The University of Sheffield

Dagmara Starczewska

Pressure transients in water distribution networks: understanding their contribution to pipe repairs

Thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

2016

Pressure transients in water networks: understanding their contribution to pipe repairs

Dagmara Starczewska

Thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Department of Civil and Structural Engineering
The University of Sheffield
Sheffield, UK.

December, 2016

To My Family

Abstract

Drinking water infrastructure functions to provide a service to meet customer demands and health requirements. Pipe repairs are one of the biggest challenges of ageing water infrastructure in the UK and world wide. Pressure transients resulting from sudden interruptions of the movement of the water can be caused by routine value operations. In a single pipeline one extreme event can burst a pipe. However the occurrences and impact of pressure transients in operational water distribution systems were not currently fully understood.

This research developed new insights and understanding of pressure transient occurrences and their contribution to observed pipe repair rates. A large scale field monitoring program, including deploying and managing high-speed (100 Hz) instrumentation for 11 months, was designed and implemented to cover 67 district metered areas (DMA) subdivided into 79 pressure zones. In total 144 locations were monitored. The data was analysed using a novel method, termed transient fingerprint. This allowed the identification of discrete pressure transients and their three fundamental components (magnitude, duration and numbers of occurrences) leading to a quantitative interpretation of pressure transients.

Evolutionary polynomial regression modelling was used to assess the impact of directly measured pressure transient data in context with static pressure, age, diameter and soil variables on 64 cast iron pipes. The analysis suggested that high magnitude, short duration repeatedly occurring pressure transients can have an adverse effect on the pipes. The extrapolation of pressure transient analysis into 7978 cast iron pipes showed inconclusive results suggesting that more accurate pressure transient data is required for each pipe in the network. Additional analysis carried out on 25 asbestos cement pipes, with actual measurements of pressure transients for each pipe, confirmed an adverse effect of pressure transient on water network observed in cast iron pipes.

This research has provided an understanding of the occurrence of pressure transients that has implications on pipe management strategies. Mitigation techniques to locate pressure transient sources based on the project outcomes could be utilised to better manage distribution systems and ultimately reduce future pipe replacements and associated costs.

Acknowledgements

My thanks firstly go to my great supervisors Prof Joby Boxall and Dr Richard Collins for their help, support and guidance thorough this entire process.

Special thanks for the support of the industrial partner Yorkshire Water for assistance with providing field locations and asset data. Thanks to the Department of Civil and Structural Engineering for financial support.

I would like to thank all amazing people I have met on the way. Special thanks to my dear friend Dr Jose-Juan Toriz-Garcia for endless advice and support. My friends Dr Daniel Paredes-Soto and Dr Miguel Juares for their valuable insights. Bryan Penny for his friendship and shearing his views. All my other friends for being with me all this time.

I would like to thank colleagues from Pennine Water Group (PWG): Dr S.Mounce, Dr S.Husband, Dr W.Shepherd, , Dr V.Speight, Dr S.Doncaster, J.Butterfield, S.Jones, Dr S.Fox, Dr M.Al-Safar, Dr M.Abdel-Aal, Dr J.Zhu, Dr G.Vesviano, M. Muthusamy, A.Sriwastava, G.Sailor, P.Raven, Dr W.Furnass, Dr F.Sonnenwald, Dr M.Rubinato, Dr K.Fish and everyone from D106 and D105 for creating a great work environment and support.

Finally, thanks to my family for their support, patience and understanding through the whole process.

Publications

- Starczewska D., Collins R., Boxall J.B., A method to characterise transients from pressure signals recorded in real water distribution networks, Proceeding of the 12th International Conference on Pressure Surges, Pressure Surges, Dublin 2015
- Starczewska D., Collins R., Boxall J.B., Occurrence of transients in water distribution networks, Proceeding of the 13th International Conference on Computing and Control for the Water Industry, CCWI, Leicester, 2015
- Starczewska D., Collins R., Boxall J.B., Transient behaviour in complex distribution network: a case study, Proceeding of the 12th International Conference on Computing and Control for the Water Industry, CCWI, Perugia, 2013

Publications in preparation:

- Starczewska D., Collins R., Boxall J.B., Application of pressure transients to modelling of pipe failure rates, to be submitted in Water Research
- Starczewska D., Collins R., Boxall J.B., Characterisation of pressure transients recorded in real water distribution system, to be submitted Journal of Hydraulic Engineering

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

The function of drinking water distribution systems is to meet customer demands and public health requirements. The ageing underground assets are in varying degrees of structural and functional deterioration which often results in increased frequency of pipe repairs and significant water losses. Water companies report 0.14 repairs/km/year. 30% of the supplied water in the UK is lost through leaks (Savic and Banyard, 2011). Pipe repairs can have dramatic consequences not only in terms of water loss, but also costly repair of structural damage to surrounding infrastructure due to associated flooding. The protection of the infrastructure, management and control of failure rates have become high priorities for water companies. Companies face the challenge of significant financial investment to upgrade or replace deteriorating pipe infrastructure to continue delivery of safe drinking water and meet leakage reduction targets. Accurate assessment of pipe performance and condition is typically constrained by the location (i.e., underground, buried) of the majority of water networks. To reduce leakage while maintaining current infrastructure, the causes of pipe failures need to be better understood.

Identification of the crucial factors that influence pipe repairs has become the focus of widespread research (Shamir and Howard, 1979; Constantine et al., 1996; Lambert, 1998; Kleiner and Rajani, 2002). Many factors which could influence the condition of the network have been identified; yet despite extensive research there is no definite factor that causes pipe failures. Currently identified factors do not fully explain the patterns of pipe failures which occur within networks. This suggests that there may be other underlying contributory factors not commonly recognised.

Pressure transients can occur within a few seconds in any hydraulic system as a result of rapid changes in velocity. The potential causes of pressure transients are generally well understood as valve operations, pumps switching or even sudden increase in demand (Kirmeyer et al., 2001;

Misiunas et al., 2003), but the phenomenon itself has not been fully associated with contribution to observed pipe repair data, particularly in complex, operating networks. This could be due to the fact that pressure transients have been recognised as a factor associated mainly with damage to water machinery (i.e. pumps (Wylie, 1965; Morris, 1967)). From the literature, it is known that catastrophic failures are primarily due to extreme events observed usually in a single pipeline (Bergant et al., 2006). A series of such failure events were documented by Jaeger (1948). Alternatively, lower intensity events may induce fatigue over long term exposure. Subsequently, the level of deterioration occurring in distribution systems due to the presence of transients is an unknown quantity. Smaller magnitude, repeated pressure transient events could also impose excessive loading on a pipe, successively weakening it and eventually contribute to its failure. Furthermore, if the pipe has been already weakened, for instance, by corrosion, the exposure to cyclic loading of pressure transients (whether significant or not) could also accelerate its failure.

To date, the existence of the pressure transient phenomenon is water networks has been questioned and not fully documented. This has widely persisted due to the lack of hard evidence and knowledge about the occurrence and magnitudes of not only extreme events but also lower magnitude pressure transients in complex water networks. Difficulties in obtaining high-resolution pressure data have been mainly caused by a lack of equipment, technology and knowledge of high-speed pressure monitoring. Current advances in sensor technology have enabled continuous and long-term field monitoring, making such data obtainable. Effective pipe failure modelling techniques could therefore be employed to re-evaluate causes of pipe repair records where transients, along with deterioration due to corrosion, may be seen as a crucial factor.

1.2 Thesis layout

Chapter 2 provides background literature describing firstly pressure transients, causes and current knowledge about pressure transients in complex networks. Following sections address current knowledge about pipe failure factors and review of pipe failure modelling techniques. The aims and objectives of this research are stated in Chapter 3. Chapter 4 contains the design and implementation of a fieldwork pressure transient data collection. Chapter 5 presents the method to identify and characterise discrete pressure transients from continuously recorded time series data. In this chapter fundamental pressure transient characteristics were identified as magnitude, duration and number of occurrences and applied to real data. Chapter 6 presents modelling methodology based on different type of cases incorporating pressure transients and other variables: static pressure, age, diameter and soil. Chapter 7 presents combined results from Chapters 4, 5 and 6. The discussion across all chapters are presented in Chapter 8. Chapter 9 contains final conclusions and

recommendations for future work.

Chapter 2

Literature Review

This literature review presents and summarises findings from investigations into pressure transients in Water Distribution Systems (WDS) and the associated impact due to their presence. Pressure transients, sometimes referred to as surge or water hammer, are pressure waves in WDS. Pressure transients generally occur in any hydraulic system where rapid changes in velocity occur due to a change of kinetic energy of the fluid into potential energy. The phenomenon is not well understood in WDS and it has been hypothesised that these fast changes in pressure have the potential to damage the infrastructure and reduce the operational pipe life expectancy and increase repair cost.

This literature review will first present the current knowledge regarding pressure transients, then highlight important field studies before summarising their findings. Particular importance will be given to the techniques used in burst modelling studies to discern the causes of pipe failures. This is considered valuable as the most of these techniques have not been applied to investigate pressure transient data.

2.1 Water network hydraulics

Water distribution systems provide a safe and high quality continuous supply of water regulated by standards and meeting demands of their customers. The official body who regulates water industries in England and Wales for flow and pressure standards is the Office of Water Services (Ofwat), a total of 21 current water companies.

The UK water network is subdivided into distribution management areas (DMAs) which are hydraulically defined areas introduced in 1990s to provide more effective operation and maintenance management and to control water loses. The continuous measure of static pressures and flows at strategic (critical) points within (DMAs) or otherwise pressure control zones is a current water industry practice to satisfy regulations (Savic and Banyard, 2011).

2.1.1 Steady state pressure

The minimum static pressure which water companies are required to maintain is 0.7 bar (7 m) (Ofwat, 2008) or 1 bar (10 m) at nine l/min (`http://www.ofwat.gov.uk/publications/water-pressure/`). Ofwat requires water companies to annually report low pressures in the level of service indicator known as the Director General Measure 2 (DG2)(Ofwat, 2004). Water Industry Act 1991 (HMSO, 1991) merely requires a pressure sufficient to reach the top floor of every building. In addition, to meet fire-fighting demands systems have often been over designed (over sized) (Savic and Banyard, 2011). Large diameter pipes are subjected to lower flow velocities and high pressures. The maximum static pressure values can be over 100 m and higher values can exists (Lambert, 1998). Static pressures are constrained by local topography (Lambert, 1998).

Some studies (Lambert, 1998; Thornton, 2003; Morrison, 2004; Boxall et al., 2007) have noticed that internal static hydraulic pressure is probably the single and the most important factor controlling pipe failures (e.g., leakage). To reduce and control leakage pressure management schemes were introduced across the UK since the early 1980s. Pressure reduction schemes are now widely implemented across UK water companies by installing pressure reducing valves (PRVs) at the inlet of susceptible DMAs (Savic and Banyard, 2011). The installation of PRVs within a network requires ongoing maintenance and monitoring which, if not provided, may result in unstable network conditions (Savic and Banyard, 2011).

2.1.2 Pressure transients

Extreme changes in flow conditions in a network have the ability to create considerable high and low dynamic pressures in pipe systems, proportional to the change in velocity if sufficiently rapid. The fluctuating oscillations can frequently exceed steady-state operating pressures in magnitude, with the lowest pressures being negative (Gullick et al., 2004). The changes in pressure and velocity travel around the system as waves, which are known as pressure transients. In complex WDS as the wave passes a network feature, e.g., junction, pipe joints or valve, part of its energy is reflected back to the source, part is transmitted further and small proportion is absorbed by the feature (Lighthill, 1978).

