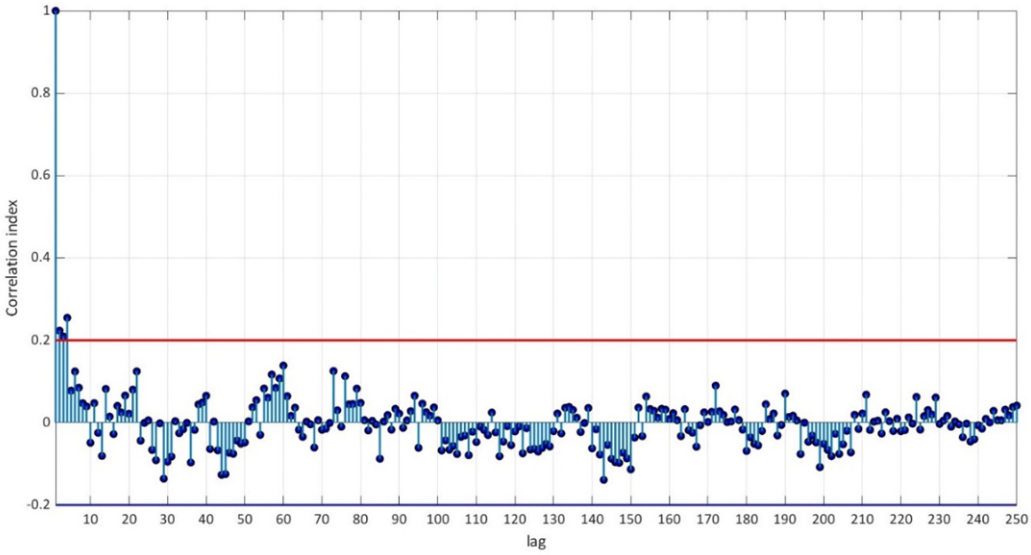


Figure 10. ACF function plot for weekly lags. (a) for 10 week lags to check for burst causing bursts. While (b) is for 250 weeks to check for seasonality.

In the ACF plot, it becomes evident that, apart from the 0 lag (bursts within 1 week), none of the autocorrelation indices exceeded 0.3. This suggests that, at a weekly time scale, burst events resulting from different causes do not exhibit strong temporal associations or patterns. The relatively low autocorrelation coefficients indicate that the occurrence of burst events in one week is not significantly influenced by the occurrence of burst events in previous weeks. This finding is essential for understanding the temporal behavior of burst events in relation to their causes, as it suggests that the network's burst occurrences are not strongly interdependent at a weekly resolution.

3.1.5 Pressure Causing Bursts

The final test conducted was a Pearson Correlation Analysis, examining the potential relationships between the number of burst events and various pressure parameters and behaviors. The pressure properties investigated for the same week included the sum of pressure over time, the mean pressure value, maximum pressure, minimum pressure, as well as the sum and maximum values of pressure fluctuations. To assess these relationships, the Pearson correlation test was employed. The results of this analysis revealed that there were no significant correlations for any of the examined pressure properties across the six available sensors. In the vast majority of cases, the correlation coefficient (r value) did not exceed 0.1. This suggests that there is minimal to no linear association between burst events and the pressure characteristics monitored by these sensors. These findings imply that the number of burst events within the network does not exhibit a strong linear relationships with the selected pressure parameters and behaviors. The limited correlation indicates that other factors or variables may play a more prominent role in influencing burst occurrences within the network.

3.1.6 Case study conclusions

In this case study we can see a clear difference in the behavior of different materials. While only 40% of the network consists out of AC it is responsible for about 67% of all spontaneous bursts. This can (partly) be due to age as the AC pipes are also considerably older than the PVC pipes. Bursts are not evenly distributed through time and we observe a higher burst count in different months for different years with the noticeable absence of the winter months. Burst clusters (bursts within 7 days of each other) are not necessarily close to each other. Also there was no clear indication that bursts are linked in time. Lastly the correlation between bursts and different pressure attributes was always low

3.2 Case Study 2: Evides II

In this case study, attention is turned towards the network of the area of Evides II case study, which is known to have been affected by a transient event on the 3rd of November 2022. The primary aim of this investigation if a transient event leads to a statistically significant higher number of bursts. However, before delving into the analysis of the transient event's impact, an initial exploration is conducted by examining the composition of the network's pipes. Factors such as pipe age and material composition are specifically investigated to gain insights into the network's normal behavior, with a focus on the prevalence of older pipes, frequently of the AC material, known to be more susceptible to burst incidents.

Adding to the context, data was received indicating the presence of 280 bursts from January 2017 to March 2022, providing a comprehensive overview of burst events over this period. Additionally, access has been granted to network data within a geopackage format, which includes details such as pipe length, material, diameter, and more. The availability of the complete dataset of bursts was expected to aid in determining whether the increase in bursts following the pressure event is statistically significant or merely a random occurrence.

Understanding the baseline characteristics of the network of the area of the Evides II case study, combined with a comprehensive dataset of burst events and specific information about the pressure event, serves as the foundation for this analysis. This holistic approach will enable meaningful conclusions to be drawn about the impact of the transient event on burst occurrences within the network and provide valuable insights for further investigation.

3.2.1 Network pipe composition

The water distribution network of the Evides II area encompasses pipes of diverse materials that have been progressively installed over an extended period. Consequently, there exists a considerable variance in the ages of these pipes, with a substantial presence of Asbestos Cement (AC) pipes. As Figure 11 illustrates, AC pipes constitute 34% of the total network's pipe lengths, while (PVC) pipes account for 55% of the network. The remaining 11% comprises pipes crafted from other materials.

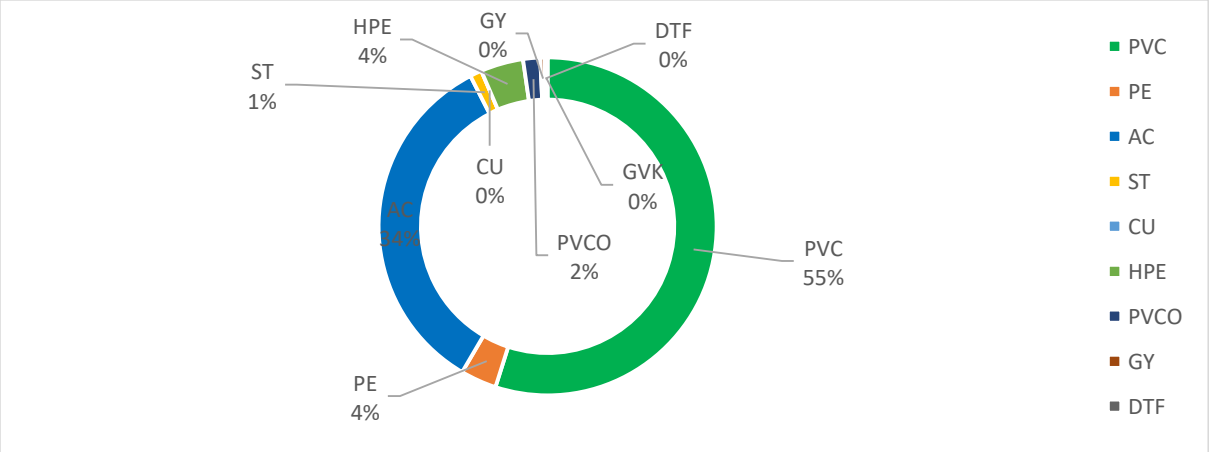


Figure 11. Evides II area pipe length percentage

A box and whisker plot representing the age distribution of pipes in the network is presented in Figure 12. AC pipes exhibit an age range spanning from 28 to 123 years, with an average age of 65 years. AC pipes with an age of 123 years point to data errors (AC was only introduced in the Netherlands in 1928 (BTO 2003.039 (Slaats & Mesman, 2004))) but this does not influence our analysis. In contrast, PVC pipes have an average age of 37 years, with an age range from 1 to 123 years (here also 123 years seems to point to unknown data point). Given the presence of aging pipe compositions, particularly the 34% share of AC pipes, the impact on the network's integrity and the incidence of failures needs to be explored. Their diameters must be examined to determine their significance in the main pipeline system or if they are historical remnants within the secondary pipes in the network.

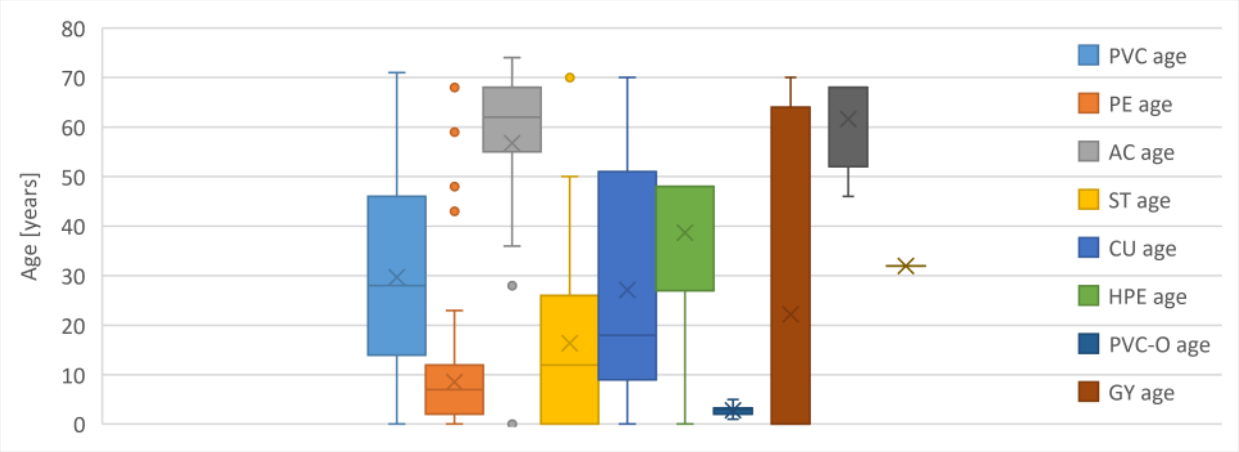


Figure 12. pipes age in 2023 per material in the network of the Evides II case study.

Figure 13 depicts the box and whisker plot illustrating the diameter distribution among various material groups. AC pipes show slightly larger diameters compared to the PVC and PE groups, although they still fall within a similar size category. The data reveals that GY (cast iron) pipes exhibit a wide range of diameters, ranging up to 300 mm, but predominantly falling between 100 and 250 mm. In contrast, ST pipes are characterized by larger diameters, typically within the 300-400 mm range. Meanwhile, DTF pipes feature smaller diameters.

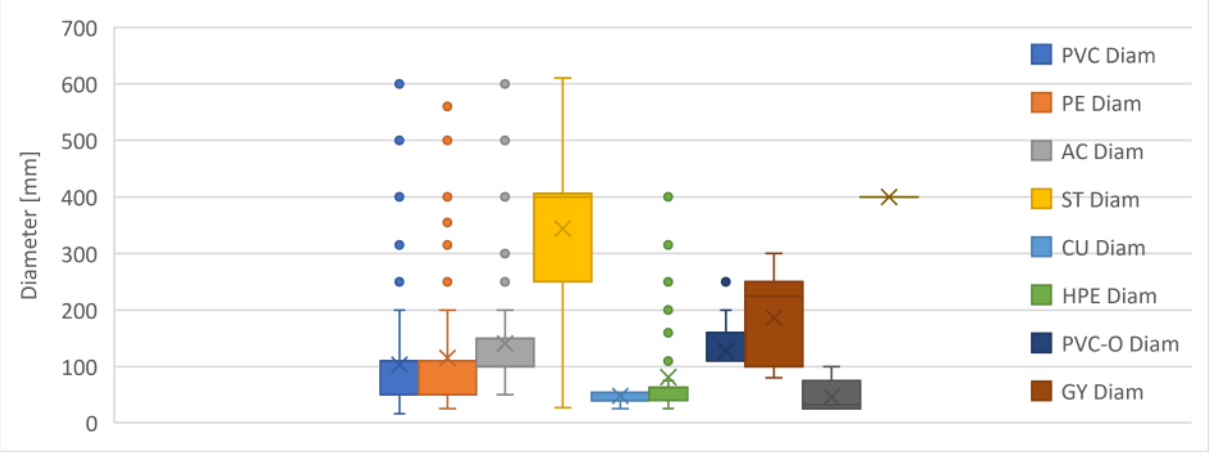


Figure 13. diameter box and whiskers for the pipes in the network of the Evides II case study.

In this water distribution network, the pipe bursts dataset covering the years 2017 to 2022 was provided. This data provides valuable insights into the frequency and seasonality of burst events within the network. Key findings from this analysis include the average bursts per month, variations in the number of bursts across different months and years, and the identification of months with heightened burst activity. In the period spanning 2017 to 2021, this network recorded a total of 280 bursts. These burst events were categorized into two primary types: spontaneous bursts and external bursts, referred to as "derden" bursts. Among these bursts, 231 were identified as spontaneous, while 49 were categorized as external.

Figure 14 presents a material-wise breakdown of the burst causes, revealing that AC pipes were responsible for 60% of the total bursts during this period, establishing them as the primary contributors to burst incidents. In contrast, PVC pipes, the second most prevalent material, accounted for 30% of the total bursts. Despite the lower overall number of bursts compared to case study 1, the data still highlights a predominant association with AC pipes in burst events.

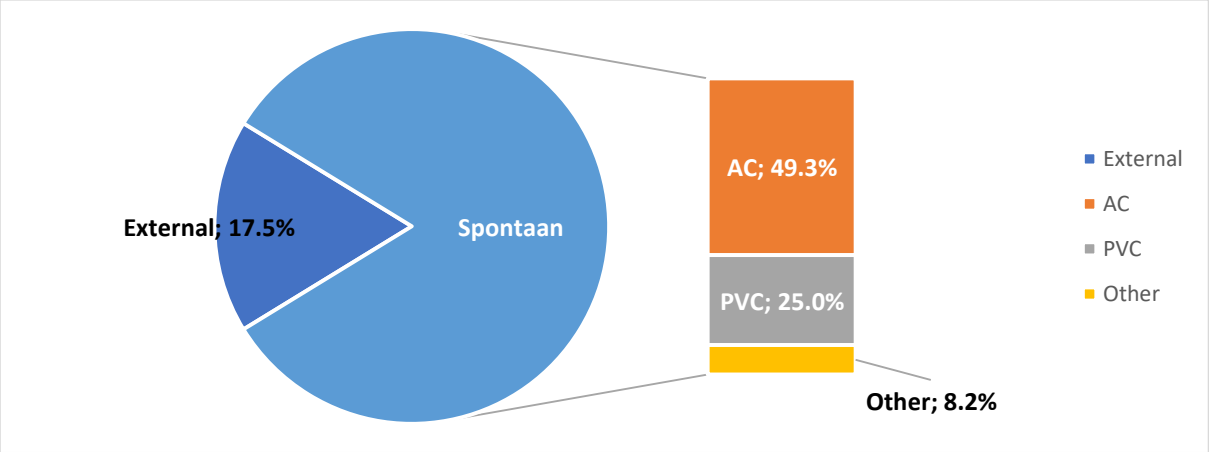


Figure 14. material percentage from bursts

These findings emphasize the critical role of pipe material composition, especially the presence of AC pipes, in influencing the incidence of bursts in this network.

3.2.2 Spontaneous Bursts Temporal Distribution

Exploring the temporal distribution of bursts is crucial to understanding their patterns and potential triggers. Figure 15 depicts the monthly distribution of bursts throughout the recorded period, illustrating the respective contributions of AC and PVC pipes for each month. Notably, the month of November 2021 is highlighted on the graph. It is evident that the AC-to-PVC burst ratio deviates significantly in this particular month. In November 2021, all the spontaneous burst incidents are associated with PVC pipes, a departure from the typical pattern where AC pipes predominantly contribute to burst numbers in other months. This observation indicates a noteworthy shift in burst behavior, emphasizing the significance of examining the factors that may have led to this deviation, such as the transient event previously mentioned.

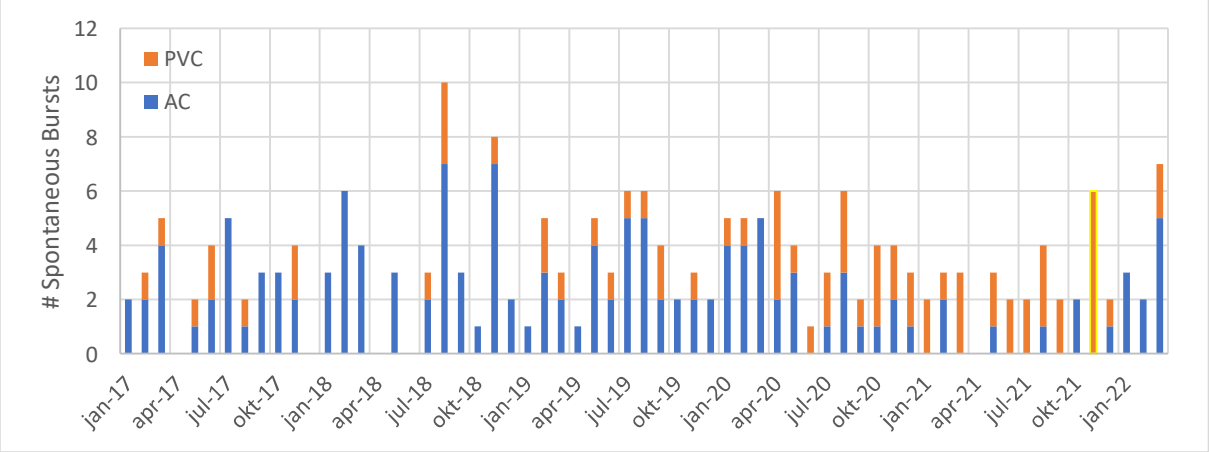


Figure 15. monthly bursts over the recorded time period

The data exhibited in Figure 15 underscores the dynamic and evolving nature of burst occurrences within the network. Notably, there is a substantial deviation observed from one month to the next, and this irregularity suggests that no clear and predictable pattern can be confidently concluded from the available data. The absence of a discernible pattern could be attributed to the complex interplay of various factors that influence burst events. Nevertheless, amid this variance, a prominent trend emerges – the significant role played by AC pipes in contributing to burst incidents. Despite the lack of a clear pattern, the data consistently underscores the importance of addressing and monitoring the integrity and performance of AC pipes within the network, given their considerable influence on the burst occurrences. For a more comprehensive visual representation of burst distribution, a surface graph has

been generated, as depicted in Figure 16. Within this graph, the highest number of bursts are distinctly represented in Purple, and an intriguing pattern unfolds.

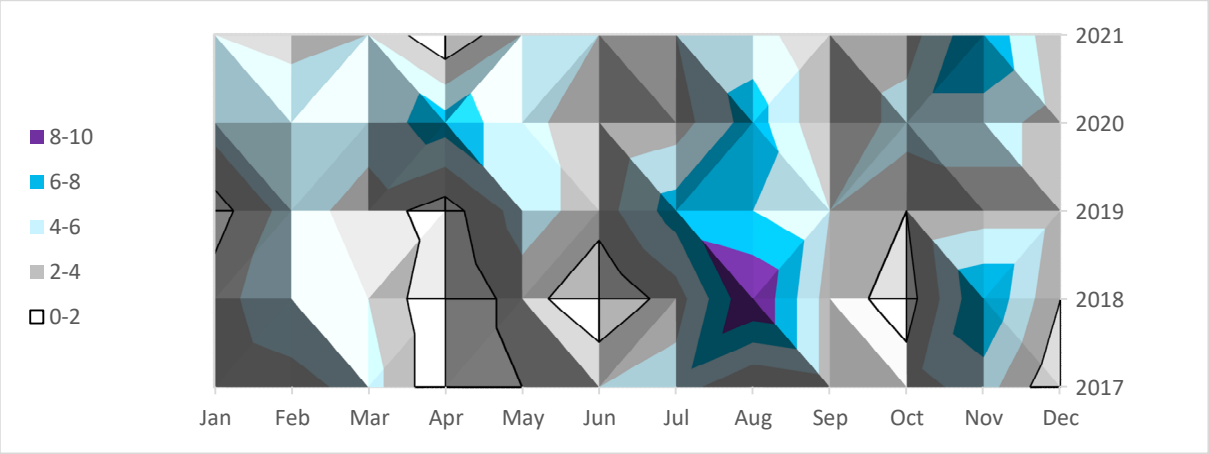


Figure 16. surface graph depicting number of bursts per month

Notably, the zenith is reached in August 2018, with 10 bursts recorded within this single month. It is important to mention that August 2018 was also an abnormal month for the first case study with 16 bursts. The surface graph serves as an invaluable tool for assessing burst distribution across the temporal spectrum. It offers a dynamic portrayal of how burst incidents have manifested over time. While individual months may exhibit sporadic fluctuations, this graphical representation allows us to discern more profound trends and shifts within the dataset. One discernible observation is that, beginning in 2019, there is a notable uptick in bursts from March to June. This suggests a potentially evolving dynamic in burst behaviour during these months. Conversely, the period from July to October consistently witnesses a substantial volume of burst incidents, suggesting a recurring pattern of heightened burst activity during these months. This detailed analysis, facilitated by the surface graph, provides valuable insights into the temporal variations of burst occurrences.

The data set, spanning five years, demonstrates variations in the number of bursts observed in different months and years. For instance, in 2020, there were 55 bursts in total, while 2018 saw 43 bursts, illustrating differences in annual leakage patterns. Moreover, the distribution of bursts across months exhibits seasonality, with certain months experiencing a higher number of burst events compared to others. In order to gain a more profound understanding of the network's seasonal behaviour, Figure 17 illustrates the average monthly baseline. This visualization provides a valuable perspective on the network's typical performance patterns throughout the year. By presenting the average monthly baseline, it becomes easier to discern any recurrent trends, variations, or anomalies in the network's behaviour.

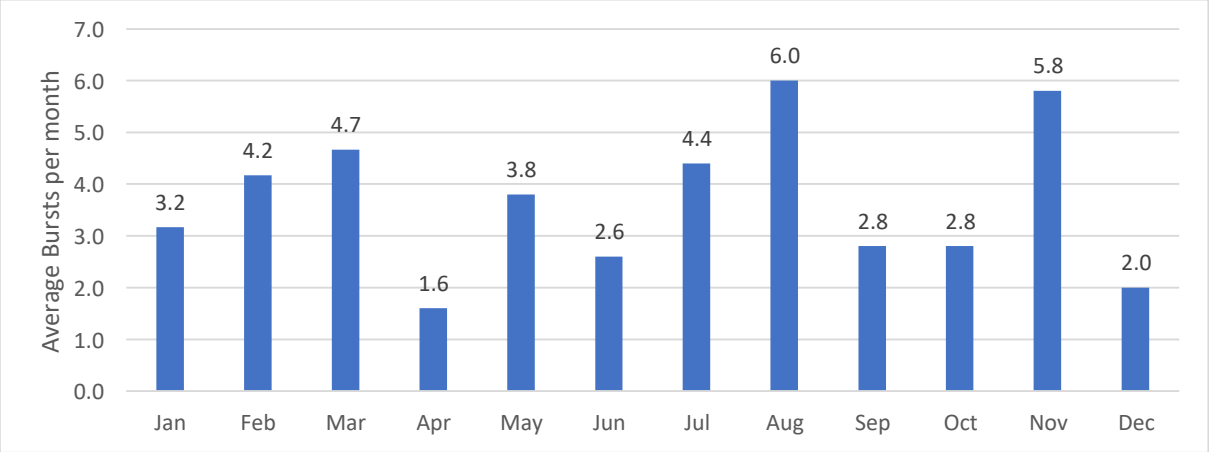


Figure 17. The dynamic distribution of dynamic bursts over 5 years period

Figure 17 illustrates the dynamic behaviour of average burst occurrences over five years. On average, burst frequencies vary from month to month, with the highest monthly average recorded in August and November, reaching around 6 bursts per month. Some months clearly exhibit higher burst activity than others, which may be influenced by a combination of factors, including potential weather changes, shifts in community behaviour, fluctuations in tourist activity, or even scheduled maintenance and development work within the network area. While the precise causes remain speculative due to the absence of specific data, the observed monthly variations emphasize the need for vigilant network management and planning. Effective resource allocation is crucial to mitigate potential disruptions and ensure network reliability, especially during periods of heightened burst activity.

3.2.3 Spontaneous Bursts Frequencies Analysis

The Poisson Distribution Analysis is performed to asses the probability of spontaneous bursts per month. The PDA is a fundamental statistical technique employed to examine the probability and recurrence of specific events within a given dataset. In the context of this study, PDA was applied to the burst data from Evides II, focusing on the monthly frequency of burst incidents. The primary goal was to estimate the likelihood of observing a particular number of burst events per month and subsequently compare this Poisson-expected distribution to the observed distribution of actual burst events. This comparative analysis sought to identify any substantial deviations between the actual distribution, which was derived directly from the dataset, and the Poisson-expected distribution.

The probability of experiencing a specific number of bursts per month can be calculated using the Poisson distribution. In this scenario, attention is directed towards the occurrence of 6 bursts per month, noted in November 2021 amid suspicions of a transient event. the results are presented at Table 1.

Table 1. Probability of having number of spontaneous bursts per month according to the Poisson distribution.

burst/month	P(X=xi), λ = 3.574	P(X<=xi)	P(X>xi)
0	2.8%	2.8%	97.2%
1	9.9%	12.7%	87.3%
2	17.8%	30.5%	69.5%
3	21.3%	51.8%	48.2%
4	19.1%	70.9%	29.1%
5	13.7%	84.6%	15.4%
6	8.2%	92.8%	7.2%
7	4.2%	97.0%	3.0%
8	1.9%	98.9%	1.1%
9	0.7%	99.6%	0.4%
10	0.3%	99.9%	0.1%

According to Table 1, the probability of encountering 6 bursts per month is 8.2%, indicating a recurrence interval of approximately 12 months (or around 1 year), which is not uncommon. However, when Table 2 is examined for the number of spontaneous bursts per month in PVC pipes, the probability decreases markedly to 0.1%, corresponding to an occurrence once every 1000 months (or approximately 83 years). Beside November 2021, no other month experienced more than 4 spontaneous PVC bursts. While April 2020 experienced 4 spontaneous PVC bursts. This discrepancy suggests that the event observed in November 2021 diverges from the usual pattern. As a result, a strong argument can be made that the pressure transient indeed influenced the incidence of pipe bursts.

Table 2 Probability of having number of spontaneous bursts per month in PVC according to the Poisson distribution.

PVC burst/month	P(X=xi) λ = 1.132	P(X<=xi)	P(X>xi)
0	32.2%	32.2%	67.8%
1	36.5%	68.7%	31.3%
2	20.7%	89.4%	10.6%
3	7.8%	97.2%	2.8%
4	2.2%	99.4%	0.6%
5	0.5%	99.9%	0.1%
6	0.1%	100.0%	0.0%
7	0.0%	100.0%	0.0%
8	0.0%	100.0%	0.0%
9	0.0%	100.0%	0.0%
10	0.0%	100.0%	0.0%

3.2.4 Burst Causing Burst Analysis

An ACF correlation test was performed on the burst events with a weekly resolution, starting with only 10 and moving on to 250 weeks. That is to examine two critical questions. (a) is there a weekly correlation between the bursts, and (b), is there a seasonality correlation aspect observed in the data. As depicted in Figure 18.(a), indicate that there is no substantial correlation between these events on a weekly basis. Moreover, Figure 18.(b) shows that there is no detected seasonality pattern.

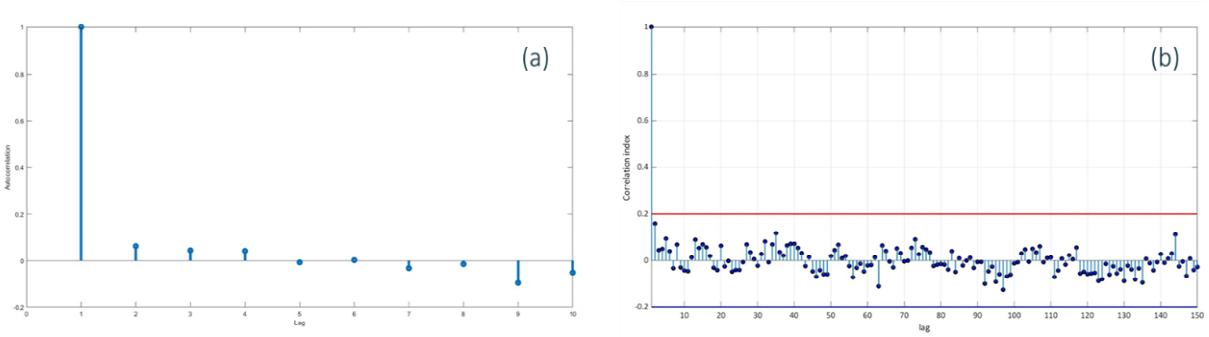


Figure 18. ACF function plot for weekly lags. (a) for 10 week lags to check for burst causing bursts. While (b) is for 250 weeks to check for seasonality.