Fluid properties, along with the structural characteristics of the pipe, govern the behaviour of transients in WDS. Exact equations simulating the maximum pressure change and wave velocity are in the form of complex coupled hyperbolic differential equations (Wylie and Streeter, 1978; Chaudhry, 1987; Massey, 2006). The widely accepted Joukowsky equation (Thorley, 1991) gives the change in transient pressure head caused by the instantaneous change in velocity. As the instantaneous change in velocity is not physically possible the equation gives an idealised situation

:

$$\Delta H = \pm \frac{c\Delta V}{g} \tag{2.1}$$

Where ΔH = change in pressure (m), c = wave speed (m/s) ΔV = change in velocity (m/s), g = acceleration of gravity (m/s^2). Wave speed can be found using 2.2 (Wylie and Streeter, 1978):

$$c = \sqrt{\frac{\frac{K}{\rho}}{1 + \frac{K}{E}\frac{D}{e}}} \tag{2.2}$$

Where: K = bulk modulus of the fluid (N/m^2), ρ = density of the fluid (kg/m^3), E = elastic modulus of the pipe material (N/m^2), D = internal diameter of the pipe (m), and e = pipe wall thickness (m).

The Joukowsky equation applies to a single pipe system as represented in Figure 2.1. To complete one period, Figure 2.1 (a), a wave travels *4L*, where L is a length of a pipe, as seen in Figure 2.1 (b).

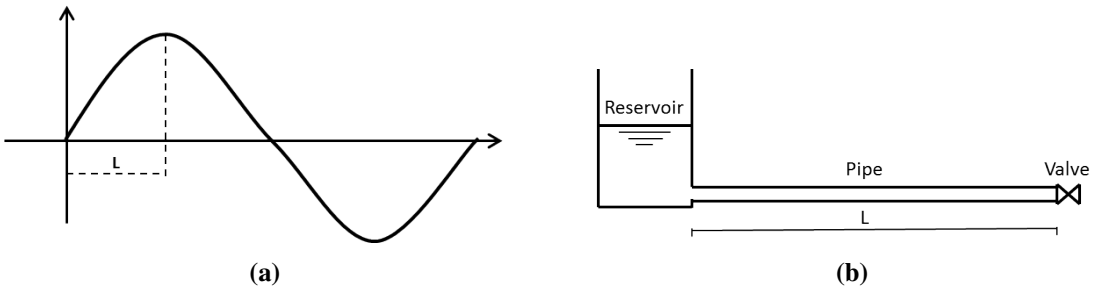


Figure 2.1: Schematic of a pressure transient wave in a simple water network consisting of a reservoir, a single pipeline and a valve, (a) a one period of transient wave, (b) a simple water network.

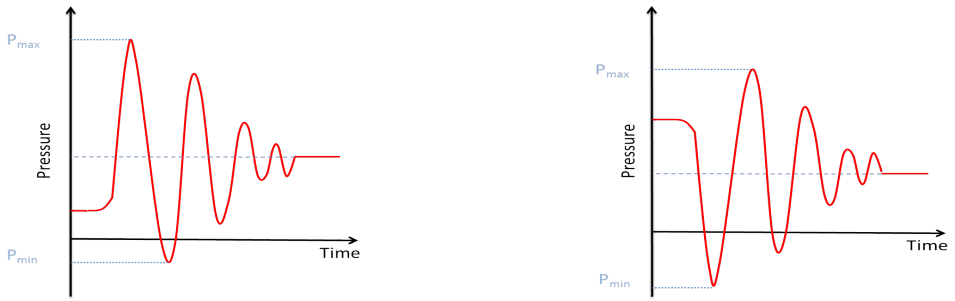
One period of the pressure wave T is equal to *4L* as seen see Figure 2.1 (a) and is directly related to the length of the pipe represented by a reflection time formula 2.3 (Thorley, 1991):

$$T = \frac{4L}{c} \tag{2.3}$$

Where: T = duration of one period of the pressure wave (s), L = pipe length (m), c = wave speed (m/s).

Transient upsurge (Figure 2.2 a) and downsurge (Figure 2.2b) can be described as a sudden change in velocity causing, an initial pressure rise/fall due to the change of kinetic energy into potential (pressure) energy in the fluid. The subsequent pressure oscillations decrease in magnitudes as energy is dissipated, mainly through frictional losses (Massey, 2006).

Pressure transients were also interpreted by energy approach to evaluate a system performance. Internal energy in the pipe during transient event was related to water density, cross-sectional area, acceleration due to gravity, wave speed and head rise (Karney and McInnis, 1990). The rate of



(a) **(b)**

Figure 2.2: Examples of pressure transients. (a) Upsurge and (b) downsurge pressure transient schematics with frictional loses leading to damped pressure oscillations.

change of internal energy was then related to power.

Some researchers (Kaliatka et al., 2006; Bergant et al., 2008a,b) identified that one of the parameters affecting the pressure wave is the visoelastic behaviour of the pipe wall. The authors concluded that visoelasic plastic pipes attenuate pressure transients fluctuations and increase the dispersion of the pressure transient event. According to the authors, leaks and blockages contribute to damping of transient pressure wave.

2.1.3 Causes of transients

Rapid changes in the fluid velocity in a pipe generating pressure waves indicates that there are many possible causes of transients. Fully documented causes of transients are mainly associated with valve operations, pump regimes, changes in demand or pipe bursts (Misiunas et al., 2003). Other researchers claim that improper operation of surge protection devices also cause transients (Kroon et al., 1984). Among all currently known causes, changes in pump operation are believed to account for the majority of transient related problems in WDS. Kirmeyer et al. (2001) provided an extensive list of transient causes, including start up and shut down of a booster pump, altitude valve closure, sudden change in demand, surge/feed or storage tank draining, opening or closing of valves (including fire hydrants) and flushing operations. The majority of these procedures can be expected to occur within normally functioning operating WDS. However, some of the events highlighted by the authors are consequences of the following unforeseen actions:

- power failure causing pump trips,
- air release valves malfunctioning,
- pressure relief valves malfunctioning,
- air valve shut,
- check valve shut,

- break in a pipeline.

The above list, although not exhaustive, highlights the wide range of possible causes of pressure transients in WDS.

2.1.4 Severity of transients (single pipeline)

Large dynamic pressures may exist in a water network for only a short period of time; however, the shock loading effect can potentially damage vulnerable assets and cause pipe failures and deformations (Rachid and Mattos, 1994).

The earliest studies reporting damage due to the presence of extreme transients were related to hydraulic resonance in hydro-power plants (Wylie, 1965) and commissioning of pumping stations (Morris, 1967). Jaeger (1948) has presented a number of accidents due to pressure transients in power plants. His detailed studies of accidents and failures of penstocks revealed that unforeseen operational events are the main culprits of failure. This study presented cases describing pressure transients causing an air valve malfunction, followed by flooding and damage of pipes, dangerous vibrations due to the presence of air in the penstock and an occurrence of a dangerous crack due to a column separation effect. The author also highlighted a case where a dangerous vibration was caused by cavitation. Cavitation, which refers to the breaking of a liquid column, is related to the formation of void pockets of vapour when liquid boils or vaporises due to the significant drop in pressure. Numerous failures have been reported as a result of damaging pressures occurring when these cavities collapse. An extensive record of such failures can be found in Bergant et al. (2006). The authors present a comprehensive historical review of the pressure transients with column separation from late 19th century (when it was discovered) until the late 20th century.

Another study (Morris, 1967) has presented cases where damage, located in the close vicinity of pumps, was related to operation of check valves during pump shut down. Described was the rupture of a 12-inch cast iron main caused by a severe surge, which followed a valve being slammed shut. An eight-hour shut down was required to undertake costly repairs.

Kroon et al. (1984) focused on the adverse effects of transients induced by rapid collapse of void spaces during a pump start-up. This can happen when void spaces, such as vapour cavities or large bubbles of air or gas, exist downstream from the pump. The authors investigated a number of pump start-ups and spin-downs, which led to cavitation. The discharge side often experienced the pressure drop reaching vapour pressures, which may have a significant adverse impact on the whole system. Considerably large cavities occurring during rapid spin-down enable the re-joining columns of liquid to accelerate over this distance. Pressure upsurge observed on the suction site did not cause such significant issues. It was concluded that pump spin-down, resulting from

power failures, can perhaps account for the majority of pipe failures. Avoidance of water column separation has become a general rule in hydraulic design involving fluid transients (Bergant et al., 2006). Nevertheless, engineers and researchers became aware that severe pressure upsurges can cause deformation of pipe walls (Rachid and Mattos, 1994). Previously unexplained engineering problems of pipelines coming off their supports were explained by pressure transients induced pipe motion (Tijsseling and Heinsbroek, 1999). Unfortunately, the research effort was devoted mainly to pumps and large transients.

Identification of failures occurring in the vicinity of pumping stations, which have experienced pump trips, made water utilities and researchers aware of the threat of large transients in this vicinity. The high number of accidents soon became a reason for a subsequent detailed investigation of the transient phenomenon in laboratories. Focus was mainly on analysing pressure transient behaviour in a single pipeline (Rachid et al., 1994). These studies gave greater insights into the potential adverse effects of transients. Previously developed transient modelling techniques (Wylie and Streeter, 1978) have enabled a comparison of experimental results with computational ones. Due to the complex nature of transients, applications of transient algorithms were, however, only successful in a single pipeline set-ups. In WDS pressure transients can also create considerable dynamic pressures that exceed steady state condition. Therefore, if pressure transients are not fully considered in the pipeline design process, then the maximum design pipe loading capacity may not be sufficient to withstand them (Jung et al., 2007).

Greater understanding of pressure transients in a single pipeline enabled the development of numerous transient control and mitigation strategies: from changes in the distribution system, through to wave speed reduction and to installation of surge tanks and air chambers (Wylie and Streeter, 1993). Relief valves, by their ability to control pressure surges, became widely used for the protection of pumping stations. The relief valve's principal mechanism for surge protection is flow relief. Surge tanks have also become primarily used to control transients in WDS. These devices, by giving the flow an alternative storage volume during an upsurge event preventing the expansion of a pipe and the compression of the water column during a pressure transient. In downsuge events, surge tanks can provide water to the network.

2.1.5 Hydraulic transient analysis

The understanding of the severity of transient events resulted in consideration of pressure transients analysis in engineering applications. A comprehensive investigation of transient algorithms was not possible until the 1960s (Hager, 2001) when the advances in computational technology enabled a numerical analysis of the phenomenon. Researchers started to develop computational solutions

to fluid flows in pipes. In this period the most famous model was developed (Streeter and Wylie, 1967) in which a transient solution was undertaken using Eulerian based Method of Characteristic (MOC) and Lagrangian based Wave Characteristic Method (WCM), which was initially termed as the Wave Plan Method (Wood et al., 1966). Subsequent studies, such as one by Karney and Mclnnis (1992) successfully applied transient algorithms to a computational model of simplified small residential WDS. Both methods were rigorously compared by Ramalingam et al. (2009). In this study it has been shown that relative efficiency of the WCM is greater than MOC. For many researchers to date, the MOC (Edwards and Collins, 2013) or the WCM (Ferrante et al., 2009) is the most commonly used numerical solution for modelling pressure transients. In addition, the most of the available on the market software incorporate one of these methods.

Computational hydraulic transient analyses are mainly undertaken in the design stage of pipe systems and focus particularly on the protection of pumping stations. Researchers, who were aware of the impacts of the transient, exhorted to carry out water hammer analysis also for every major system design or operation change (Kroon et al., 1984). Most commercial software packages utilise the classic pressure transient theory, which allows for consideration of the design of surge suppression devices but leads to dangerously imprecise results, as highlighted by Ramos et al. (2004b) and Covas et al. (2005). A pressure transient analysis carried out in the operational stage is essential for diagnosis of malfunctions or for understanding the causes of a pipe burst, which was also highlighted by Soares et al. (2013). Unfortunately, much of the successful research into computational transient modelling was conducted in laboratories where simple pipe reservoir arrangements were investigated. Even though some researchers have in some cases presented more complex models for plastic pipes, such as HPPE and PVC, considering their viscoelastic behaviour (Ramos et al., 2004b) and mechanical damping effects (Stephens et al., 2011), there have only been limited attempts to create more complex models by the application of transient algorithms to a full scale distribution system (Ebacher et al., 2011). Other researchers (Kwon and Lee, 2011) have used a reliability analysis based on load and resistance function, to estimate the probability of pipe failure resulting from pressure transient flow in WDS. The model presented by Kwon and Lee (2011) showed that simulated unsteady flow increased the probability of pipe failure significantly. This research used a small modelled network where transients were generated by sudden changes in demand.

The phenomenon of pressure transient and its possible impact on WDS has not been widely understood. This was evidenced by the number of pipe breaks caused by the existence of pressure transients (Kroon et al., 1984)

The complexity of WDS complicates the analysis of transient behaviour. The reflections from

junctions and loops of a system give the potential for superposition of two or more waves. This may be constructive or destructive, depending on the phase of the travelling waves (Gartenhaus, 1977).

Karney and McInnis (1990) noted that many incorrect simplifications and assumptions are commonly made to manage the analysis more efficiently. This seems to be a persistent challenge to these days as researchers (Jung et al., 2007) still find this to be a current and unsolved issue.

2.1.6 Transient data acquisition

Pressure transients are phenomena that are not fully understood, especially when the complexities of WDS are taken into consideration. One of the reasons is the short, but potentially extreme nature of a pressure transient, which may only last a few seconds.

Historically pressures in water pipes were recorded using pen and charts. Early field research into transients often used pressure-transmitting devices, where information was logged on strip charts. When computers become available the digital storage processes replaced this traditional method. In the study by Friedman et al. (2005) the two methods were shown to be comparable. Investigation of low and negative pressures in live WDS showed that the pen charts tended to record higher values than digital data loggers. Today the use of digital pressure data loggers is established. Prior to the advent of computers in the early 90s, little was known about transient behaviour in complex networks. This was mainly due to the lack of confirmatory studies, which could verify computational simulations.

Current collection of pressure data from water distribution systems in the UK is usually undertaken at a 15 minute resolution to satisfy regulation ((Ofwat, 2008), see section 2.1.1). This data is usually manually collected at monthly intervals (Savic and Banyard, 2011). 15 minute conventional pressure monitoring data is unsuitable as this recording frequency is insufficient to capture travelling pressure transient waves. This practice gives insights into steady state system conditions and only allows for the capturing of changes which occur over periods of minutes and hours. Current understanding of transients indicates that only high frequency and preferably continuous recording can reveal the existence and characteristics of transients in WDS.

In places where power is available e.g., pumping stations and service reservoirs, data is typically collected through supervisory control and data acquisition (SCADA) on-line systems. From this arrangement water utilities can record some pressures at relatively low frequencies (Friedman et al., 2005). Various online communication technologies have been developed for data collection and transfer at low power (Stoianov et al., 2006). The continuously recorded time series pressure data is analysed by examination of threshold levels and these are set by regulatory standards and

site-specific operational experience (Savic and Banyard, 2011).

According to the latest research on high frequency data, which evidences the occurrence of transients in operating water distribution systems, high frequency monitoring may be essential for characterising and quantifying pressure transients and for evaluating their associated impacts. Continuous monitoring using a wireless sensor network was successfully used in field by Stoianov and Nachman (2007). This work was used primarily for real time burst detection and location in a simple network. This idea was extended by Srirangarajan et al. (2010) in trials in a live distribution systems. To acquire shape and magnitude of pressure transients a frequency of not less than 10 Hz was identified (Savic and Banyard, 2011).

Some field studies have used data loggers at a frequency of 20 Hz (Besner et al., 2010; Friedman et al., 2005; Gullick et al., 2004). The work of Kirmeyer et al. (2001) investigated a number of sites where pressure was measured at 1 Hz and 20 Hz. Pressure transients were monitored from two to 43 days at various locations, but higher recording frequencies were only used for a short period of time, when specific field activities took place. Lower recording rates (1 Hz) allowed pressure transients to be captured during longer duration recordings. Unfortunately, the pressure loggers used in this study were not time synchronised and pressure traces were not monitored simultaneously. This made it difficult for the pressure traces from different locations in the same system to be compared and analysed. The successful use of 1 Hz and 20 Hz recording rates was applied by Gullick et al. (2004). In this study low recording frequencies allowed longer logging periods, whereas higher frequencies provided shorter, but more precise data. 1 Hz frequency recording enabled weeks of data to be stored. For short (e.g., hours) recording periods a 20 Hz frequency was used. Work by McInnis and Karney (1995) used recording frequency of 32 Hz. Available memory space, however only allowed a continuous recording of pressures for 8.3 min. Creasey and Garrow (2011) used 10 Hz recording frequency to detect pressure changes over a nine-month recording period. Pressures were, however, not recorded continuously but only when their values fell low enough to trigger loggers. Data from each logger was collected monthly. In the study by Besner et al. (2010) pressures were recorded continuously for a few months from hydrants at 1-4 Hz.

There is little research employing high speed data logging to make a general assessment of all transients (in terms of their magnitudes and frequencies) in complex WDS. Popular use of lower recording rates at 1 Hz (Kirmeyer et al., 2001) for longer pressure monitoring were shown to be sufficient to capture high and low transients events. The low frequency of the continuous recording were not necessary a preferred choice but rather a practical to optimise power consumption for longer duration monitoring and data storage.

More recent studies have shown the ability to record field data at high rates of 100 Hz for short durations (Hampson et al., 2013), 300 Hz (Stoianov and Nachman, 2007), 500 Hz (Misiunas et al., 2005; Stoianov and Nachman, 2007) (for 5 sec every 5 min) and 2kHz (for an initial phase, then reduced to 250 Hz) for burst detection using wavelet analysis continuously recorded for about 2 hrs (Srirangarajan et al., 2012). Such high frequencies have been successively applied but only for a short recording periods. High frequencies in these research were mainly required by analysis method investigated (e.g., wavelets).

Despite these challenges, modern technology has enabled researchers to firstly gain insights and secondly to use transient waves to investigate various engineering issues. Subsequent studies have used the large amount of information that the propagation of transient waves through the system provides. This has resulted in the application of transient based methods for burst detection (Misiunas et al., 2004), leak detection (Guo et al., 2012) and also location of transient sources (Hampson et al., 2013).

2.1.7 Field studies into transient and instrument location

Walski and Lutes (1994) focused on confirming the existence of transients and then finding their causes. This study conducted a larger scale monitoring (covering one city). Located in a real network in Austin (Texas), portable pressure monitoring devices enabled the researchers to find the source of low pressure events. The culprit was the change in velocity due to pump start and shut-down. Field tests also confirmed that numerous low pressure events, which were experienced in the city, were due to transients.

One field study which used digital pressure recording devices was by McInnis and Karney (1995). This investigated the existence of transients in major transmission and distribution mains in the city of Calgary (Canada). The study was based on a series of engineered pump trip events. Pressure transients were measured by two high speed pressure-recording devices synchronised with the computer time. The comparison of field recordings with calculated results showed that a larger energy dissipation of transient waves was observed in the real system. The authors therefore suggested that one of the reasons for the lack of field data was the complex nature of WDS that contributed to the damping of transients.

Subsequent research into transients in live WDS focused on documenting the existence of low and especially negative pressures (Karim et al., 2000, 2003; Gullick et al., 2004). The attempts to characterise the occurrence of pressure transients in WDS were undertaken by monitoring more sites and for increased durations. These studies have confirmed that the most of the recorded low and negative transients are associated with pump operations. The work of Kirmeyer et al. (2001)

focused on monitoring two live water distribution systems. In the first system, ten loggers were selected for monitoring of pressure. These were monitored individually, using one pressure logger per site. Mainly sites with a history of low pressure events, spiking pressures, high pressures and sites without previous pressure problems were monitored. Two sites had experienced an increase in pipe failures and repairs. The remaining two sites were chosen to monitor distribution and transmission mains. The second distribution system was monitored at four locations for a short time, mainly to investigate transients caused by rapid hydrant opening. The results from the monitoring showed that no negative pressure waves were detected at the time of monitoring, but low pressures were seen at three sites.

The existence of low and negative pressures was also investigated in UK distribution systems (Creasey and Garrow, 2011). In this study three locations were monitored with four or five loggers placed at increasing distances into the WDS. At the first site low pressures recorded were associated with the repair of large bursts. At the second location, a large industrial user was suspected to generate large flows and therefore the potential of transient events existed. During a two-week recording period, no low pressures were detected, however underlying 40-second oscillations were noticed. These produced 20 m pressure head changes. At the third location 5 loggers were installed over a five-month period. A large transient recorded at this site was associated with a pump trip. This event caused a 90 m pressure drop. Other low transients were associated with flushing events. The authors concluded that the overall risk of the existence of very low pressures was minimal.

Gullick et al. (2004) extensively monitored eight distribution networks to determine the number and locations of low and negative pressures in WDS. Fifteen low and negative pressure events were detected, twelve of which resulted from pump shut downs. Two were due to sudden increases in demand and/or booster pump operation. One more event was detected; the cause was not determined.

Negative transient pressures have been reported in Canadian WDS (Besner et al., 2010). Over the first four months of a recording period, eleven negative pressure events were recorded at twelve sites. Four were the result of the repair to a transmission main. Three were detected in sections which were isolated for repair work. The remaining four showed transient events caused by power failure at the treatment plant. A subsequent second monitoring, carried out over 13 months, resulted in seven negative pressure events, but not all of these events could be explained.

Other methods have been developed for optimal sensor location in WDS to detect bursts or leaks and improve on the current water industry practice of placing pressure sensors at a critical point inside DMAs (Farley et al., 2008, 2013). This was followed by other sensor placement

methodologies using artificial intelligence (Mounce et al., 2010). The method based on a graph theory focused on transient source detection also investigated the optimum sensor placement (Hampson et al., 2013). The optimum placing of sensors within complex WDS is application dependent. To capture pressure transient events, the most studies used a single measuring location. For more complex analysis (e.g., burst, leak detection) an increased number of instruments is required, e.g., a minimum of three for simple and relatively small networks Srirangarajan et al. (2010).

2.1.8 Summary

It has been shown that large pressure transients can damage water infrastructure. The impact of small to medium pressure transients, however, has never been investigated or evaluated so remains unknown.

The literature review showed the collection of large scale high speed pressure data is necessary as currently available data (e.g., 15 min resolution or only low and negative events) is insufficient to reveal a) the existence of pressure transients and b) which information about their possible damaging effect should be analysed.

Field based methods historically were limited by available power and storage capacity due to the large volume of data collected using high frequency continuous monitoring. For instance, a sample frequency of 20 Hz is sufficient for visual identification of transient events providing a reasonable representation for the profile of a transient wave, but it may not be adequate to gain sufficient information for subsequent analysis. Other studies for pressure transient source location required higher frequency recording. For instance, pressure transients source location in real WDS was possible using a frequency of 100 Hz.

While a large number of potential causes of transients are known, field measurements recording the system pressures associated with these causes have not been sufficiently documented. Commonly recognised sources were pump operations and industrial users. Many studies highlighted that although pressure transients were observed, their direct causes remained unidentified. This review has found that no existing studies have undertaken long-term monitoring of large sections of distribution systems at a high recording frequency that allows the capture and characterisation of the wide variety of pressure transients.

2.2 Signal analysis

2.2.1 Noise

Pressure monitoring instrumentation provides instantaneous readings which can be subject to noise and may not be representative for the actual pressure signal, see Figure 2.3. The reduction of noise is usually undertaken to improve quality of raw data and subsequent analysis.

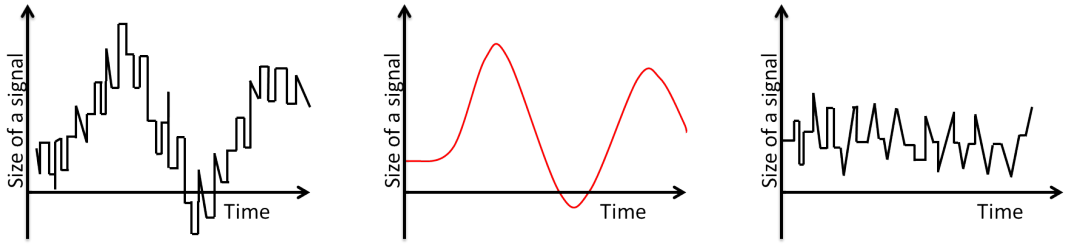


Figure 2.3: The characteristic of the dynamic pressure data obtained from pressure data loggers. (a) Real signal from a measuring device (with noise). (b) Ideal signal. (c) Noise.