In the ACF plot, it becomes evident that, apart from the 0 lag, none of the autocorrelation indices exceeded 0.3. This suggests that, at a weekly time scale, burst events resulting from different causes do not exhibit strong temporal associations or patterns. The relatively low autocorrelation coefficients indicate that the occurrence of burst events in one week is not significantly influenced by the occurrence of burst events in previous weeks. This finding is essential for understanding the temporal behavior of burst events in relation to their causes, as it suggests that the network's burst occurrences are not strongly interdependent at a weekly resolution.

3.2.5 Pressure Causing Burst Analysis

The results of this analysis revealed that there were no significant correlations for any of the examined pressure properties across the six available sensors. In the vast majority of cases, the correlation coefficient (r value) did not exceed 0.11. This suggests that there is minimal to no linear association between burst events and the pressure characteristics monitored by these sensors. These findings imply that the number of burst events within the network does not exhibit a strong linear relationship with the selected pressure parameters and behaviors. The limited correlation suggests that other factors or variables may play a more prominent role in influencing burst occurrences within the network.

3.2.6 Case study conclusions

In this case study we also see a clear difference in the behavior of different materials. Here the network consists for 34% out of AC while it is responsible for about 49% of the bursts. Also here this can be (partly) due to age as AC is considerably older than PVC. Here also the bursts vary with time and different years show different peaks with August and November on the high side and April and December on the low side.

Looking into the transient event that happened on the beginning of November 2021 (which saw 6 bursts, all on PVC pipes) we see that, while 6 bursts have a recurrence time of roughly once every 1 year, the recurrence time of 6 bursts on PVC pipes is approximately every 83 years. Suggesting that there is indeed a correlation between the transient event and the 6 bursts that month. Also for this case study there no correlation was found between bursts and bursts (in time) and neither between pressure attributes (excluding the transient event) and bursts.

3.3 Cases study 3: WML

3.3.1 Bursts Stats

The initial phase involves a meticulous examination of the five-year dataset on pipe bursts provided by the water utility WML. The primary focus is on pertinent factors such as pipe materials, operational age, and the causative elements behind each burst. The underlying hypothesis posits that older pipes are inherently more prone to spontaneous bursts compared to their recently installed counterparts. Furthermore, the distribution of bursts over time is scrutinized, considering the potential influence of monthly variations in human activity, alongside external parameters such as traffic, temperatures, and soil freeze. It is suspected that the system operation and hydraulic instability is what is causing the spontaneous burst, this theory will be tested in this section. First, let's consider the burst pipes stats. Looking at the burst demographic in Figure 19.

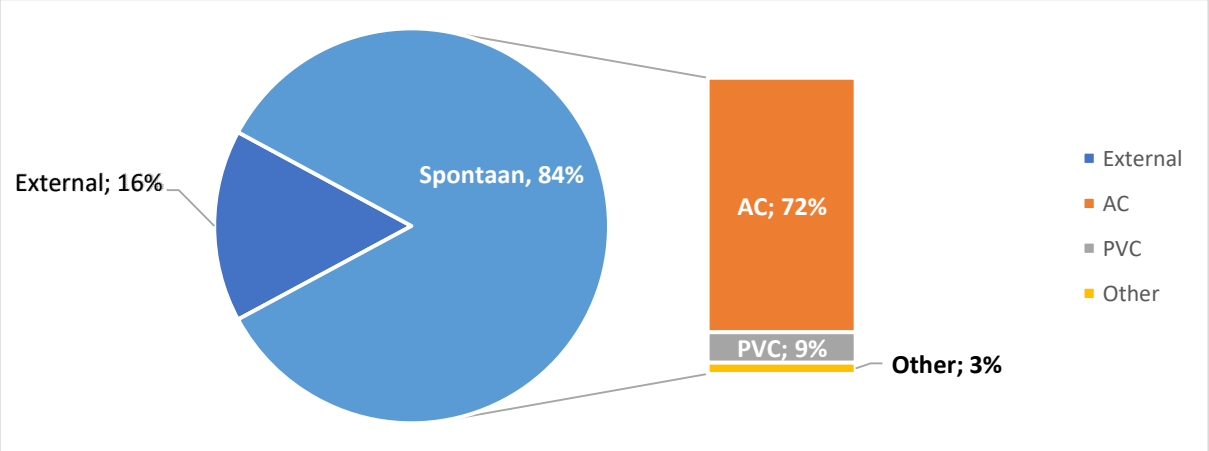


Figure 19. Bursted pipes material groups

A pivotal consideration in understanding burst dynamics involves analyzing the statistics of burst pipes. Figure 19 presents a demographic overview of burst pipes, revealing that only 16% of the total incidents over five years resulted from external accidents. Remarkably, the predominant share, amounting to 84%, is attributed to what is termed spontaneous bursts. Delving deeper into spontaneous bursts unveils a significant correlation with pipe material. Specifically, asbestos cement pipes contribute to 72% of the total bursts, while PVC accounts for merely 9%, with other collective pipe materials contributing 3%. This underscores the crucial role played by pipe material, warranting further exploration.

Upon assessing the age distribution categorized by pipe material, Table 3 elucidates the maximum, minimum, and average ages of burst pipes under distinct categories such as AC-spontaneous, PVC-spontaneous, spontaneous, and external.

Table 3. AC, PVC, Spontaan and External bursts percentage

Age	AC spontaan	PVC spontaan	Spontaan	External
Max	68	66	68	68
Min	42	5	5	2
Average	60	48	58	37

Table 3 demonstrates that AC exhibit bursts in pipes with average age of 60 years, in stark contrast to the 48 years observed in the PVC category. The minimum age for AC pipes is 42 years, whereas PVC pipes register at a mere 5 years. This disparity in age distribution is expected, given the cessation of AC pipe installations after the 1990s.

Consequently, existing AC pipes in the network surpass their operational life expectancy. To visually represent the age distribution across categories, a box and whiskers plot (Figure 20) is generated.

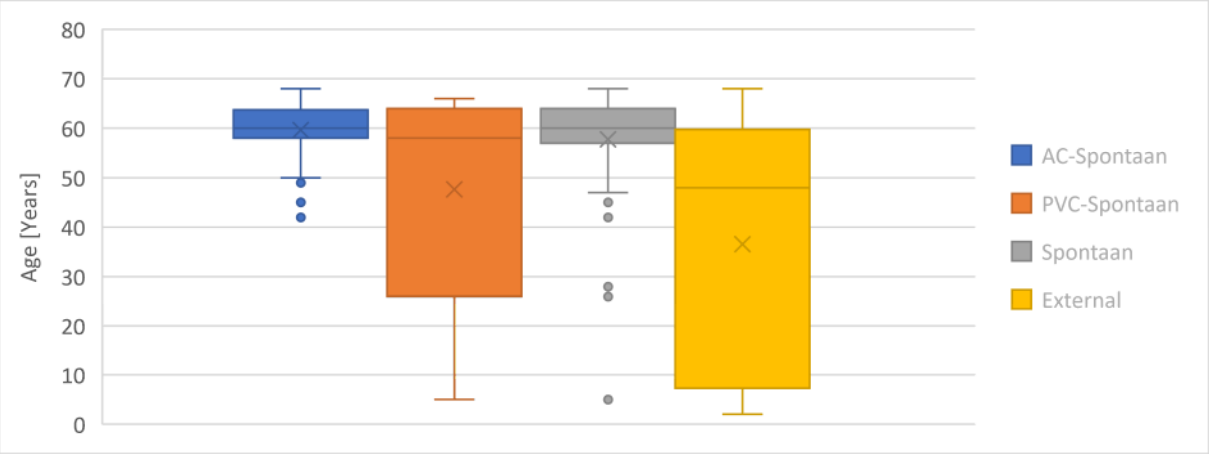


Figure 20. bursted pipe age box and whiskers

Figure 20 elucidates a notably narrow age range for damaged AC pipes, situated prominently on the age axis. Conversely, PVC pipes exhibit a broader age spectrum, even when only spontaneous bursts are being considered, encompassing significantly younger pipes. This reinforces the assertion that older pipes, particularly AC pipes in this context, are more susceptible to spontaneous bursts. Importantly, the prevalence of AC pipes in the dataset does not imply their inherent inferiority or superiority; rather, it is a reflection of their historical installation.

Pipes impacted by external factors exhibit a wide age range, reaching a maximum of 68 years, and encompassing pipes as young as 2 years. Logically, a discernible correlation between pipe age and external factors is absent. In conclusion, the analysis underscores a clear association: the older the pipe, the higher its susceptibility to spontaneous bursts.

3.3.2 Bursts Spatial Distribution

A temporal distribution analysis was conducted to explore patterns, such as uniformity or recurring yearly and monthly patterns. Figure 21 underscores the dynamic and evolving nature of burst occurrences within the network. Notably, there is a substantial deviation observed from one month to the next, and this irregularity suggests that no clear and predictable pattern can be confidently concluded from the available data.

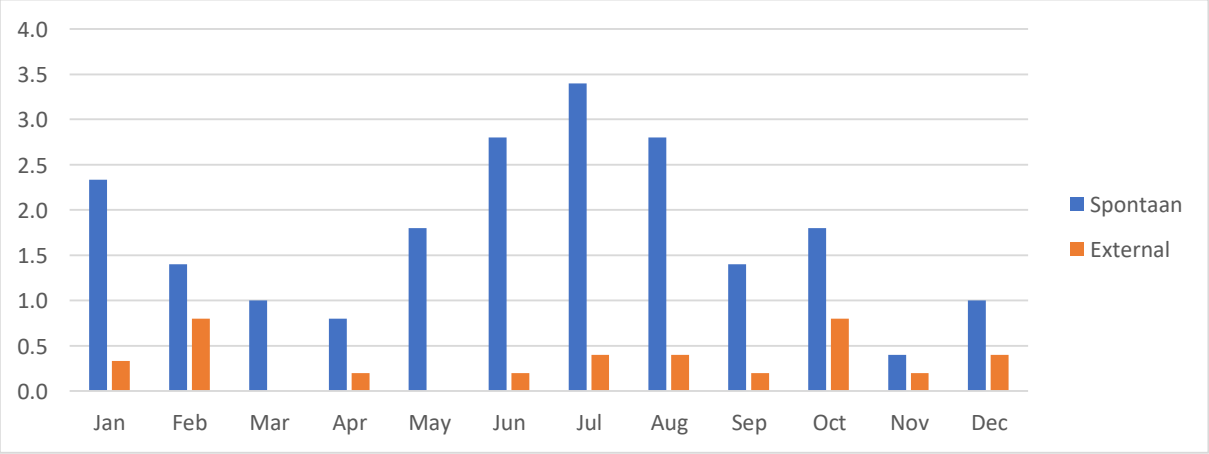


Figure 21. average number of spontaan and external bursts per month

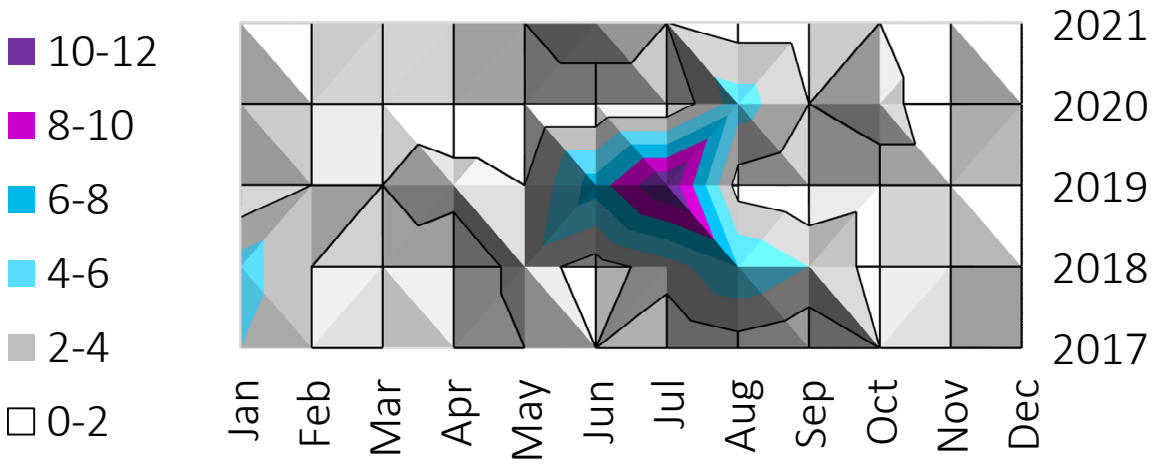


Figure 22. Surface for the number of bursts per month over the recorded period.

As shown in Figure 22, the months of Jun and July of 2019 were exceptionally active for both spontaneous and external categories. Spontaneous bursts accounted for 31 in total, 12 bursts in July alone. Burst caused by external factors accounted for 20 in total, from which 9 occurred in July. It is clear to see that the summer months are more active, but not by a great margin.

3.3.3 Temporal Analysis of Outliers

To comprehend the dynamic behavior of the water distribution system, we extended our analysis to evaluate the number of outliers over a five-day period. This temporal span commences five days prior to a burst event and extends until the event is detected. The rationale behind this extended timeframe is to capture the evolving behavior of the system, acknowledging the temporal lag between a failure occurrence and its subsequent detection. Recognizing the inherent delay in detection, this approach allows for a more comprehensive understanding of the system's behavior over time.

The methodology involves comparing the observed number of outliers during this five-day window to the yearly average of outliers, for each sensor separately. Subsequently, events are categorized based on whether the number of outliers during the specified period exceeds these yearly average for at least one sensor. The findings of this analysis are presented in Figure 23-25.