The source of noise is due to the physical measurement from the instrument and also other operational events. Previous studies have preprocessed the raw pressure signal to remove noise (Srirangarajan et al., 2012). The study used de-noised data to detect simulated burst event in a real water network.

The individual measurements of the pressure from the device may not agree well with each other. In this case the precision of the measurement is addressed. The precision can be indicated by the width of the data distribution, which describes how much variation occurs between successive measurements. Precision is also known as a measurement of random noise (Smith, 1997) or random error.

In discrete time the 'white noise' is represented by discrete signal samples, which are considered to be uncorrelated and have the same probability distribution (often called independent and identically distributed random variables). A special case is when the signal has a normal distribution (zero mean and constant variance) and it is known as Gaussian white noise. If these random variations are a representation of 'white noise', the signal should have equal power i.e. flat/constant power spectral density over the range of frequencies. A power spectral density (or energy spectral density) is a distribution of signal power into frequency components. Strong and weak variations in a signal may be then identified by their contribution to total power.

2.2.2 Representative length of the data (stationarity of the data)

Stationary processes - arise when the random mechanism producing the process does not change with time (Grenander and Rosenblatt, 1984). According to the author these are often happening in technology and in the physical sciences. In other fields (e.g. economics), however, it is often assumed to hold if period of time is not too large and if the systematic component is isolated in an appropriate way. Box et al. (1994) say that stationary time series can be described by its mean, variance and autocorrelation function, or otherwise by its mean, variance, and spectral density function.

The stationary process is a special class of stochastic process based on the assumption that the process is in particular state of statistical equilibrium (Box et al., 1994). There are two types of stationarity. First one is called a strict (or strong) stationarity, the second one weak or stationarity of order 2.

Strict stationarity is the example of the strongest form of stationarity. A process is stationary when its properties do not change (are unaffected) for any period of time. That is, if the joint probability distribution of any set of observations is the same for all times, i.e., is unaffected by shifting all the times of observation forward or backward by any time period. That means that stationary time series has statistical properties, which are constant over time. Therefore, the stochastic process has a constant mean which defines the level about which it changes, and a constant variance.

If the data were sampled for too short time length the variability in the sampled means and variances are expecting to vary (scatter). When the sampling time length is increasing the degree of deviations between the sampled means and variances should converge to a constant value indicating achievement of stationarity. At this point the average value of estimated means and variances should be the same around a constant horizontal line.

2.2.3 Representations of time series data

Current methodologies for assessing pipe repairs are based mainly on a multivariate approach. This method investigates effects of all variables on the problem studied, where each variable is represented by a single value (Kennedy and Turley, 201). Characterising transients requires a time-series approach. To employ that data format in a model, analyses of the time signal are necessary, which may return potentially useful information, such as trends and seasonality (Lewicki and Hill, 2006). These methodologies mainly include representing time-series features as latent variables (not directly measured), so that dynamic properties can be expressed in a ‘static form’, such as the form of a single variable appropriate for modelling (e.g., mean, median and

mode). That encoded variable is consistent with multivariate analysis and can be incorporated into the model. The 'static' representation of time-series data should consider the representation of features which lead to the event, such as number of significant transient events during specified time intervals or number of loading cycles for small transient events. These could be done by applying a rainflow algorithm, developed by Matsuishi and Endo (1968) for random loading. This algorithm reduces a spectrum of changing stress into stress reversals. It is currently unknown whether it is large pressue transient events that affect water networks and cause pipe failures, or if long term fatigue loading of repeatedly occurring pressure transients induces a greater impact and risk.

2.2.4 Detection of peaks and troughs

Peak-trough detection techniques have been successfully developed and applied to characterise features in complex time series data in other fields (Du et al., 2006; Coombes et al., 2003). These techniques have enabled researchers to count significant events in time series data and are generally fast and easy to implement on large datasets. Such peak-trough detection algorithms have not been applied to pressure transient time series data. However, direct unconditional application of these techniques will not provide the sufficient characterisation of all magnitudes and rate of occurrences of transient events observed in operational WDS. This is because high magnitude pressure transient waves frequently comprise of smaller waves resulting, for instance, from reflections at junctions.

There are numerous methods for peaks and troughs detection. The use of each depends on the application.

(a) zero-crossing – the common approach for peak detection is the use of differences in the change of the slope sign (from a positive to a negative or vice versa). When the zero-crossing is detected the local maxima and minima can be found. This method, however, detects all peaks and troughs, whether strong/significant or not. A common approach require the noise reduction process before peak/trough algorithm to be applied.

(b) wavelet analyses are capable to deal with the fact that pressure transient waves are series of discontinuous events. Wavelet analysis has been successfully applied in transient analysis for peak detection (Du et al., 2006), leak detection (Ferrante et al., 2007) and burst detection (Srirangarajan et al., 2012) and also in other fields for singularity detection in electrocardiogram signals (Jalil et al., 2010). Wavelet analysis has proven its use as a tool for the detection of signal discontinuities and for denoising a signal in a number of contexts (Liu et al.,

2003; Placidi et al., 2003; Olhede and Walden, 2003). However, the method require a high recording rate and was successively applied for short recording periods.

(c) peak and troughs detection algorithm – based on the algorithm (MATLAB script: `http://billauer.co.il/peakdet.html`). The algorithm detects peaks (finds the local maxima and minima) in a vector. A point is considered a maximum peak if it has the maximum value, and was preceded (to the left) by a value lower by delta.

2.2.5 Correlation measures

To assess the discrete pressure transient events against orginal time series signal various methods can be used. One method to assess it is a sum of squares of residuals ($S_r{}^2$). A small value of $S_r{}^2$ indicates a good fit of the model to the data. Young's Coefficient of Determination (R^2) (Young et al., 1980) is another measure. For this measure high value indicates a good fit to the data. The Mean Squared Error (MSE) together with the Mean Absolute Error (MAE) are the major measures of performance of forecasting models (Pearn and Laboratories, 1991; Harmer et al., 2008). Disadvantage of using the MSE can be seen in the tendency to heavily weight the outliers. This is a result of squaring each of them, which gives heavier weights on large errors in comparison to the small ones. This can be important to consider as for some applications this effect was undesired and has led to the use of alternative error measures such as MAE. Both methods measure a vertical distance rather than a horizontal one.

The Root Mean Square Error (RMSE) is a scale dependant measure. Therefore, it is not suitable in comparing across data sets that have different scales. RMSE is one of the most popular scale dependant measures currently used when dealing with time series data (Hyndman and Koehler, 2006). Some researchers noted that the use of RMSE is related to its popularity within statisticians (Ahlburg, 1992) and it is also frequently used in electronics (Smith, 1997). RMSE is often preferred because it is in the same scale as data, therefore, allows comparing different methods applied to the same set of data. However, for comparing methods across different data sets Normalized Root Mean Square Error (NRMSE) is preferred.

2.2.6 Summary

Various techniques to evaluate continuously recorded time series signal have been investigated. This included noise reduction followed by different techniques to identify discrete events from a time series data through correlations measures.

2.3 Pipe failures

Pipe failures may have an economic consequences with costs related to the actual repair and other associated charges. These are often accompanied by significant social (e.g. supply interruptions, damage to the properties, traffic) and environmental (e.g. water and energy loses).

2.3.1 Current state of water distribution systems

Water distribution systems are currently facing the challenges of maintenance, replacement and upgrading. The main concern and area of focus is the number and severity of pipe failures and leakages, which increases as the water infrastructure ages; as 65% of current UK water infrastructure was constructed in the 1960s (Savic and Banyard, 2011). The repair and prevention programmes, together with optimisation of available financial resources, require more rigorous studies and research into the causes of pipe failures, in order to minimise costs.

As water infrastructure has undergone deterioration, researchers have been able to analyse different factors and reveal more information about pipe failure causes. These factors have provided a basis for determining sufficient rehabilitation and maintenance strategies, to meet current and future demand requirements. Traditional approaches initially have focused on investigations of the historical pipe failure data. The repair rates were then analysed and insights were revealed into the historical failure patterns. Failure statistics were implemented together with recommendations and methodologies for the improvement of water infrastructure performance. The criteria for deciding whether pipes should be replaced or not were outlined by Stacha (1978) and became commonly used. The criteria for replacing pipes were not only targeting decisions on pipe replacements (O'Day, 1983) but also a rehabilitation planning (Sargaonkar et al., 2013). In addition, it was concluded (Clark et al., 1982) that once a pipe fails and requires maintenance, future maintenance requirements increase exponentially and its maintenance costs rapidly outreach a cost of its replacement.

2.3.2 Pipe failure factors

Selection of pipe failure factors depends mainly on data availability and it is usually a complex task. Some research (Constantine et al., 1996) listed factors, which according to them affected pipe failure rates:

- Environmental conditions (i.e., soil type and overhead traffic),
- Operating condition (i.e., water pressure),
- Asset features (i.e., material and pipe size).

Material: Current water distribution systems have been built from a variety of materials requiring different construction techniques. In the past, the majority of water infrastructures were constructed using cast iron and asbestos-cement materials. For these reasons, the majority of studies in the existing literature were based on data obtained from ferrous pipes (cast iron, ductile iron, etc.) (Kettler and Goulter, 1985; Kleiner and Rajani, 2002) as these have been in use for longer.

It was found that, for cast iron pipe materials, overall failure rates increase with time and decrease with increasing diameter (Kettler and Goulter, 1985). Cast iron pipes are often prone to graphitisation, which weakens the pipe wall. However, even weakened pipe often has enough material strength to allow it to function and sustain external and internal loadings. In addition, in the study commenced by O'Day (1983), the method of manufacture of cast-iron pipes was found to influence pipe tendency to fail. The later-introduced steel pipes have many advantages, mainly in high tensile strength and high compressive strength. An important disadvantage of using steel pipes is their propensity to external corrosion; consequently internal and external protections are required.

Current installation process covers a variety of plastic pipes such as Unplasticized Polyvinyl Chloride (UPVC), Polypropylene (PP), Polyethylene (PE) also known as LDPE, MDPE, HDPE (low, medium and high density) but structural problems have been encountered even with these types of materials. Consequently, researchers have commenced some analyses of failures in plastic pipes. For instance, Davis et al. (2007) have developed a physical model to estimate failure rates in buried PVC pipes as they age, using the UKWIR National Mains Failure Database from 17 UK water utilities. In this mode, average failure rates were estimated, which were then compared with actual failure rates. This comparison showed good agreement (95% confidence limit) between both rates. However, lack of actual operating pressures from the UK water utilities represents the substantial limitation of this approach. Further research to obtain these values is needed to fully validate the model.

Material lining: According to the researchers ductile pipes are mainly lined internally with cement mortar and externally coated or covered by plastic sheaths to resist external corrosion. Over the years cast iron pipes have been exposed to different quality of water which may have caused its deterioration due to biological, mechanical and chemical processes. Deterioration of external and internal pipe material has two forms: structural and surface. Seica et al. (2004) evaluated the strength of exhumed cast iron pipes confirming that corrosion and air pockets in the pipe material exist and contribute to reducing a pipes structural integrity by acting as stress concentrators and crack instigators. Large variations in pipe strength made it difficult to predict

failure rates based on pipe material.

Spray lining such as cement mortar of 4 mm thickness was historically preferred material lining. This lining was often affected by water quality and chemistry causing lime in mortar bleaching away leaving the sand residue (Savic and Banyard, 2011). Semi structural lining - epoxy or polyurethane (PU) lining has been therefore introduced as a supplement. This type of lining could be sprayed in 1 mm thick layer. Other structural lining materials can have a thicker layer and therefore reinforce the structural property of the pipe.

Age: The majority of early historical studies considered age as the most relevant variable explaining pipe failures (Shamir and Howard, 1979; Walski and Pelliccia, 1982) and some relationships were found between age and pipe failures. These studies used regression analysis where pipe failure data was usually plotted against time, and regression was used to develop an equation which gives the number of failures as a function of time (Constantine et al., 1996; Shamir and Howard, 1979; Walski and Pelliccia, 1982; Walski et al., 1986). These studies developed an equation describing the increase in pipe failure occurrences over time for a particular pipe or for the entire network. For instance, the exponential relation to age was reported by Shamir and Howard (1979), and Walski and Pelliccia (1982). Powered relation between pipe failures and age has been reported by Kleiner and Rajani (2001). These two relationships have been used to establish the best time for a pipe replacement. Some insights into how age parameter may affect pipe failures can be found in life time models (Andreou et al., 1987). This research has reported that the failure rate will be constant after the third burst. This seems to be unrealistic finding which may be due to some artefacts in their analysis. This is also in contrast with some regression models (Shamir and Howard, 1979) suggesting that failure rate may increase with time according to an exponential or power law.

Some authors made a distinction between pipes with and without burst history (Andreou et al., 1987). It has been found that it agrees with some statistical findings (Goulter and Kazemi, 1988) and explanation of life cycle of the pipes refereed to as a 'bathtub' curve (Andreou et al., 1987; Kleiner and Rajani, 2001). Particularly in plastic pipes, pipe failures can be described by a bath tube curve. This is harder to observe in cast iron pipes where there is usually not enough data to detect the first, steep bend of the 'bathtub' curve.

According to other researchers age alone was not concluded to be a significant factor when analysing pipe failures (O'Day, 1983). This opinion was consistent with Walski and Pelliccia (1982); Clark et al. (1982); Kettler and Goulter (1985); Goulter and Kazemi (1988) which concluded that age should not be the only variable assessing pipe condition and explaining pipe failures. Also Lei and Saegrov (1998) have claimed that age of pipes do not play an important role when

assessing their remaining lifetime, and Røstum (2000) that age cannot be used as the only measure determining a pipe replacement. In the study conducted by Andreou et al. (1987) it has been shown that pipes which have a tendency to fail at lower ages performed better than pipes which experienced failures at higher ages. However, a historical research by Dyachkov (1994) reported that some pipes which had exceeded their service period were in satisfactory condition and caused no problem in operation.

Age can be considered as an indicator of a length of time a pipe was exposed to different factors (pressure, soil loading, or corrosion, both internally and externally). Time makes a significant contribution to the detrimental processes such as corrosion, fouling, soil subsidence. Age as a factor may also indicate the construction processes at specific time (e.g., short life time of pipes which had been laid at the end of the Second World War, due to shortage of cast iron material which had been used to produce ammunition). Age is difficult to separate from other influencing factors. In addition, some time dependent factors (such as cyclical, recurring environmental factors or weather conditions) may mask the effect of underlying ageing patterns. This is usually a case when limited data is available. Considering all the above, age of the pipe may be an effective indicator of the state of a water network. Some studies characterised age as 'useful life' which is a duration after which it is no longer fit for purpose.

Age is often associated with pipe deterioration process over time and some studies have shown constant increase of pipe failures over time due to deterioration of the pipe (Kleiner and Rajani, 2001). Therefore, in pipe failure analysis, age is usually used in combination with other factors.

Diameter: Early studies on pipe failures identified diameter to be important in the analysis of pipe failures (Walski and Pelliccia, 1982; Clark et al., 1982). Some researchers (Kettler and Goulter, 1985) reported a strong inverse relation of diameter with pipe failures. This study, commenced in Winnipeg over a 10-year period, has shown strong negative linear correlation with diameter. This led to the conclusion that small pipes are more prone to failures then larger ones. Other studies (Boxall et al., 2007; Yamijala et al., 2009) have also reported that higher failure rates were typical in smaller diameter pipes and shown that burst rate decreases when pipe diameter increases. The high frequencies of failures for small diameter pipes indicated low beam strength of these pipes when exposed to loads. The pipe ability to resist the forces is measured by second moment of area. Therefore, the pipe with decreasing diameter shows decrease in the second moment of area (Skipworth et al., 2002). Large diameter pipes usually have thicker walls which enable them withstand high circumferential loads. These pipes, however (if made of metal) are subject to corrosion or other degradations which may weaken their structural strength. It has been found (Walski et al., 1986) that smaller pipes were prone to circumferential failures whereas large

transmission pipes made of steel suffered mainly from corrosion holes.

The accurate prediction of pipe failures is challenging as the change in diameters is uncertain (as the uncertain are blockages in the pipeline). It has been shown (Bergant et al., 2008b) that in some cases the presence of unknown leaks and blockages may cause additional damping.

Pipe length: Some authors have included length as an important additional explanatory variable (Le Gat and Eisenbeis, 2000) while explaining pipe failures. Pipe lengths are not uniform and can significantly vary within a network. Longer pipes are representative for rural areas, where transmission mains of uniform characteristics tend to be used. Longer pipes may be more representative in rural areas with relatively low variability in the surrounding environment.

In terms of the density of connections, shorter pipes can have more tendencies to fail than longer ones This behaviour was attributed to greater number of joints and fittings per distance observed in areas of high population density and led to a conclusion that high connectivity in high density urban areas with short pipe lengths would have a higher burst rate (Boxall et al., 2005). Higher burst rates were also associated with variability in surrounding environment. If connections are considered as a point of weakness, then it can be perceived that shorter pipes experience higher burst rates than longer pipes. In addition, external impact (soil conditions, traffic) may differ along long pipes, which can influence their burst rates. The relationship of bursts to the length of the pipe may be expected to be linear, as the longer the pipe the more of it was exposed to potentially adverse conditions. However, this relation has not often been observed.

The majority of studies include length of pipes in the failure data as rate. Walski et al. (1986) suggested that to make any comparisons meaningful a failure data should be presented in units of number of pipe repairs per unit length and also per unit time. This will give quantity which is referred to as pipe repair rate.

Static pressure: steady state pressure was identified as important factor in pipe failures (Brandon, 1984; Skipworth et al., 2002). It was found that reduction in static pressure help to reduce water loss through leaks (Lambert, 1998), see section 2.1.1. Majority pipes in water distribution systems are exposed to varying static pressures during a typical day, seasonal variation or even due to changes in a system performance. The relationship between these gradual changes and pipe failures are not well understood.

Some research has investigated the total stress in a pipe as a combination of axial stress and hoop stress which are in turn combination of external loads, internal pressures, temperature differential and longitudinal bending (Tesfamariam and Rajani, 2007). Longitudinal failures are recorded to be predominantly a result of internal pressures (Kleiner and Rajani, 2001). Steady state pressures are important factor considered during a pipe design. Increasing demand can be

one of factors acting as internal load on a pipe.

Soil: Historically, old pipes were buried directly in soils. Later, these were laid typically in sand. Plastic pipes (introduced later) have been generally put into bedding rather than directly into soils. It can therefore be assumed that soil may not have a direct affect on plastic pipes. However, soil data may be valuable factor when assessing more brittle materials and may be even more important with increased age of pipes.

Soil data can be expressed in various forms. Constantine et al. (1996) has found that soil type is the most important variable in explaining variation in failure numbers from asset to asset. The soil classification mentioned in this research was: sand, clay, loam etc. This study showed that the soil type influences corrosion in a complex manner, which is still not fully understood. Morris (1967) has included soil as one of the important factors causing pipe failures. He found it so valued that he recommended that soil surveys should be carried out before the network design process.

Soil impact is related to decrease in volume when soil is dry and increase when moist, related to weather changes. The relationship of soils with weather conditions during a year has been widely reported. Baracos and Hurst (1955) have reported that water pipe failures in Winnipeg were observed from September to January. According to Liu et al. (2012) increased soil movement and increased loads on pipes occur when soil freezes. This is to be expected especially if soil is highly permeable. This finding was with an agreement with previous study, which reported higher pipe repairs in winter months due to frost penetration (Walski et al., 1986).

Some studies noticed that pipe failures appeared after summer when dry soil conditions were common or prior to spring thaw. Hudak et al. (1998) have shown that in Texas pipe failures were observed during extreme dry periods in expansive soils.

Especially clay soils of small particle size and platy shape allow water particles easily to be attached. These soils shrink and swell as a moisture level changes. Pipes laid in such soils are subjected to deformation imposed by ground movement. Kleiner and Rajani (2000) noticed that swelling and shrinkage of soils can increase vertical load on buried pipes and seasonal swelling and shrinking has been identified by many research as one of the causes of pipe failures. This can be especially important in the UK, where clay soils are widespread. Soil fracturing is a way to describe this natural ground movement due to seasonal shrink and swell of soils. Soil fracturing can also describe possible adverse effect of ground movement potential if presented in a scale from very low to very high.

Other terms to convey soil impact would be through its workability when soil can be physically manipulated for cultivation purposes. However, workability of soil does not represent its precise condition as it highly depends on a human factor, e.g., the operator and available machinery.

Soil workability (sometimes called trafficability) is directly related to the condition of land for cultivation. It is a measure of the optimum water content at which the greatest proportion of small soil clusters are produced by working the land. Clay soils are especially difficult to work as the soil clumps and it is hard to separate. On the contrary, sandy soils are easier to work. When soil is wet then it gets cohesive and forms wetter, large clods, with poor trafficability. However, when soil is dry it can produce dry large clods (soil gets compacted and it is more difficult to plough). High soil workability allows agricultural cultivation without causing detriment of soil structure by compaction or smearing (Cooper et al., 1997).

Clay soils are considered as being corrosive to pipes (Rajani and Zhan, 1996). Soil Corrosivity (relative corrosivity), was established based on five major classes: soil acidity, sulphide content and aeration and wetness. The assessment of soil corrosion was previously undertaken by corrosion status index (Kumar et al., 1984). Soil corrosivity and soil fracture potential are two main factors gathering effect of corrosion and ground movement on pipes (Cooper et al., 2000). Corrosive soils are often culprits of pipe external corrosion. Soil movement due to expansion or shrinkage impose further external loads. However, the investigation of soil data in some studies (Boxall et al., 2007) showed lack of consistent dependence with burst rate.

Temperature: Temperature as a factor may have an impact on pipes. Due to reduction in temperature the volume of water expands, frost heave in the soil which surrounds the pipe, and increased brittleness. The development of frost heave is usually influenced by rate of freezing, surface cover, mains depth intensitivity and duration of cold water. Generally, there is a lack of strong correlation between main failures and temperature. Dehghan et al. (2008) showed that random water pipe failures should be considered as non-stationary processes as they can be influenced by seasonal variations of temperature.

Fatigue: Studies into mechanical behaviour of materials has revealed that damage mechanisms can occur when metals are subjected to repeatedly applied stress/loading: periodic or variable (Lemaitre, 1984). This kind of damage was categorized as fatigue to distinguish it from other types of failure. This study has shown that inelastic deformation of metals can be one of the mechanisms responsible for initiation and growth of microcracks and cavities. This damage mechanism was later applied to the water industry to evaluate the integrity of its structural components. Usually in the fatigue analysis load (magnitude) is the most important factor, however, how long this load is applied is also important (D'Souza, 1966) as the stress factor alone gives little indication of safety of a material. For some materials when small load is applied then longer time is needed for the material to deform (e.g., weeks, months). It is usually calculated as strain rate (D'Souza, 1966). Pugsley (1966) reported that for aeroplanes peak loads are calculated per hour per flight for safety.

Safety load calculations were also applied for other structures, e.g. bridges, tall buildings, to give maximum allowable strain rate (D'Souza, 1966).

'Rainflow counting'(cycle identification techniques) – is a preferred method in fatigue (cumulative damage analysis). Pre-processing of the loading signal is required before applying this method, to extract maximum and minimum peaks and put them in classes. This requires peak- trough detection by level-crossing, peak counting, range-pair counting. The level-crossing counting is typically counted each time when the positive slope of the load are registered at levels above the mean. There is no difference if the positive or negative slope crossings are counted. Counts at each level are cumulative. The levels are constructed by counting the first possible largest cycle, followed by the second largest. Peak counting method counts peaks above and below the relevance levels (reference level is usually a mean) and reports them separately. The most common method is to first count the largest possible cycle for maximum and minimum peaks, followed by the second largest cycle. In the simple-range counting method each range (positive and negative) are counted as one-half cycle. The rainflow counting method has been used as a standard method in fatigue analysis for the railway, aircraft and automotive industries.