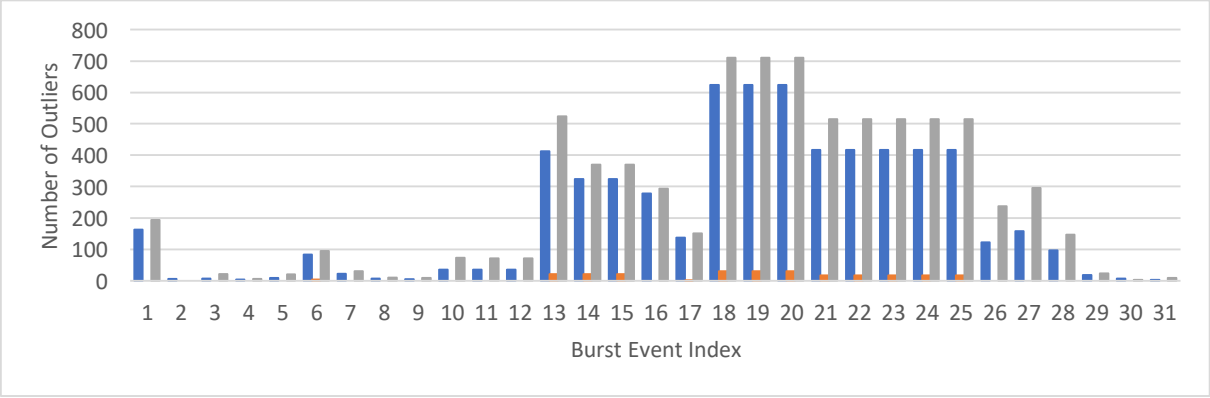


Figure 23. Number of outliers per event for 2019 over different sensors. Each sensor is represented by a different color. 31 events and 3 sensors in total- one sensor is not functioning well.

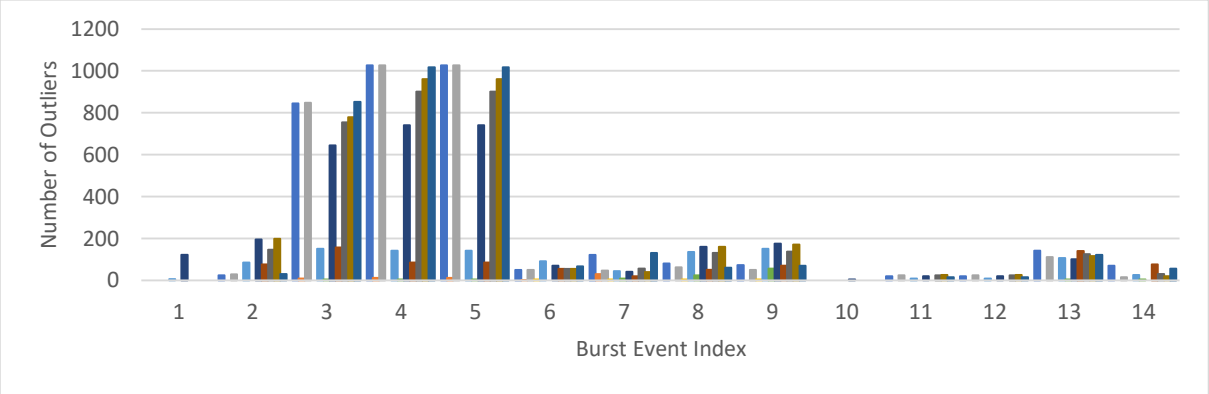


Figure 24. Number of outliers per event for 2020 over different sensors. Each sensor is represented by a different color. 14 events and 11 sensors in total.

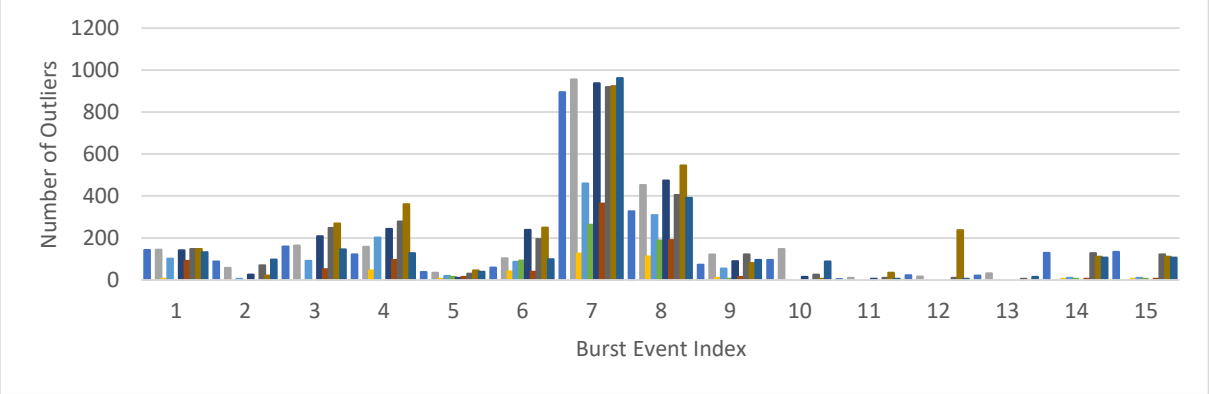


Figure 25. Number of outliers per event for 2021 over different sensors. Each sensor is represented by a different color. 15 events and 11 sensors in total

Furthermore, to gain a holistic perspective on the behavior of pressure outliers over an entire year, we conducted an analysis by calculating the cumulative outliers for each day over the subsequent five days. This comprehensive approach allows us to construct a map illustrating the evolving pattern of outliers throughout the year. By juxtaposing this cumulative outlier map with the timeline of burst events, we aim to discern any discernible correlation between periods of heightened pressure anomalies and the occurrence of bursts.

The outcomes of this analysis are represented at figures 26-28, and provide insights into the temporal dynamics of pressure anomalies and their potential role as precursors to burst events. The subsequent correlation analysis with burst timelines is poised to shed light on the interconnected nature of pressure anomalies and bursts, facilitating informed decision-making for system optimization and reliability.

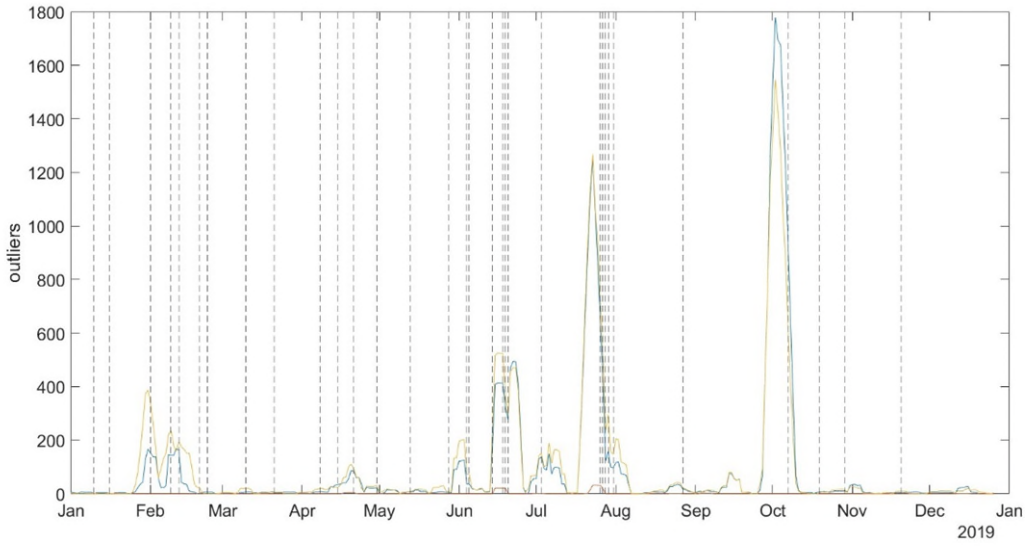


Figure 26. Five days outliers over time from all sensors over 2019, with burst indicated with dashed lines. Each sensor is represented by a different color.

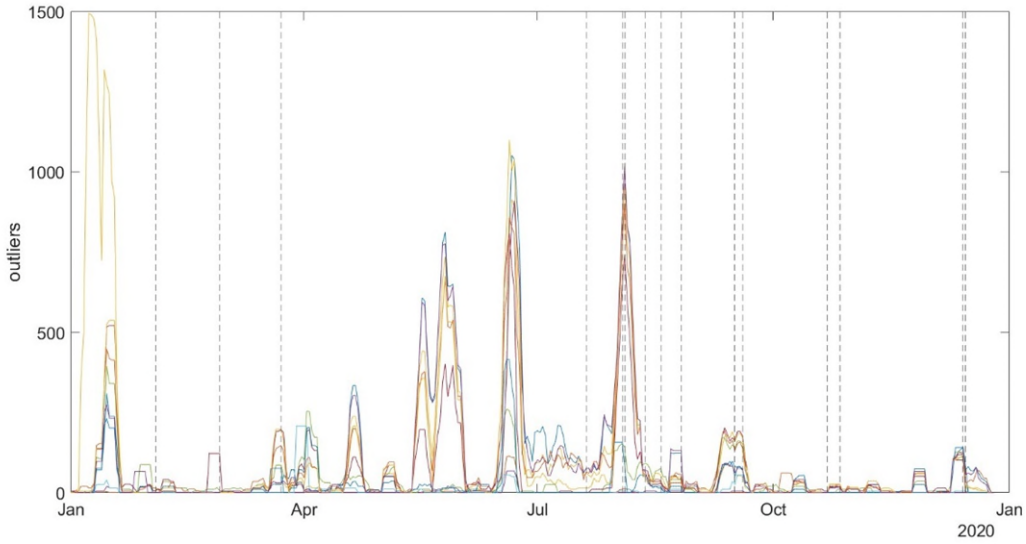


Figure 27.Five days outliers over time from all sensors over 2020, with burst indicated with dashed lines. Each sensor is represented by a different color.

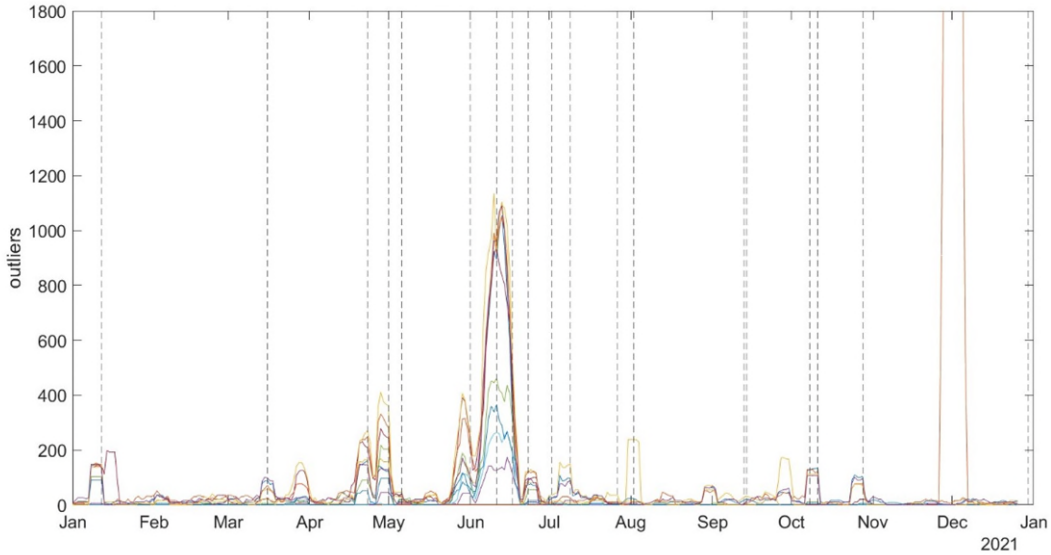


Figure 28. Five days outliers over time from all sensors over 2021, with burst indicated with dashed lines (zoomed in). please see original graph plot in supplementary A. Each sensor is represented by a different color.

Upon scrutinizing Figures 26-28, a discernible correlation emerges, indicating that a considerable number of bursts coincide with the occurrence of outliers. This observation underscores the significance of pressure anomalies as potential precursors to burst events within the water distribution system. However, a nuanced perspective is crucial. While the correlation is evident, it is equally evident that not every instance of abnormal pressure events (characterized by a high number of outliers) results in a burst, and vice versa. The complex interplay between pressure dynamics and burst occurrences suggests that additional factors and conditions contribute to the manifestation of bursts. The results of the burst event probability test are summarized in the Table 4.

Table 4. Burst event probability between the years 2019-2021, based on the pressure outliers analysis.

Year	Real Positives (test 1)	False Positives (test 2)	Random (test 3)
2019	58%	22%	28%
2020	79%	41%	46%
2021	87%	22%	33%
Overall	70%	29%	36%

The results indicate a consistent pattern of higher real positive percentages (70%) compared to random detection rates across the years (36%) tested. This suggests a significant correlation between burst events and pressure anomalies, with the detection system performing better than random chance. Additionally, the relatively low percentages of false positives (29%) demonstrate the effectiveness of the burst event detection methodology in minimizing erroneous detections during non-burst periods. This duality in the relationship necessitates a more comprehensive understanding of the contextual factors that influence the transition from abnormal pressure conditions to observed burst events. Exploring these factors will be instrumental in refining predictive models and enhancing the effectiveness of early warning systems within the WML water distribution network.

The challenge lies in disentangling the intricacies of the system behavior, identifying additional contributing factors, and discerning the thresholds beyond which pressure anomalies manifest as burst events. Further investigation and analysis will be crucial for gaining a deeper understanding of the complex dynamics governing the water distribution system, ultimately facilitating more accurate prediction and mitigation strategies.

Mean RMSE Analysis:

The pressure data over the three-year period was visualized through individual plots for each year, providing a background against which the sensors were overlaid for the five days surrounding burst events. These plots revealed a discernible correlation between pressure levels and bursts. Statistical analysis further supported these visual observations, indicating that, in certain cases, pressures in the five days preceding a burst recording in the system exceeded the average by a factor of 5 times the RMSE. This finding emphasizes instances of abnormal pressure conditions, offering a quantitative measure to identify deviations from typical values.

However, it is imperative to acknowledge that the reliability of sensors varies. While the correlation between high and low pressure and bursts is evident, the reliability of sensors plays a crucial role in the accuracy of these observations. Some sensors exhibit a higher degree of reliability than others, a factor that warrants careful consideration in the interpretation of pressure data. Figure 29 provides a visual representation of these analyses. The identification of reliable sensors and the quantitative assessment of pressure anomalies contribute to a comprehensive understanding of the factors influencing burst occurrences.