Through stress-strain (S-N) curves the fatigue life is calculated by Palmgren-Miner rule. This has allowed researchers to hypothesise that pressure peaks due to transient events can cause the deformation and rupturing of pipe walls (Rachid et al., 1994). The authors performed pipe lifetime computational simulations in which an additional variable, representing the damageable elasto-viscoplastic behaviour of a single metal pipe, was included. Following a number of simulations, it was shown that if the pipe damage value approaches the critical value, rupturing can occur even with a slight increase in the amplitude of pressure. This increase could happen, for instance, due to the superposition of the signal. Shortly after this study, it was pointed out that, in reality, pressure transients comprise several peaks, which could also be sufficient to damage a pipe (Rachid and Mattos, 1998). Therefore, the cyclic nature of transients should be taken into account to actually describe the presence of transients and their effects on the pipe. Attention was given to pipe material degradation as induced by inelastic strains. This resulted in the mechanical model describing pipe behaviour. The loading, subjected to a stainless steel pipe, was assumed to be due to changes in the internal pressure, generated by a rapid valve closure. The critical damage value, incorporated in the model, depended on the pressure history and hardening law. The numerical model presented by the authors estimated the accumulation of the damage and the predicted lifetimes of pipes that are subjected to pressure transients. In comparison to plastic pipes, metal pipes are more subject to fatigue failures whereas plastic pipes can fail at the joints where leaks tend to occur. These early studies enabled researchers to link transient cumulative dynamic

impact with the structural integrity of a pipe, but were not validated by experimental data.

Many studies have investigated fatigue strength or the cyclic life of materials (Little and Jebe, 1975). These utilised the basic material data, which may be obtained by testing specimens in the laboratory, either under constant load (stress) or constant strain (displacement). The resistance of pipes to such impact depends on the tensile strength of the material and pipe wall thickness (Skipworth et al., 2002). Fatigue behaviour, however, depends on many factors and it is not possible to include all of them during laboratory experiments. In the laboratory, data are obtained using plain test specimens manufactured to a high degree of accuracy, which are then subjected to load patterns of constant amplitude in a non-hostile environment. It needs to be clear that these conditions are substantially different from a real environment (Little and Jebe, 1975).

The existence of smaller magnitude transient events have not been sufficiently documented and analysed as forces that could potentially damage pipe material. There is the potential that shock loading due to such transient events may also have a significant impact on WDS, leading to pipe failure as a result of fatigue. From this the cyclic nature of transients should be taken into account to actually describe their behaviour in a pipe (Brevis et al., 2014). Such data is currently not available. None of the reported studies have investigated fatigue effect caused by pressure transients on WDS.

The literature review has revealed that live water distribution systems are vulnerable to numerous *other factors* affecting their structural integrity. Identification of the most significant factors for pipe failures, and the associated remedial measures, has become of interest to many researchers. Numerous factors are thought to cause pipe failures, and many of these have been conclusively proven: internal variables (water temperature and quality); structural pipe characteristics (length, diameter and pipe material); and external (environmental) conditions. Shamir and Howard (1979) and later Clark et al. (1982) proposed the classification of the main factors into four groups:

- prior to installation — the quality and characteristics of pipes and their components including manufacturing defects or damages due to handling and inadequate design,

- installation techniques of buried pipes — which often may trigger high failure rate due to installation problems,

- the environment in which pipes are buried, such as type of soil, external loads, traffic, roadwork, and temperature differential,

- operating service (e.g., operating pressure and pressure transients) including standard of service (how well the system is being operated and maintained).

Some researchers (Kleiner and Rajani, 2001; Goulter and Kazemi, 1988) pointed out that failures in neighbouring pipes can trigger failures due to e.g., sudden change of pressure in the system. Lambert (2000) added to this list local factors such as topography and number of connections. Following this, Royer (2005) characterized the main factors for the occurrence of breaks into three main groups: chemical stressors, physical stressors and other factors. Under the chemical stressors group, he identified internal and external corrosion. Pressure transients was listed under the physical stressors group, along with other factors such as:

- traffic loads,
- soil loads,
- point loads,
- thermal stress resulting from temperature differences between water, pipe and soil,
- other damages such as dig-ins, and damage to external coating and internal linings during transport, storage and installation.

Under the other factors group he registered ageing, pipe flaws and installation defects. Also Morris (1967) identified that pressure transient may be one of possible causes of pipe failures. However, pressure transients were never proved to cause damage in complex networks and these recommendations were based on previous pipe failures caused by pressure transients in the vicinity of pumping stations.

2.3.3 Summary

Number of factors affecting the pipe repairs have been identified in the literature. Commonly accepted factors were age, diameter and static pressure with age being shown as a complex factor which depends on combining impact of numerous parameters affecting a pipe during its service life. None of them, however, was fully able to explain pipe repairs. Therefore, there is a potential that adding an extra parameter, e.g., pressure transients may help better describe pipe repairs and account for what has not been previously possible to measure. Current advances in technology allow such data to be obtained.

2.4 Approaches to modelling of pipe failures

The occurrences of pipe failures were previously investigated through various type of modelling. These included physical, statistical (regression based), neural network and evolutionary models.

2.4.1 Physical models

The literature review showed that pipe burst occurrences can be investigated, through physical modelling based on modelling the physical phenomena that lead to the occurrence of a pipe failure (Rajani et al., 1996; Rajani and Zhan, 1996; Zhan and Rajani, 1997; Rajani and Makar, 2000; Kumar et al., 1984). Physical models (deterministic or probabilistic) estimate pipe failure through the simulation of mechanical (e.g., soil - pipe) interactions. This approach is very challenging, as obtaining sufficient data combined with a good understanding of current pipe condition for every pipe is usually not possible. It is particularly difficult when trying to account for inconsistencies in e.g., ground conditions. Furthermore, it is not possible to obtain a complete load history (internal or external) even for a single pipe. The incorporation of this modelling requires the knowledge of complete pipe history of external and internal loading condition. In addition, some pipes date back to the Victorian era making these requirements more challenging. Therefore, these models are generally used when modelling major transmission mains or gas and petroleum pipes (Rajani and Kleiner, 2001). There has also been an attempt to physically model pipe-soil interactions (Rajani et al., 1996). Davis et al. (2007) have developed a physical model to estimate failure rates in PVC pipes as they age. This work was based on the UKWIR National Mains Failure Database from 17 UK water utilities. The model assumes the occurrence of cracking from inherent defects of the pipe wall, which can propagate under operating conditions. Despite these successes, modelling of the entire WDS is much more challenging as data required for these analyses are usually not directly available and expensive to obtain.

2.4.2 Statistical

Various strategies and algorithms have been developed to classify pipe failures in WDS. The most historically established and currently used classification tools rely on regression-based analyses. The explicit rules form a fundamental element of this approach. The majority of prediction models are built on multivariate regression or a regression approach evaluating various input factors and assesses their correlation with the target variable. The main advantage in using this approach is that regression models are suitable for short time burst histories, and therefore historically were widely used for providing insights into a pipe break's parameters. A significant limitation of regression based models is the assumption that pipe repairs constitute a continuous dataset following a continuous probability distribution.

The application of a simple regression model for a pipe failure analysis was presented by Shamir and Howard (1979). The authors developed exponential and linear models. An advantage in using this approach is relatively small data requirements allowing various factor aggregations.

However, neither goodness of fit was provided, neither significance of coefficient used. Other limitation of this model was omitting factors such as environmental, pipe failure history or all pipe characteristics.

Clark et al. (1982) have developed two regression models to describe a first failure and accumulated failures of pipes. Their multiple-regression provided greater insights into the factors contributing to pipe failures. Therefore, these models are more suitable for network repair and renewal than the model of Shamir and Howard (1979). Its limitation is that non-failed pipes cannot be effectively integrated in the model.

2.4.2.1 Life time models

Life time models are based on the complete burst histories of each pipe in a given dataset. These models usually use all available information to predict future pipe failures (Lei and Saegrov, 1998; Andreou et al., 1987). Unfortunately majority of datasets can be old, therefore complete burst histories for these are unavailable. Lei and Saegrov (1998) proposed two models: 'counting process' and 'lifetime model'. According to the authors lifetime models can estimate the probability of a pipe failure within a specific time limit.

2.4.2.2 Generalized linear models (GLMs)

Linear modelling is predicated on various characteristics of the variable trying to be explained. The wider class of such models may be used, these are called 'generalized or mixed effect linear models' which are special cases of linear modelling. In a simple way the Generalized Linear Models (GLMs) transform continuous modelling into probability by adding a function that allows to work linear on non-linear relationships. The polynomial regression can be used because it is linear in terms of parameters. GLMs provide an extension of linear models, a generalization of conventional linear regression, for non-normal distributions of the target variable and non-linear transformations.

Many researchers have chosen generalized linear models for modelling pipe repairs. It is because pipes have two types of behaviour (they burst or do not burst). This behaviour can only be described by non-linear relationships, for instance, by Poisson distribution or binomial distribution. The use of GLMs with binary representation of repairs was presented by Yamijala et al. (2009). Another way to express probability of failure is to express repairs as categories, such as repair, no repair, about to repair. In this case, multinomial regression can be used. However, the most popular practice is treating pipe repairs as numerical, but discrete events which can be *0, 1, 2,* or *3* repair events in the specific interval. In that case repair events are simply counts. To model

the probability of a pipe having *1* or *2* or *3* repairs the Poisson discrete probability distribution is used. The use of GLMs with Poisson distribution was shown to be sufficient to reveal strong associations between repair rate and age, diameter and length (Boxall et al., 2007). The Poisson GLM is usually based on the assumption that the conditional mean is equal to the conditional variance which may not be valid for some count datasets. When conditional mean is greater than the conditonal variance the data is overdispersed and other models for handling overdispersion should be applied.

Debón et al. (2010) compared the models for evaluating the risk of failure in water supply network using receiver operating characteristic curve (ROC curve) by plotting the true positive rate against the false positive rate against various thresholds. The authors concluded that the best model was generalized linear model (GLM). They compared Cox model, accelerated lifetime model (using Weibull distribution) and GLM. Different variables were investigated as having a possible impact on risk of failure in water networks. The authors have shown that pipes which are less predisposed to failure are installed under a side walk, have low pressure, large diameter and short length. GLM and Weibull models showed that ductile cast iron pipes were less prone to burst.

Yamijala et al. (2009) used GLMs to estimate/model the likelihood of pipe repairs in the future and determine the parameters that the most affect the likelihood of repairs. Data were divided into two 6-month periods (November - April and May-October) to capture the periodic changes in soil moisture levels. Models were only applied to reduced data set consisting of those pipes that had experienced a failure during the data-recording period. The selection was implemented to prevent from an excess of non-failed pipes (zeros) above what models could reasonably predict, termed as zero-inflation problem. Therefore, use of data with excessive zeros require methods, which are able to handle this problem. However, for logistic GLM all pipes were included in analysis. The study compared four models (time linear, time exponential, GLM and logistic GLM) and concluded that logistic generalized linear model provided good estimates of pipe reliability. None of the models of Yamijala et al. (2009) provided particularly strong fits with dataset. For logistic GLM utilities would be more interested in the probability of having at least one repair than in predicting the precise number of repairs on a pipe segment. Presence of even one failure of a pipe causes costly repairs.

2.4.2.3 Bayesian statistics

Previous (regression) modelling approaches are often termed as frequentist as in contrary to Bayesian. Differences between both approaches is the interpretation of probability. Bayes theorem is based

on a prior, a likelihood function, a posterior and a probability constant and accounts for uncertainty in the data. The prior represents what is known about parameters before data is seen. In the fequentist approach there is no prior and once the model has been specified, only MLE needs to be optimised. The properties of that estimator can be derived, which is usually challenging. In a frequntist approach it is difficult to state something about variability of the estimator, in Bayesian approach this can be achieved because it is a distribution. Therefore, it can be treated as a probability distributions and any measure of uncertainty can than be used (i.e., means, variances, covariances). However, in order to estimate the prior, not only knowledge is required, but also large dataset to test the assumptions. This is not always available for all model parameters. Therefore, the estimation of the prior is a challenge. Despite this fact Bayesian approach was used for prediction of pipe failures (Economou et al., 2007; Watson and Mason, 2006)

2.4.3 Artificial Neural Network modelling

Soft computing such as artificial neural network (ANN) is an attractive option to precise computing, as its tolerant of uncertainty and imprecision. Soft computing models are subtype of statistical models that can be successfully used for the assessment of historical pipe repair data. According to Berardi et al. (2010) ANN are based on the mechanisms underlying the behaviour of neurons. They can be roughly defined as general purpose regression techniques. ANN main feature is the capability of fitting the data so that they are also known as *universal regressions*.

2.4.4 Self Organizing Maps

Kohonen Self Organising Maps (SOMs) are one of Artificial Neural Network (ANN) approach for clustering and visual exploration of data. SOMs utilise self-organising process as form of unsupervised learning described by Kohonen as vector quantization and projection algorithm. SOMs do not give a mathematical expression at the end and neither give any statistic which could evaluate the outcome. SOMs aim at reproducing the topology (the same neighbours). In other words if two high-dimensional objects are very similar, then their position in a two-dimensional plane should match. SOMs use a regular grid of 'units'/neurons and into them the objects are mapped (not into a continuous space). Object is mapped to the neighbouring by the search for similarity. SOMs concentrate on the strongest comparability. This method can be considered as an analogy to the k-means clustering where every unit/neuron corresponds to a cluster. The number of such clusters is defined by the size of the rectangular grid. The method can be also seen as the data dimensionality reduction, because it provides two-dimensional view of high-dimensional data.

SOMs have been successfully used for microbial and physico-chemical data to determine which parameters affect regeneration of discolouration material (Mounce et al., 2016) and to examine relationship between water quality and asset data (Mounce et al., 2014; Blokker et al., 2016).

2.4.5 Evolutionary polynomial regression (EPR)

Genetic programming (GP) is the most widely data-driven pattern recognition techniques that was used to capture non-linear interactions between various parameters and model complex engineering problems. It is used to find the appropriate mathematical model structure to fit through a set of points. The whole population of such functional expressions is created and searched through based on how good each of them fits the data.

Evolutionary Polynomial Regression (EPR) is a method which integrates genetic programming with numerical regression to develop easily interpretable mathematical equations. Giustolisi and Savic (2006) describe EPR as a stepwise regression which integrates the effectiveness of genetic programming and search strategies with the advantages of numerical regression for estimating model coefficients. According to the categorization of modelling technologies based on transparency level (Ljung, 1999), EPR is classified as a grey box technique.

The EPR process has two stages:

(a) Search for symbolic model structures using Genetic Algorithm (GA).

(b) The estimation of model parameters, i.e., polynomial coefficients, (a_j) (for each combination of inputs) using least-squares (LS) method.

At the first stage EPR identifies the model structures. The random population of the exponents vectors is created and assigned to columns of candidate input variables matrix. These exponents are set of values defined by the user which also include zero (to exclude some of the combinations from the equation). When the population of model structures are created the second stage is performed (least-squares polynomial coefficients estimation). The least square developed model equations are then evaluated/ ranked through Multi-Objective Genetic Algorithm (MOGA) (Giustolisi and Doglioni, 2005). The obtained model equations are ranked according to the following criteria:

(a) Number of polynomial coefficients, number of input variables (explanatory variables), or both (Savic et al., 2009). The ranking is based on minimization of the number of polynomial coefficient and/or input variables in the symbolic model equation called *parsimony*. Parsimonious equations (i.e., more parsimonious model structures) with a few polynomial coefficients are preffered over long and complex equations.

(b) Accuracy (model fitness to the data). The ranking is based on maximization of the model accuracy.

MOGAs are known as population based search and optimization techniques. All model equations that corresponds to the optimal trade-off between model parsimony and fitness (Giustolisi and Savic, 2006) are determined. The returned model parsimony was identified as the main advantages in using EPR. One or both parsimony criteria can be selected by performing a two- or three-objective optimization while searching for model equations. Two objective optimization function can be either minimization of number of polynomial coefficients or minimization of number of input variables and then maximisation of model accuracy. The search results in maximization of model's parsimony. The three objective optimization performs minimization of number of polynomial coefficients, minimization of number of input variables and maximization of model accuracy simultaneously.

The ability of EPR to return a symbolic model equations became especially attractive for engineers. EPR has been widely adopted in many areas including sewer networks, water distribution systems, hydrology and structural engineering. EPR has been used to model repairs in water distribution systems on failed and non-failed pipes (Berardi et al., 2005; Kakoudakis et al., 2016), sewer systems (Savić et al., 2006). Berardi et al. (2008) have used EPR to model pipe deterioration.

EPR returns possible model equations from which the user need to select the best candidate. The selection should be based on analysis of similarities and differences between different symbolic equations. The trade off between their complexity and fit should be considered. The user should identify common terms which appear the most frequently in the model equations and which terms are discarded when structure complexity decreases (Giustolisi and Savic, 2006; Giustolisi et al., 2008). It has been highlighted that terms, which appear only once represents a weak relation with the physical phenomenon being investigated and are likely to represent error realization from the data (Giustolisi and Savic, 2009).

2.4.5.1 Goodness of fit measure

The model fit to the data is described by the Coefficient of Determination (CoD):

$$CoD = 1 - \frac{\sum_n (\hat{y} - y_{exp})^2}{\sum_n (y_{exp} - avg(y_{exp}))^2} \tag{2.4}$$

where n is the number of samples, $\hat{y}$ is the value predicted by the model, y_{exp} it the value measured and $avg(y_{exp})$ is the average value of the corresponding observations (evaluated on the n samples). This measure of fit is a relative value depending on data used. This measure of fit was used by Berardi et al. (2008); Altomare et al. (2013); Giustolisi and Savic (2006); Laucelli

et al. (2014) in the assessment of the EPR model equations. Previous studies (Kettler and Goulter, 1985) obtained high (i.e. 0.82) CoD values for cast iron pipes but lower (i.e. 0.75) for asbestos cement pipes.

2.4.5.2 User choices

General configuration of symbolic model equations can be introduced by the user with the choice of the structure of the expression depending on prior knowledge (i.e. physical understanding) about a model. There are seven possible general model structures considered when searching for a best model structure (Giustolisi and Savic, 2006). The user can specify an inclusion of exponential and logarithmic functions representing physical insights in a final model structure. However, it has been stated that if these functions are not really describing the underlying phenomenon (target variable) their exponents may be equal to zero. The user can also make a choice about input variables, maximum number of polynomial coefficients and exponents. The user can also force the LS search to only positive parameter values in a model structure. This is in addition to the unconstrained LS. Non-negative least squares implies more parsimonious model being selected with respect to the number of parameters being selected and improving the selection process.

The set of exponents (values within user-defined bounds) can be any discrete values providing that value of zero is included. EPR can then deselect the input variable from the final model structure when the variable is considered not significant enough in describing the phenomena being analysed (Giustolisi and Savic, 2006). When the exponent assigned to an input variable becomes equal to zero the variable is deselected from the model structure. A global search algorithm is performed simultaneously on both the best set of combinations of input variables and their exponents by integer GA coding, which determines the location of the candidate exponents in a form of a matrix.

The user can decide whether the input and/or output should be scaled. The choice to scale can be considered as filtering values, which may result with some information in the data being lost. Having the same order of magnitude for all data could help with the best model search process. Scaling is especially suggested for time series models, whereas for static models data may be not-scaled.

2.4.6 Summary of failure modelling approaches

The literature review showed that an explicit observation of the relationship between transients and the observed pipe repair rates does not exist in the published literature at present. The transient contribution analysis can be satisfactory performed in areas that have adequate (i.e., high

frequency) pressure data. To the best of the author's knowledge this kind of extensive monitoring has not been previously undertaken; consequently none of these analyses have been carried out. The complex problem of expressing the probability of pipe repairs can be investigated in many ways, for instance by transforming continuous modelling into a probability or data mining. With all failure prediction models they are only as good as the data provided. Different water utilities have different data storage procedures, therefore model developed using data from one water utility may not be as successful when using data from an alternative utility.

2.5 Summary

It is currently unknown whether it is large transient events that effect networks and cause pipe failures, or if long term fatigue loading of repeatedly occurring pressure transients is a greater risk. From the literature it is known that catastrophic failures are primarily due to extreme events observed usually in a single pipeline (see section 2.1.4). Smaller magnitude, repeated pressure transient events, however, could impose excessive loading on a pipe, successively weakening it and eventually contributing to its failure. Furthermore, if the pipe has been already weakened, for instance, by corrosion, the cyclic loading (whether significant or not) could also accelerate its failure. Fatigue analysis is designed to deal with cycling loading imposed on a pipe. It has been previously recognised (see section 2.3.2) as a technique which could possibly help to understand transients. There have been only limited studies investigating pipes exposed to pressure transients. The cumulative fatigue effect on a pipe, by a range of pressure transient events, has not been assessed even in a simple pipe network. Causal links between pressure transients and pipe repairs becomes much harder to determine in operational WDS, where complex pipe arrangements are present. It is mainly because the repeating occurrences of pressure transients in operational WDS remain largely unknown. Fatigue analysis has helped to understand that magnitudes of pressure transient events and number of their occurrences are important factors (along with e.g., material properties of a pipe) for the assessment of their impact on the pipe.

An important component, which has previously been omitted in pressure transient characterisations, is the duration of a pressure transient event. This component has not been considered despite the fact that greater risk of damage to a pipe may be associated with more sudden pressure transient events. Large events happening over a shorter time may have different impacts on and pose different risks to the network than the same magnitude events happening over a longer period. For these reasons time may be an essential factor to consider when analysing the impact of pressure transients. Consequently, a high potential for damage to a pipe would be a combination of pressure transient magnitude and duration with a greater contribution to damage attributed to more sudden

events (not only to events of large magnitudes).

The protection of infrastructure and control of the pipe repairs has become a high priority concern to water utility companies. Identification of the crucial factors that influence pipe repairs has become the focus of widespread research and many factors, which could influence the condition of the water network have been investigated. Extensive analyses of pipe repairs have provided insights into the possible factors that can affect pipe failure rates. Depending on the problem analysed, and often the tools available, some of these factors have been modelled to reveal a greater understanding of the pipe failure process. A majority of analyses considered pipe intrinsic factors. In this approach, pipe material was typically combined with pipe diameter and length (Kettler and Goulter, 1985; Shamir and Howard, 1979), age (O'Day, 1983; Røstum, 2000) which is usually used as an indicator of the length of time that the pipe has been exposed to other factors and soil characteristics, which were often associated with ground movements (Kleiner and Rajani, 2000). The challenges usually arose when dynamic factors such as corrosion were considered (Ahammed and Melchers, 1994; Royer, 2005; Thomson and Wang, 2009). However, despite extensive research efforts, there is no definitive answer as to which specific factor(s) cause pipe repair. According to Royer (2005), for a particular pipe characteristic such as age, dimension, joints, connections, valves, material, lining, coating, bedding, external and internal loading, there may be a few or many contributing factors that determine the occurrence of this pipe to fail. Multiple and varying causes of pipe failures increase the level of unpredictability of a pipe failure. There is, however, limited work in the literature identifying the most important factors for estimating pipe burst occurrences. Given the complex and interrelated nature of pipe repair occurrences, many factors such as physical, hydraulic and environmental (climate data, such as temperature) are considered. The factors influencing pipe repairs might include localised asset characteristics such as pipe material, age and diameter. However, hydraulic conditions, such as static or dynamic pressures may also be important. Therefore, a careful choice of factors is essential for each analysis. It is important to highlight that the contribution of pressure transients to the observed pipe repair records was not investigated in any of these studies.