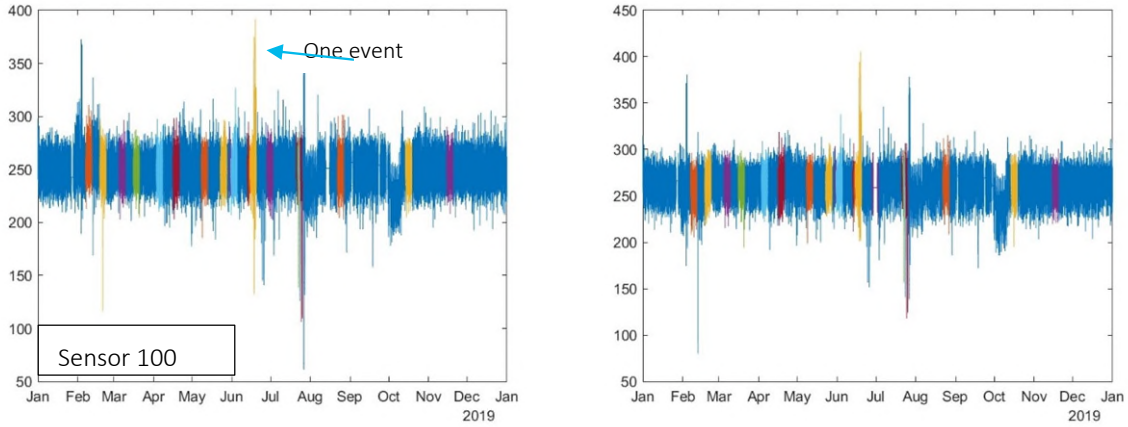


Figure 29. the sensors pressure marked with the burst events and five days prior. It shows the same clear correlation with high pressure and bursts.

3.3.4 Case study conclusions

Again in this case study we see considerable more bursts in AC pipes. While delving a little deeper into age and and burst we see that bursts tend to take place in older pipes. Also for PVC the average ages of burst pipes is 48 years. In contrast, as expected, externally caused bursts happen on pipes of any age. For this case study we again see burst frequency vary with season with a clear concentration in the summer, with July 2019 being an extreme, as well as to lesser extend January.

This case study does suggest a correlation between pressure variations (the number of pressure outliers) and burst clusters (bursts happening closely together in time). The correlation found is 70% while a random test only results in a correlation of 33% suggesting that pressure variation might indeed cause bursts.

3.4 Case Study 4: de Watergroep

For this case study we had access to burst, pressure and network data of the Watergroep for an area west and southwest of Ieper containing among others Poperinge and Heuvelland. The failure data contained 727 failures (or bursts) which occurred between august 2015 and September 2022. The network data consisted of 651 km of pipes of different material and the pressure data of 25 pressure sensors. For 19 of these sensors the location was known. The pressure sensors contained very detailed pressure data which was measured between 3-8-21 and 3-4-22.

3.4.1 Network pipe material composition

The network consists of 651 km of pipes ranging from 25 mm nominal diameter to 300 mm nominal diameter (Figure 31). The materials include among others PVC, steel and AC (see Figure 30). The majority of the network consist of PVC and a large percentage of AC (Vezel-Cement) with some smaller fractions of other materials.

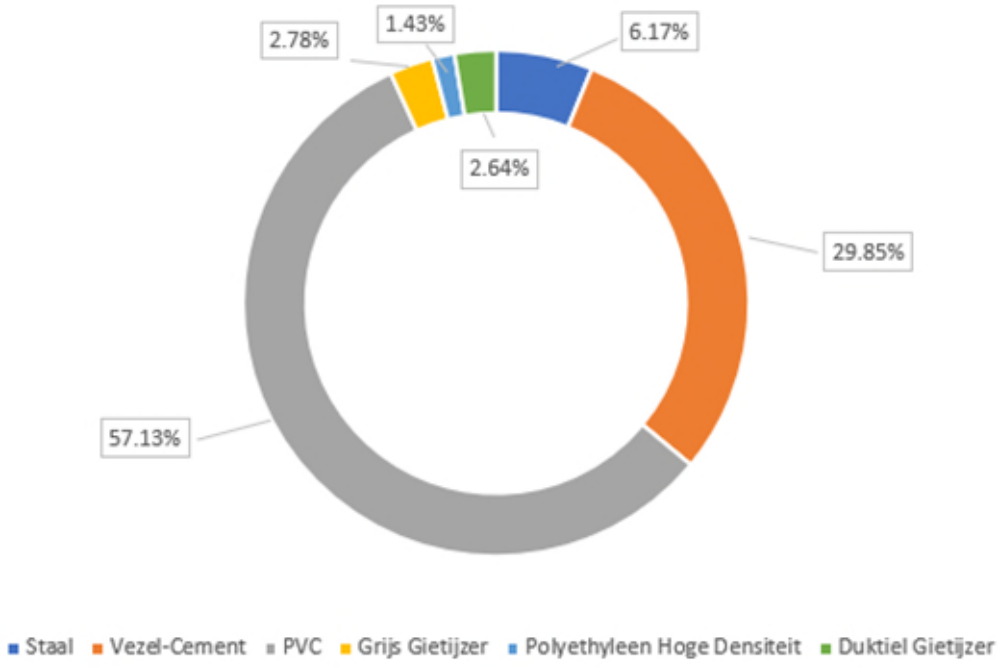


Figure 30. length percentage per material in the network. 0.2% of unknown and 0.01% of ductile cast iron Blutop were omitted from this graph.

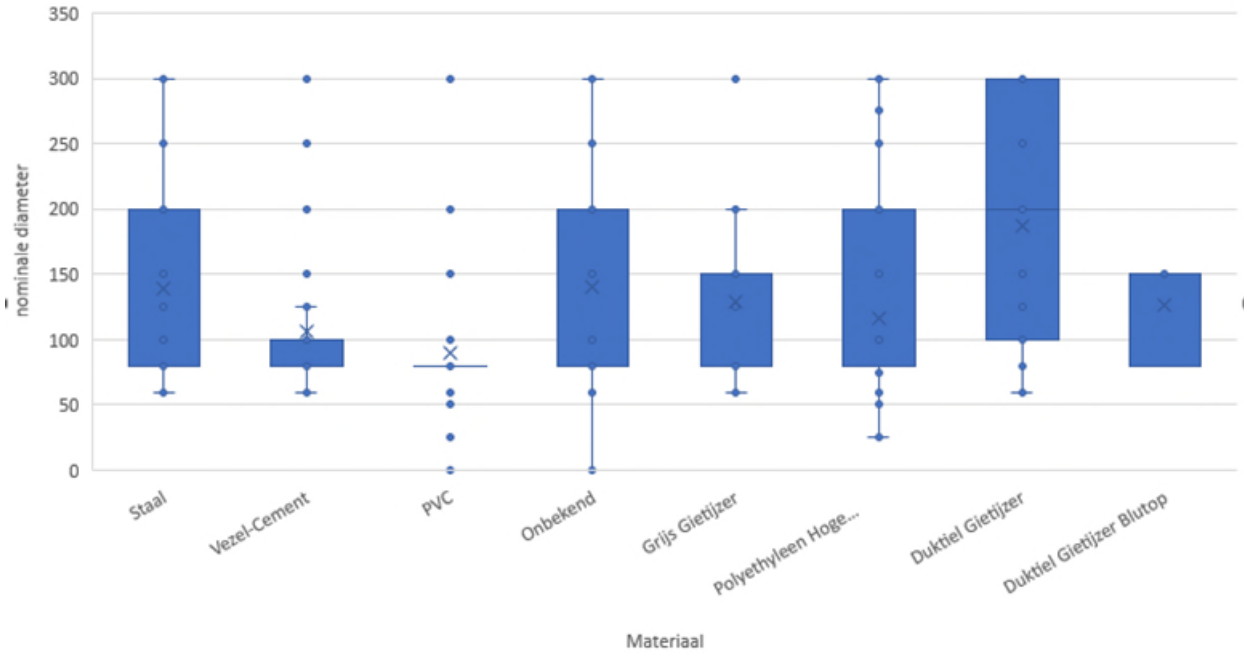


Figure 31. Whisker plot of the diameters per material

3.4.2 General burst information

There are 727 bursts in the dataset. As we can see in Figure 32 of all these bursts 5% are categorised as being caused external, 58% as internal and a large proportion (37%) as unknown. In our analysis we have included the unknown group. Leading to a dataset of 687 bursts. Unless otherwise stated, analysis has been performed with this filtered dataset of 687 bursts.

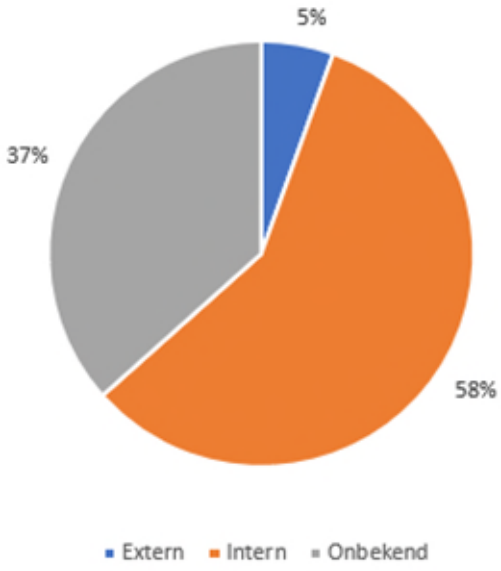


Figure 32. burst after their cause category

In Figure 33 we can see the percentage of bursts per material. When we compare this to Figure 30 we notice the large amount of bursts for steel (21% while the length percentage is only 6.17%) and in lesser amount AC (vezel-Cement 36% of the burst while its length percentage is only 29.85%).

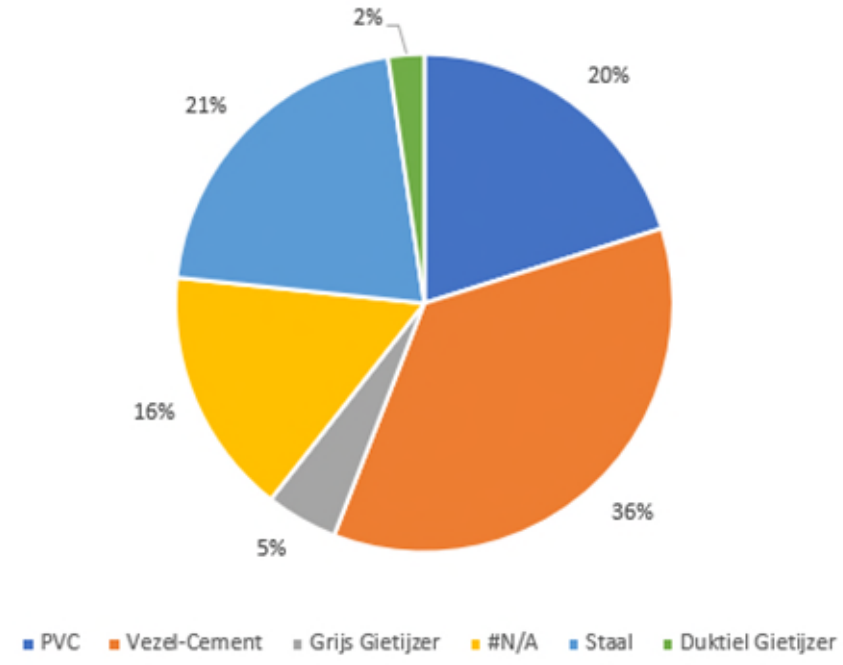
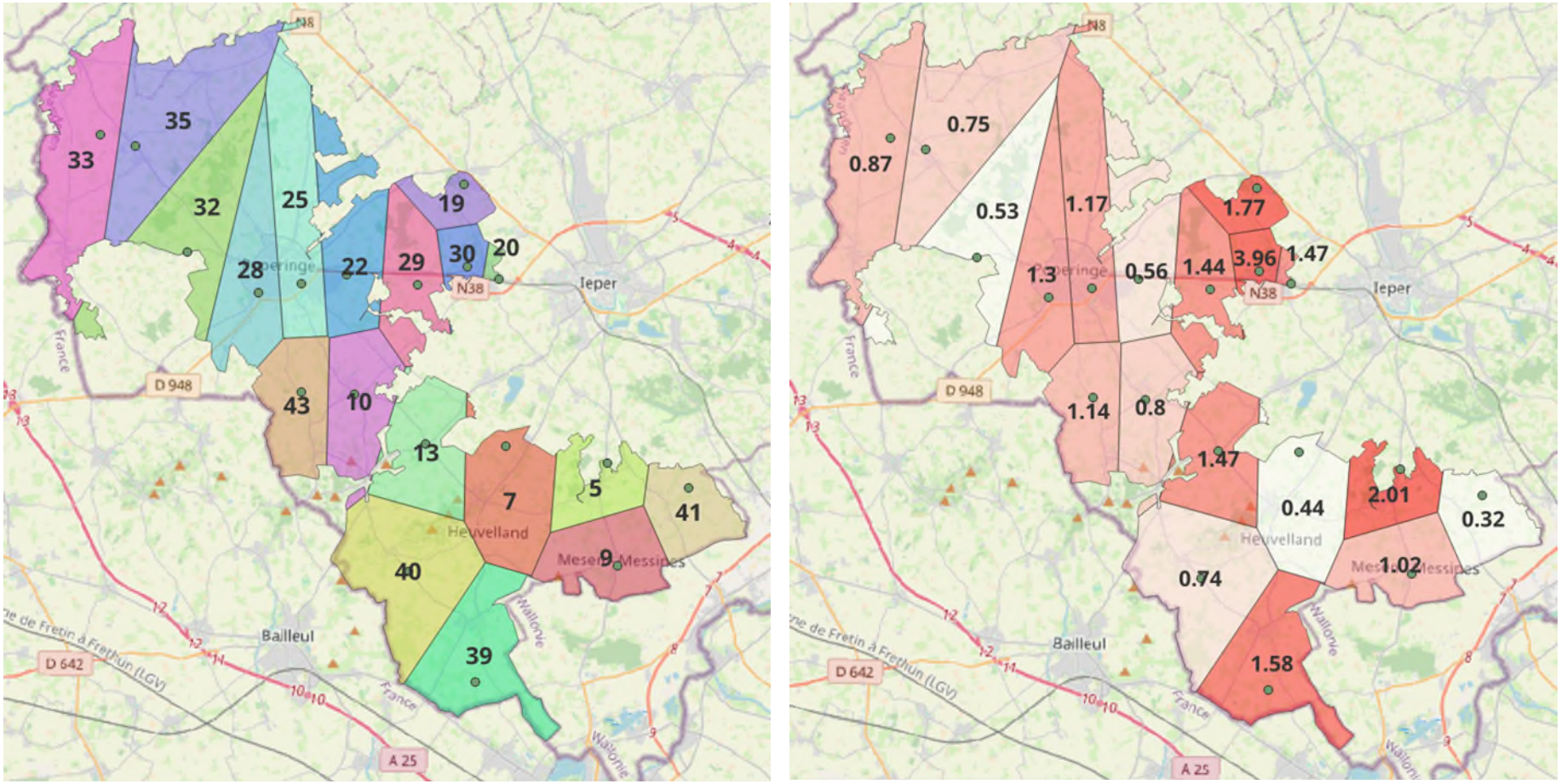


Figure 33. internal and unknown caused burst per material

3.4.3 Initial Spatial burst analysis

In this case study we have detailed pressure data of 19 sensors of which we know the location. We divided the study area into 19 regions using Voronoi polygons.

In Figure 34 we see the study area. In a) we see the different polygons with their respective ID's as we will use them in the rest of this report. In b) we see the amount of bursts per km pipe. There is one area, just west of Ieper (in the east of the case study area with ID 30) which has 3,96 burst per km that really stands out from the rest.



a) b)

Figure 34: Case Study area. The polygons are Voronoi polygons around the different pressure sensors: a) shows the different polygons and their ID. b) shows the amount of bursts per km per polygon.From here we can distinguish one very high bursts per km area namely number 30.

3.4.4 Pressure sensor information

The polygons all contain one pressure sensor which yielded detailed information. With this pressure information we created pressure “fingerprints” (see 2.7). Two examples are found in Figure 35 and Figure 36.

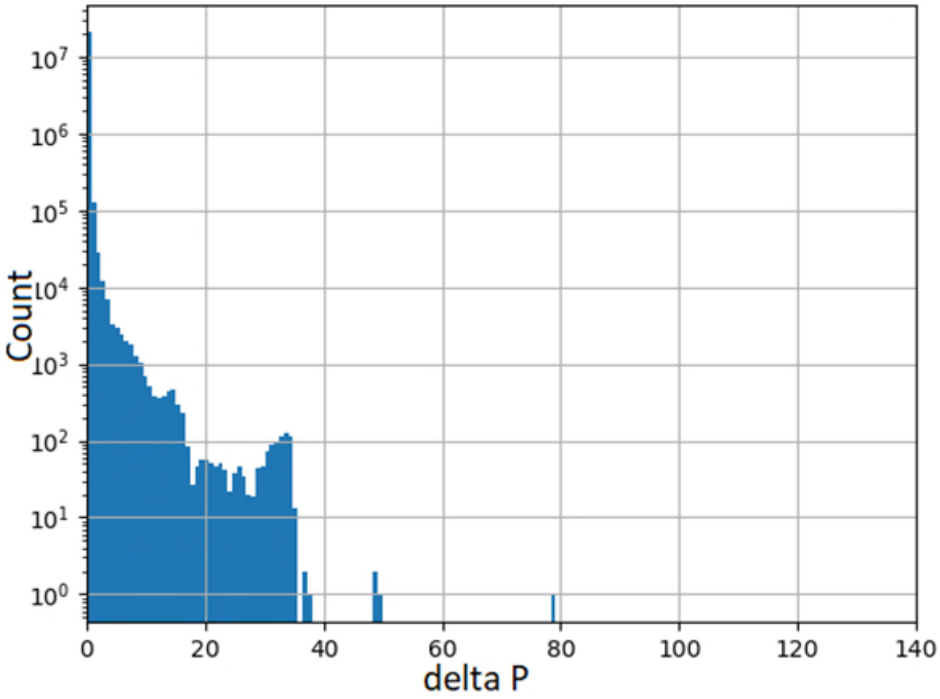


Figure 35. Pressure sensor in polygon 13. This is an example of what we consider a "busy" area. The graph shows a histogram with on the x-axis the value of delta P and on the y-axis (which is logarithmic) the amount of delta P’s with that particular value. For this sensor there is one delta P with a value of almost 80 mH2O delta P’s with values between ~30 and ~35 all appear about 100 times.

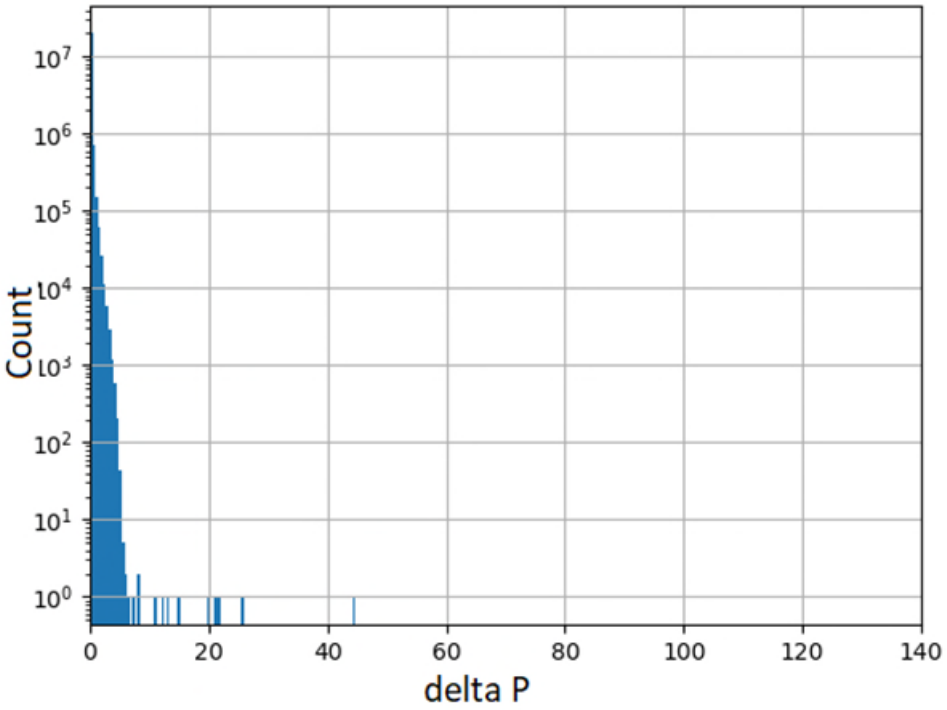


Figure 36: Pressure sensor in polygon 28. This is an example of what we consider a "non busy" area. The graph shows a histogram with on the x-axis the value of delta P and on the y-axis (which is logarithmic) the amount of delta P's with that particular value. For this sensor there is one delta P with a value of 45 mH2O The bulk of the delta P's are below 10.

We then determined the 75% quantile of all these "fingerprints". For the two examples in Figure 35 and Figure 36 these values are 8.72 (polygon 13) and 0.31 (polygon 28). All values are reported in Table .

3.4.5 Analysis of relation between the amount of pressure fluctuations and the amount of failures

Using Excel we performed a correlation analysis (considering we have a limited dataset of 19 areas) between the amount of bursts per km and the 75% quantile of ΔP. This gave the very weak correlation of 0.16 (a perfect correlation would yield 1). We then continued to split the burst per material by calculating the amount of bursts per material per polygon as well as the amount of km of that material in that polygon. Thes numbers are given in Table . Using this we could calculate the amount of bursts per material per km of a given polygon. We performed this exercise for PVC, steel and AC and we excluded polygons that contained less than a km of steel or less than 5 km for AC.

We than again performed a correlation. The results are shown in Table 5.

Table 5: correlation between the amount of pressure changes and bursts per km of different materials

	Bursts per km Steel	Bursts per km PVC	Bursts per km AC	Bursts per km (all materials)
75% quantile	0.31	0.25	0.21	0.16

The correlations are all weak ranging from 0.21 to 0.31. The correlation is highest for steel and lowest for AC but considering the amount of datapoints we should consider them all equal. In Figure 37 we show the amount of bursts per km for these three materials as well as the 75% quantile plotted in the different polygons. As mentioned, polygon 13 can be considered as the "busiest" polygon with a 75% quantile of 8.72. This is also the polygon which has the second highest bursts per km for AC namely 2.5 (the highest amount of bursts per km is polygon 40, but this only

consists of 727 meter of AC). There also seems to be a (weak) correlation for polygons 9 and 19 but there are also polygons (like 28) which have no apparent relation between bursts on AC and pressure fluctuations. The same holds for steel. Polygon 39 can be considered as a busy polygon and it has the second largest amount of bursts per km. Although polygon 30 has the most bursts per km of steel while it is not considered a busy polygon.

3.4.6 Case study conclusions

In this case study we observe that more bursts take place on AC and steel with respectively 36 and 21% of all spontaneous and unknown caused bursts while they only make out 29 and 6% respectively of the network.

In this case study we looked at areas with more or less pressure fluctuations and if those areas are more or less prone to bursts. The correlations are weak at best with the “best” correlation for steel. However considering the limited amount of data (areas) no real conclusions can be drawn from this analysis.

Table 6: data on the bursts on different materials and the total amount of km of that material per polygon. The total also includes the materials ductile and grey cast Iron and unknown which are omitted from this table

Area ID	Bursts on Steel	Km Steel	Bursts on PVC	Km PVC	Bursts on AC	Km AC	Total bursts	Total km	75% quantile of delta P per second
5	0	0	14	16.2	7	5.3	47	23.4	4.03
7	0	0.02	1	18.9	13	15.2	18	40.3	2.51
9	0	0.03	10	18.8	20	11.5	31	30.4	5.3
10	0	0	1	15.7	12	9.4	21	26.3	0.39
13	0	0.05	5	12.6	31	12.5	41	27.8	8.72
19	14	2.8	4	7.3	2	1.1	20	11.3	3.5
20	8	1.8	0	4.3	3	1.4	11	7.5	0.07
22	0	0.8	6	19.2	9	8.4	16	28.7	2.57
25	2	1.1	7	39.1	61	39.5	97	82.7	0.59
28	11	4.4	23	42.6	39	19.5	88	67.8	0.31
29	8	2.8	2	9.7	10	6.3	27	18.7	3.22
30	32	3.0	3	5.6	2	1.2	39	9.8	2.27
32	6	6.5	1	16.1	10	14.8	20	37.4	2.56

33	6	4.0	17	34.5	4	7.9	46	52.6	1.81
35	7	2.5	11	18.8	15	28.1	37	49.9	3.65
39	46	6.6	2	20.2	5	9.2	57	36.0	6.77
40	4	3.6	8	31.3	2	0.7	32	43.0	2.6
41	0	0	5	18.3	0	0	6	18.3	0.09
43	0	0.02	19	21.8	0	1.9	33	28.9	5.35

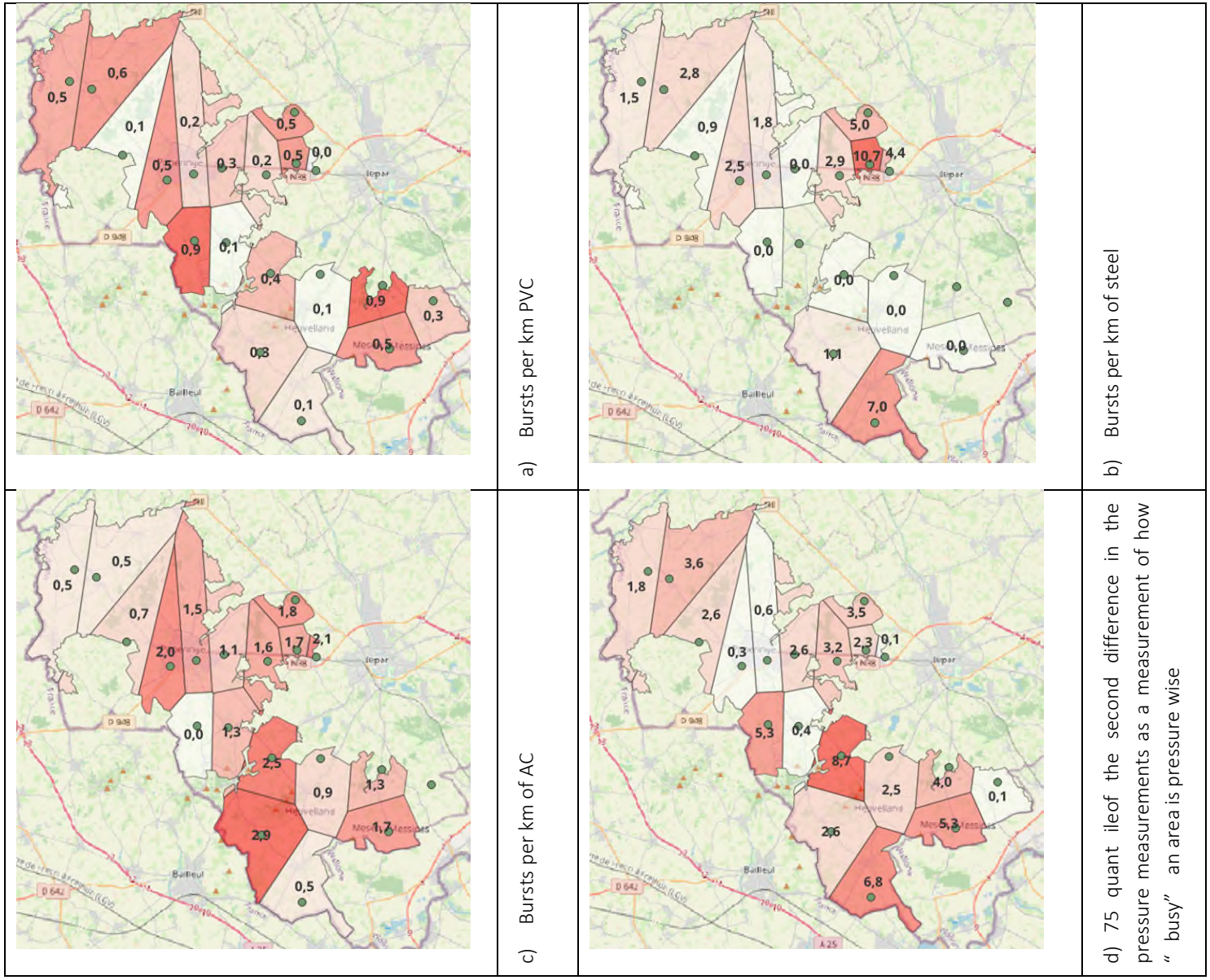


Figure 37: bursts per km of material (PVC, Steel and AC) and a measure of how much pressure changes in an area.

4 Discussion, conclusions and recommendations

4.1 Material and age

All case studies show that some materials are more prone to bursts than others with AC exhibiting a larger proportion of bursts compared to PVC. In the case study of de Watergroep, also steel stands out in the burst statistics. However the case studies of Evides I, Evides II and WML show us that AC is also considerably older than PVC suggesting that age can also be (partly) the cause of the different behavior of different materials.

4.2 Seasonality

For the case studies Evides I, Evides II and WML we see that burst events exhibit distinct temporal and spatial patterns. In the network of the Evides I area, we observed seasonality in burst occurrences, with summer months recording higher numbers, potentially due to factors such as increased water demand and temperature fluctuations. The analysis also revealed a noticeable shift in the ratio of AC to PVC bursts in specific months.

4.3 Bursts causing other bursts

In the case studies Evides I and Evides II we looked into the question whether bursts cause other bursts both using temporal and spatial analysis. The Evides I case study shows that clusters of bursts do not happen particularly close to each other in space.

4.4 Pressure causing bursts

For the case studies Evides I and Evides II we saw no significant correlation between pressure variables like maximum, minimum or fluctuations. For the case study of WML however, where we had more detailed pressure measurements a correlation between pressure fluctuations (in the form of outliers) and bursts was observed.

The investigation into the network of Evides II area revealed a notable increase in burst incidents following a specific transient event on November 3, 2021. Although the transient event is a potential influencing factor, it's essential to conduct further analyses to establish statistical significance and ascertain causality.

In the case study of de Watergroep showed weak correlations between areas with many pressure fluctuations and the amount of bursts. With the correlation with bursts in steel pipes being the “strongest” followed by Ac and PVC. However it should be noted that the correlations are weak at best and that the data is limited.

4.5 Overall conclusions and recommendations

From the case studies it cannot be statistically proven that bursts cause bursts. The case study of WML shows prove of a correlation between pressure fluctuations and bursts. Considering that the case studies Evides I and Evides II do not show this correlation it is interesting to see if that is due to the type of pressure data that was available for these case studies (being less detailed) or whether the area of WML is the exception. The Evides II case study shows a correlation that seems to be significant between a transient event and bursts on PVC pipes. However this is based on only one transient event. It is therefore recommended to look into more transient events to see if the same conclusion can be drawn there. Lastly the case study of the Watergroep shows weak correlations (based on very limited data) between areas with many pressure fluctuations and the amount of bursts. It is recommended to take a further look into this by extending the analysis to more areas.

5 Bedrijfsparagraaf

5.1 Evides, Patrick van den Ende

Voor drinkwaterbedrijven is het interessant om de correlatie te kennen tussen drukken/drukstoten en opgetreden lekkages. Door deze inzichten kunnen maatregelen genomen worden. Denk aan (uitgebreidere) pompsturingen om drukstoten te voorkomen, naleving van draaischema's bij afsluiters en afspraken/maatregelen bij klanten om heftige drukschommelingen in het leidingnet te voorkomen.

Daarnaast biedt het beïnvloeden van drukken kansen om lekkages te voorkomen en daarmee vervangingsinvesteringen in het leidingnet later in tijd te plannen. Bijvoorbeeld om op die manier een geschikter vervangingsmoment te zoeken met andere stakeholders in de buitenruimte.

5.2 WML, Wim Lafeber

Het verband tussen drukschommelingen en lekken wordt goed in kaart gebracht in de casus WML. Of dit verband causaal is, of dat beide min of meer onafhankelijk van elkaar zijn gerelateerd aan de zomer is niet helemaal zeker. Uit specifieke incidenten komt wel degelijk naar voren dat drukschommelingen kunnen resulteren in spontane lekken.

WML rolt nu al een aantal jaren slimme meters uit bij grootzakelijke klanten en doet drukmetingen in het distributienet. Uit deze data komen interessante observaties naar voren: door grillige afnamepatronen (die we direct kunnen toekennen aan specifieke afnemers) ontstaan drukschommelingen in het net, die vervolgens weer invloed kunnen hebben op de sturing op de pompstations. De aanpak is nu om in gesprek te gaan met deze afnemers en om samen te zoeken naar een oplossing om deze drukschommelingen te verminderen. Hopelijk kunnen we hiermee het aantal lekken op termijn reduceren.

6 References

Beuken, R. H., & Moerman, A. (2022). *Uniforme storingsregistratie (USTORE).* Nieuwegein: KWR PCD9.

Dash, A., van Laarhoven, K., & Wols, B. (2023). *Leidingfalen door dynamische belasting.* Nieuwegein: BTO 2023.050.

Moerman, A., & Wols, B. (2015). *Verkennend Onderzoek Verkeersbelasting en Leidingfalen.* Nieuwegein: BTO 2015.004.

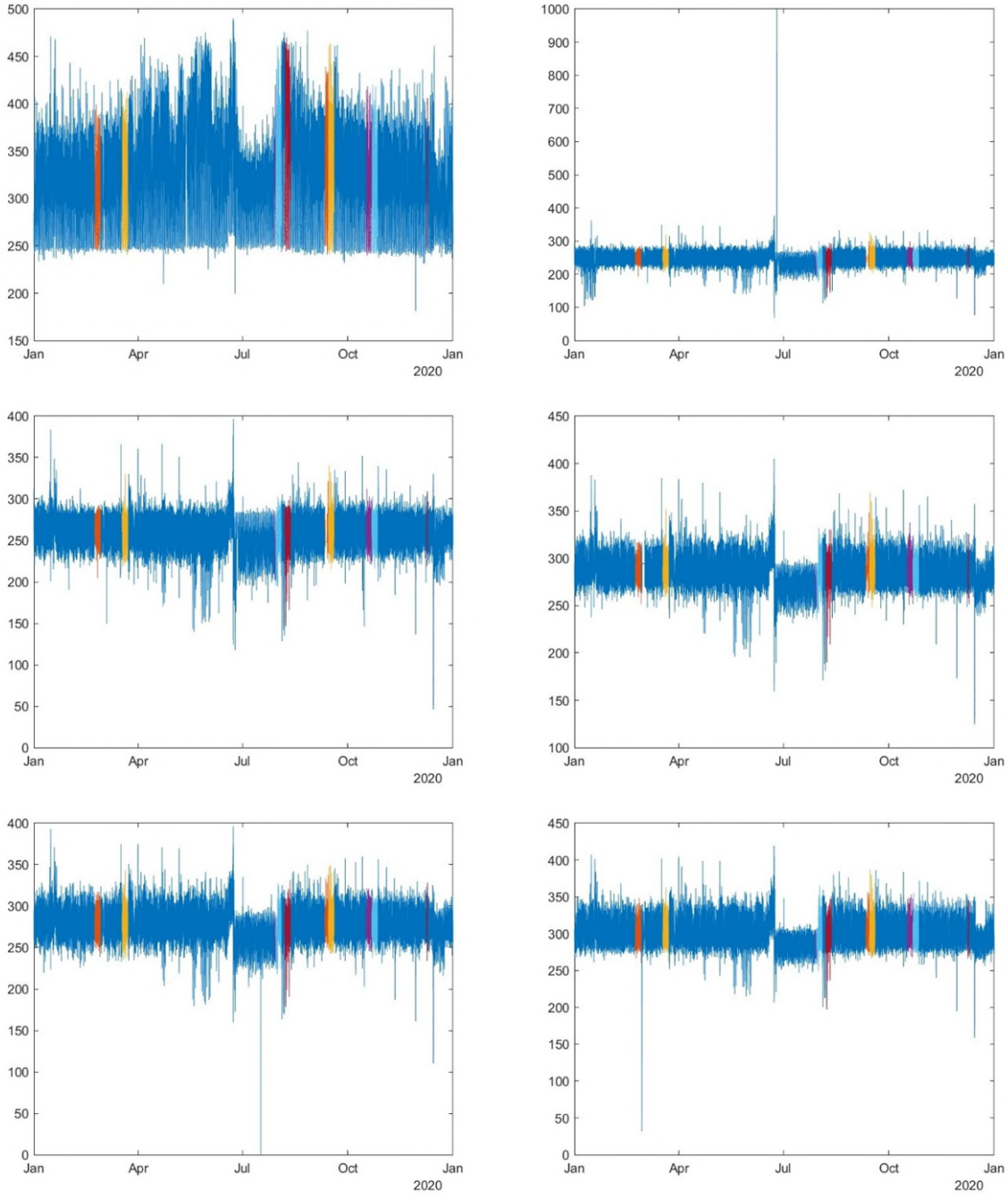
Slaats, P., & Mesman, G. (2004). *Conditiebepaling asbestcement waterleidingen - wanddikte, belastingen -.* Nieuwegein: Kiwa N.V. BTO 2003.039.

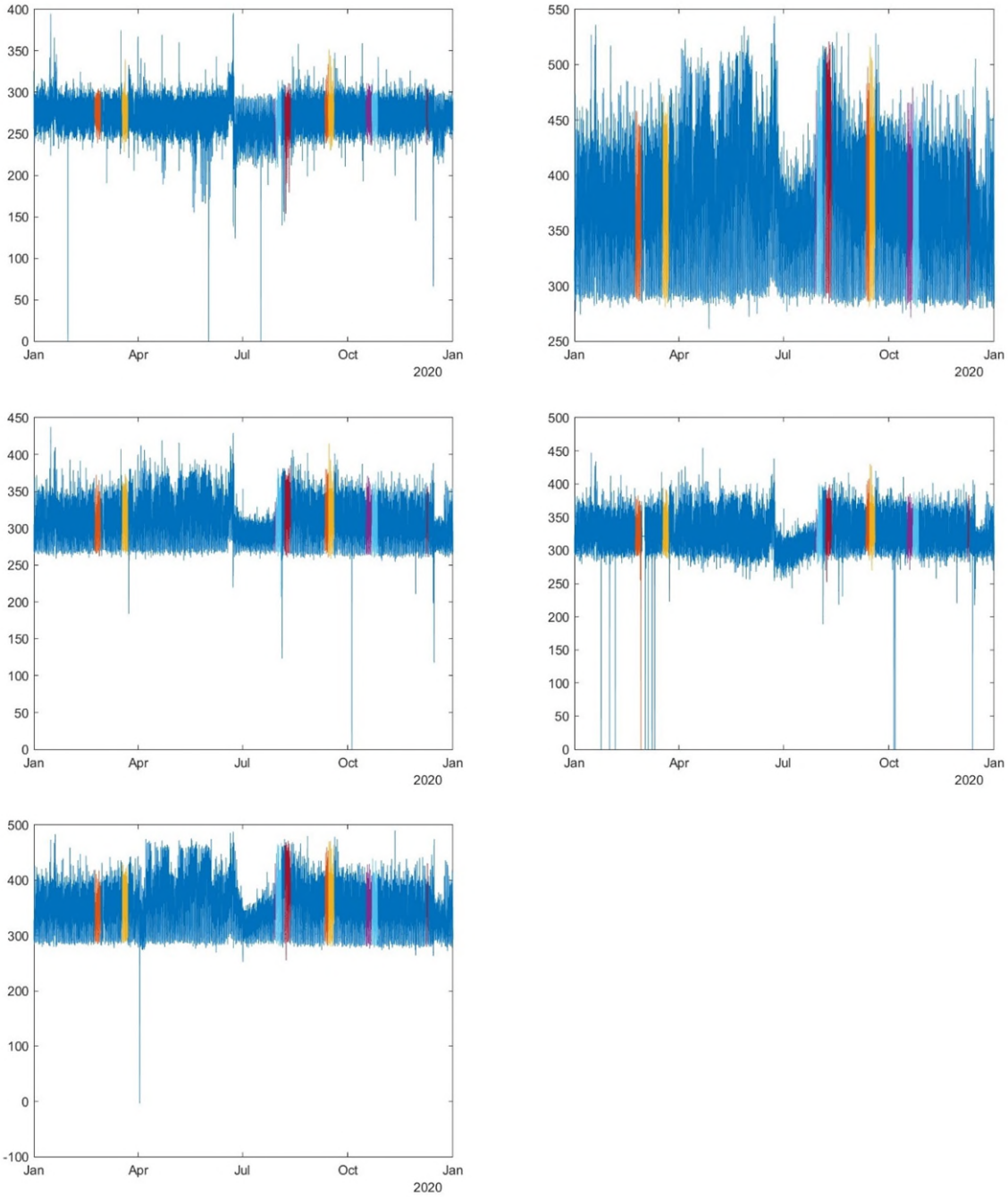
Wols, B., van Summeren, J., Mesman, G., & Raterman, B. (2016). *Fysieke kwetsbaarheid leidingen voor klimaatverandering.* Nieuwegein: BTO 2016.016.

I Appendices

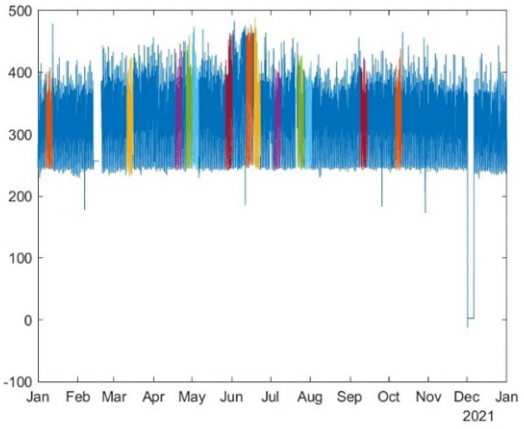
This section Includes detailed tables, charts, and supplementary information supporting the analysis and results presented in the report.

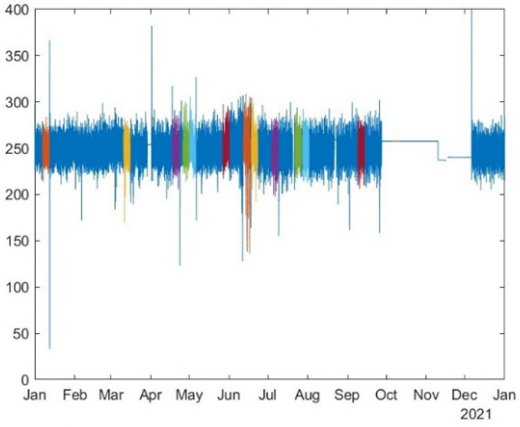
Supplementary B: Sensors plot with bursts mark for 2020

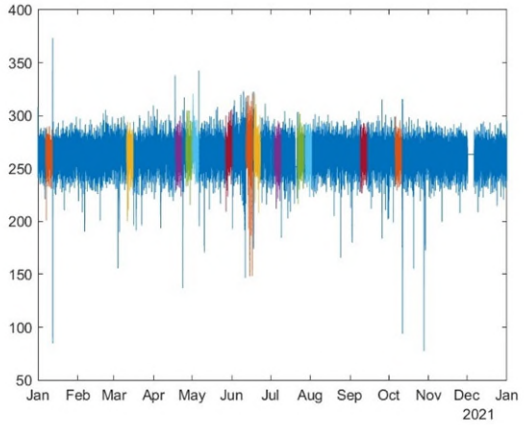


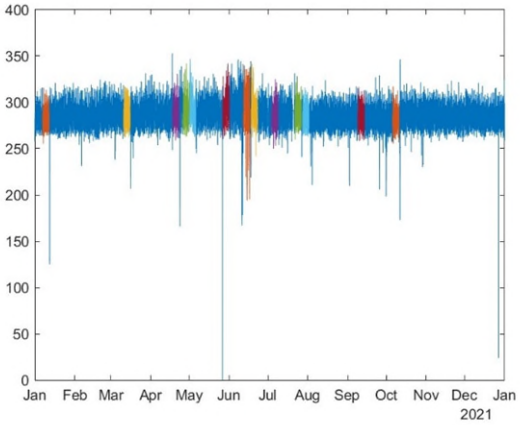


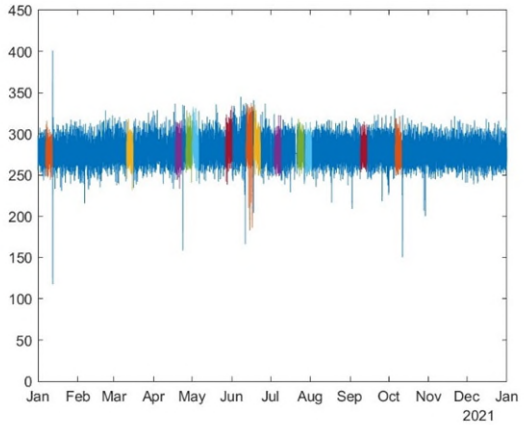
Supplementary C: Sensors plot with bursts mark for 2021

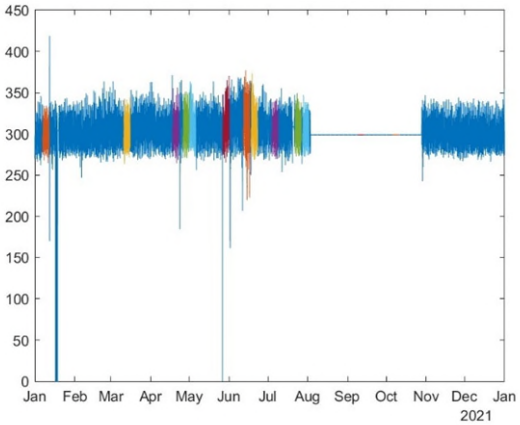

500
400
300
200
100
0
-100
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021

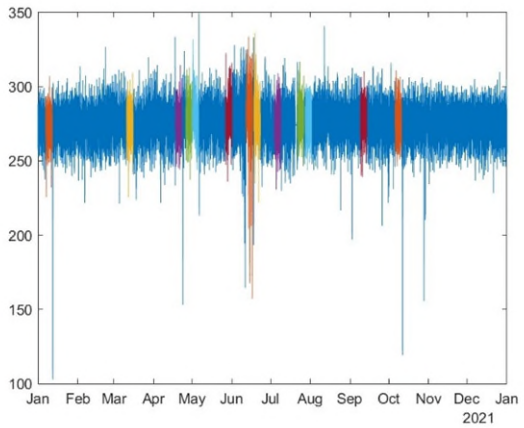

400
350
300
250
200
150
100
50
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021

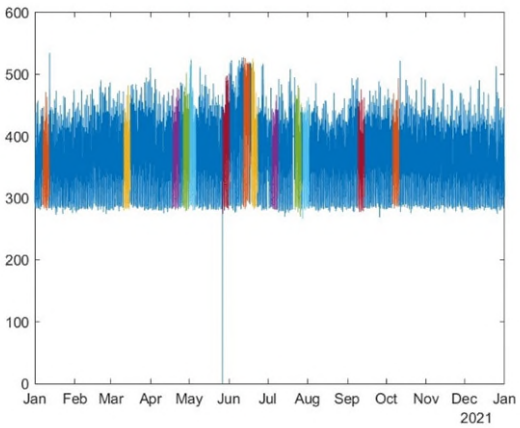

400
350
300
250
200
150
100
50
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021

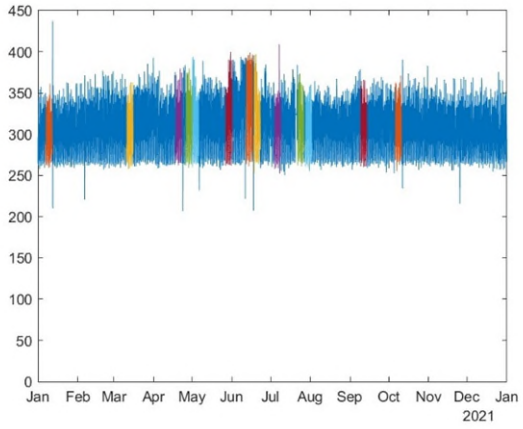

400
350
300
250
200
150
100
50
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021

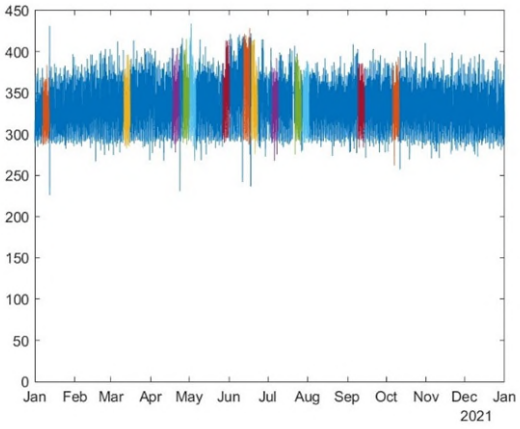

450
400
350
300
250
200
150
100
50
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021

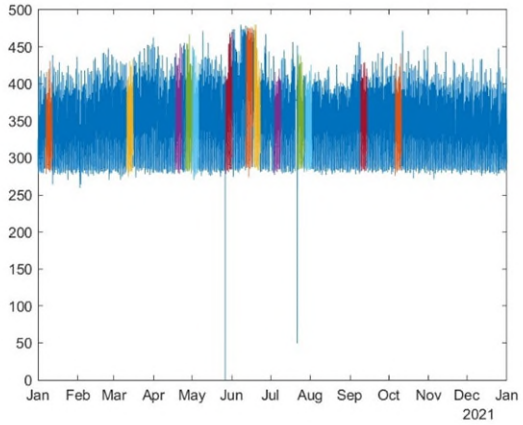

450
400
350
300
250
200
150
100
50
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021


350
300
250
200
150
100
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021

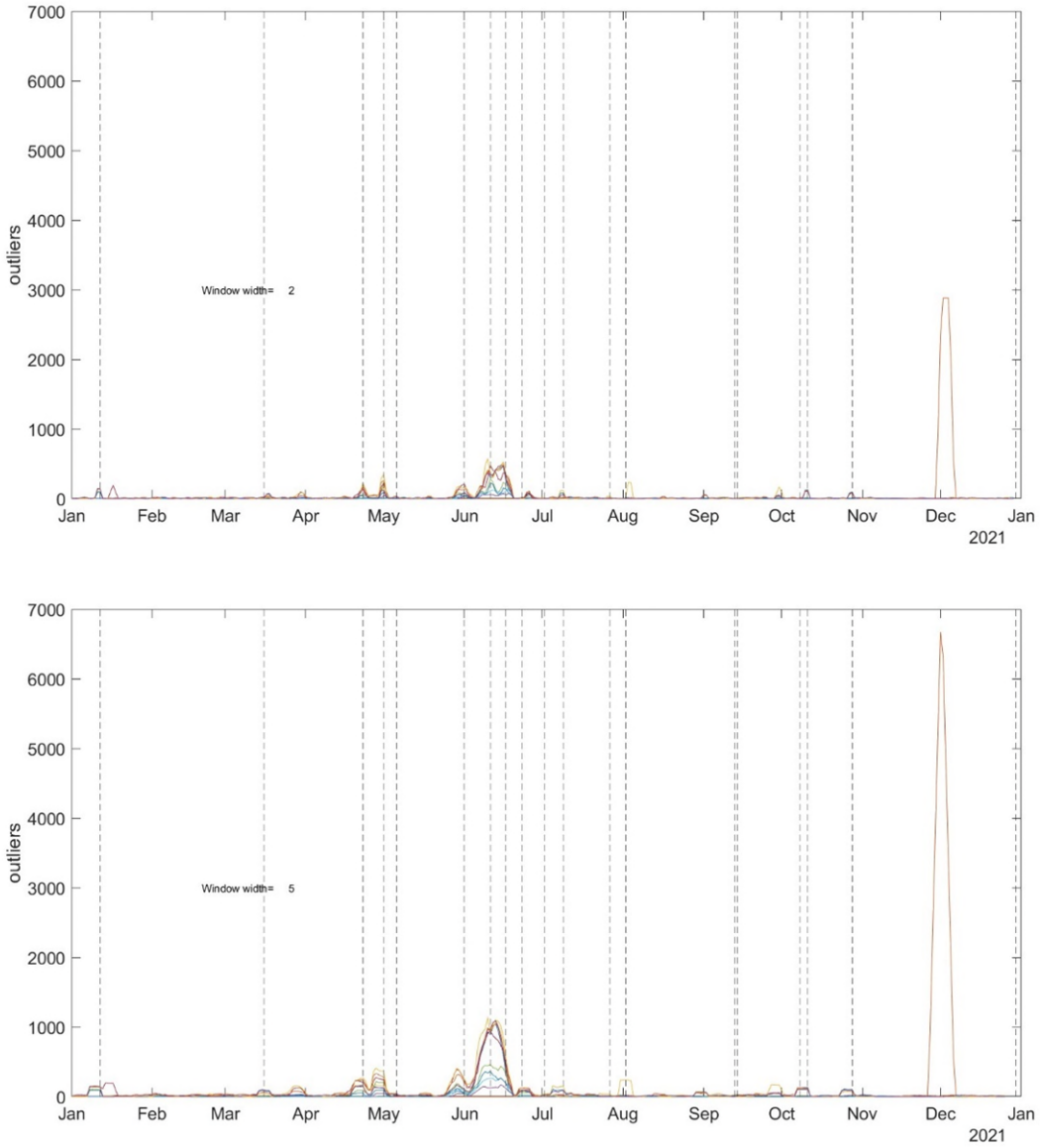

600
500
400
300
200
100
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021

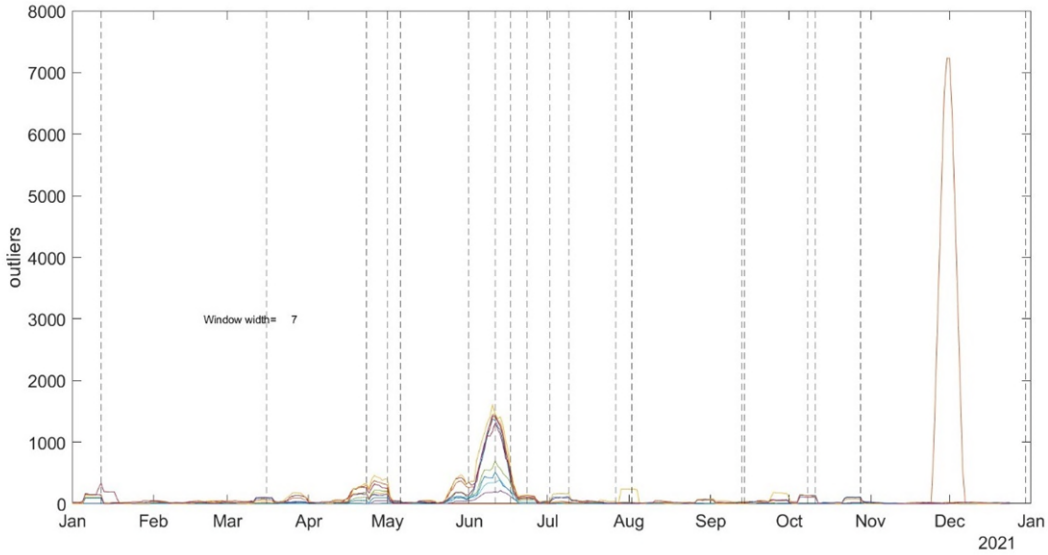

450
400
350
300
250
200
150
100
50
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021

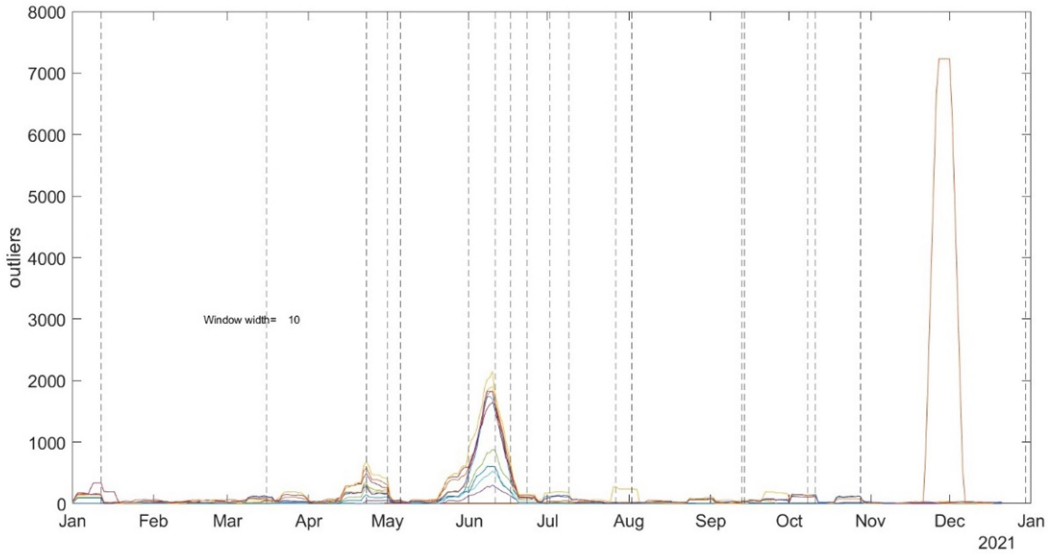

450
400
350
300
250
200
150
100
50
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021


500
450
400
350
300
250
200
150
100
50
0
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2021

Supplementary D: Outliers detected for different window widths of 2, 5, 7, and 10, with standard variation of 3. for the pressure dataset of 2021. Note that, although different y-scale, the second figure (width =5) is the same as Figure 28




outliers
Window width= 7
8000
7000
6000
5000
4000
3000
2000
1000
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2021


outliers
Window width= 10
8000
7000
6000
5000
4000
3000
2000
1000
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2021