Chapter 3

Aims and Objectives

The aim of this research is to enhance the understanding of the occurrence and impact of pressure transients in water distribution networks and to assess the extent to which pressure transients contribute to the observed repair records of water supply infrastructure in combination with other variables.

The detailed objectives of the research are to:

1. Study the occurrence of pressure transients in water distribution networks by undertaking a large-scale, long-term data collection programme and to characterise and quantify pressure transients in these systems:
   (a) collect data from water networks with diverse range of characteristics in relation to materials, pumps, pressure controls and customer types,
   (b) develop and apply novel method to characterise pressure transients in terms of their magnitudes, durations and number of their occurrences.
2. Identify and apply methods to determine the contribution of pressure transients, along with other variables, to pipe repair history:
   (a) identify, assess and critically review potential methods used for linking historical pipe repairs data to possible explanatory variables and select the method which describes pipe repair history and allows integration of different pressure transient characteristics,
   (b) identify, assess and select key explanatory variables contributing to pipe repair history.

Chapter 4

Fieldwork

4.1 Introduction

Collection of pressure data in UK water distribution systems is typically undertaken at a 15 minute resolution, the accepted standard for pressure monitoring required by regulators. To reveal the existence of any dynamic pressure changes, however, a faster frequency recording is required. Such monitoring has not been widely implemented and as a result the existence of pressure transient waves in WDS is currently largely unknown beyond special cases such as close to pump stations. This research collected data from wider water networks to provide detailed overview of pressure transient occurrences. This chapter presents how network transient data were collected through custom-built, high-speed (high frequency) pressure measurement instruments. The aim of collecting this data was to investigate, for the first time, the in depth extent of the occurrences of pressure transients in water distribution systems. The instrument location methodology and selection of a range of different operational networks enabled the collection of many examples of pressure transients. This unique data was collected to investigate transferring the recorded time series signal into a meaningful and site specific representation showing what a given pipe experiences.

4.2 Methodology

With advances in instrumentation, high speed pressure monitoring is now feasible and allows an insight into the occurrence of pressure transients in operational WDS. A limited number of pressure transient field studies have been published (see section 2.1.7). Previous studies into operational WDS transients focused on documenting the existence of low and negative pressures mainly around pumping stations. These studies highlighted that the analysis of transient behaviour

requires a time-series history at a high recording rate. For instance, it has been shown that a high sampling frequency of 100 Hz can be used for pressure traces recorded in operational WDS. Some studies have used 300 Hz and more (for details see section 2.1.6). However, it remains unknown how pressure transients may cause asset damage and how they should be measured across different operational networks to reveal this information. An evaluation of the occurrences and sizes of not just extreme but also regularly occurring pressure transients in diverse water systems needs to be undertaken. The methodology for the collection of continuously-recorded high speed dynamic pressure data was, therefore, designed to ensure capturing of random and regular pressure transients of variable magnitudes. The collection of this data would facilitate characterisation and subsequently association of pressure transients with network performance.

4.2.1 Fieldwork requirements

To achieve objective 1 (a) of this research the following requirements were identified:

1. To establish a sufficient recording frequency to firstly capture and secondly accurately and sufficiently define pressure transient waves occurring in WDS.

2. To establish a methodology to gather the data from diverse water networks (sites) considering:

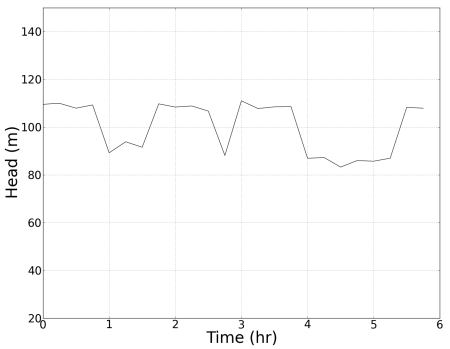
    - how many sites should be monitored,
    - how many locations measuring pressure data are required in the specific site,
    - total number of pressure measurement instruments required for the data collection.

3. To establish how long time series pressure data should be recorded from one location.

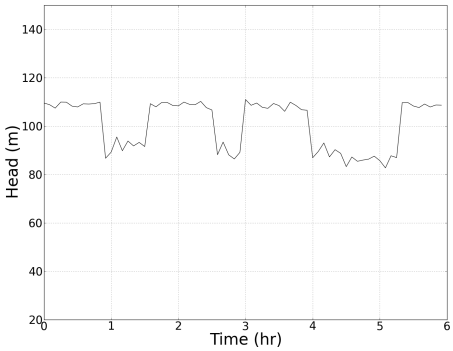
4.3 Recording rate

The frequency of the recording should be sufficient to capture and reveal how pressure transients propagate and change within networks. Therefore, for an estimation of the required recording frequency qualitative assessment of data samples from real water networks was undertaken.

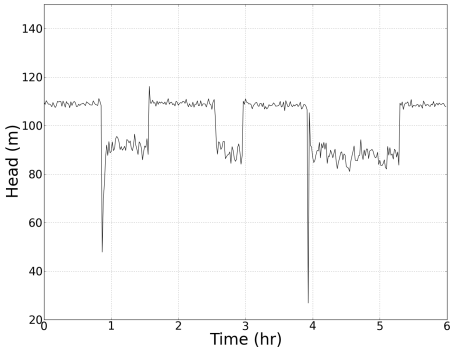
Two samples of pressure data collected from different operational WDS during initial exploratory fieldwork are presented. These samples illustrate how an increase in recording frequency above the level of 15 minutes is necessary to capture pressure transients (see Figure 4.1 and Figure 4.2). The figures show that a interval of 1 second visually captured 'naturally' occurring pressure transients (see Figure 4.1 (e) and Figure 4.2 (e)).



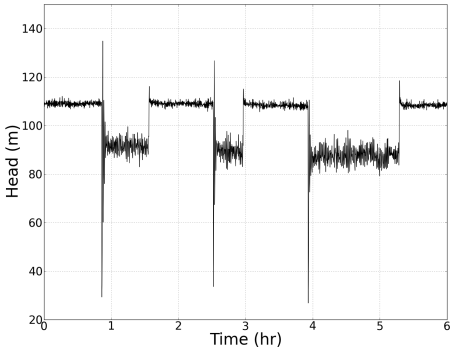
(a) recording interval of 15 min



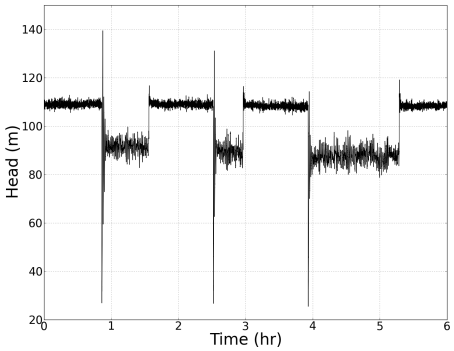
(b) recording interval of 5 min



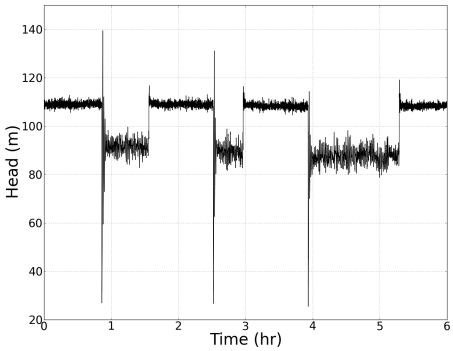
(c) recording interval of 1 min



(d) recording interval of 10 sec

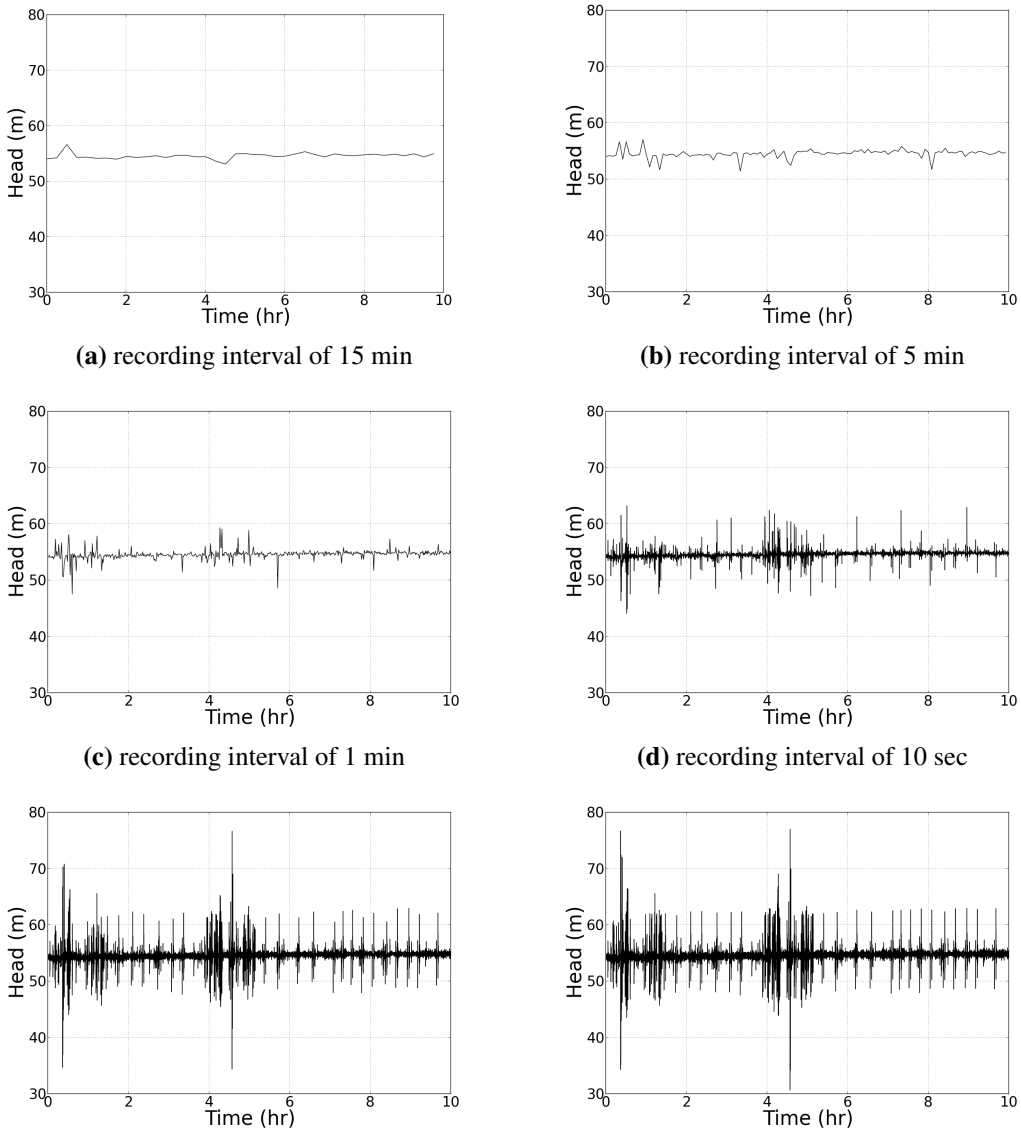


(e) recording interval of 1 sec



(f) recording interval of 0.01 sec

Figure 4.1: Down-sampled pressure traces recorded from an operating WDS at the same point.



(a) recording interval of 15 min

(b) recording interval of 5 min

(c) recording interval of 1 min

(d) recording interval of 10 sec

(e) recording interval of 1 sec

(f) recording interval of 0.01 sec

Figure 4.2: Down-sampled pressure traces recorded from an operating WDS at the same point.

Further examples of pressure data, presented in Figure 4.3 and Figure 4.4, show in greater detail different shapes and magnitudes of pressure transients that occurred in operational WDS with a recording intervals of 10 sec, 1 sec and 0.01 sec.

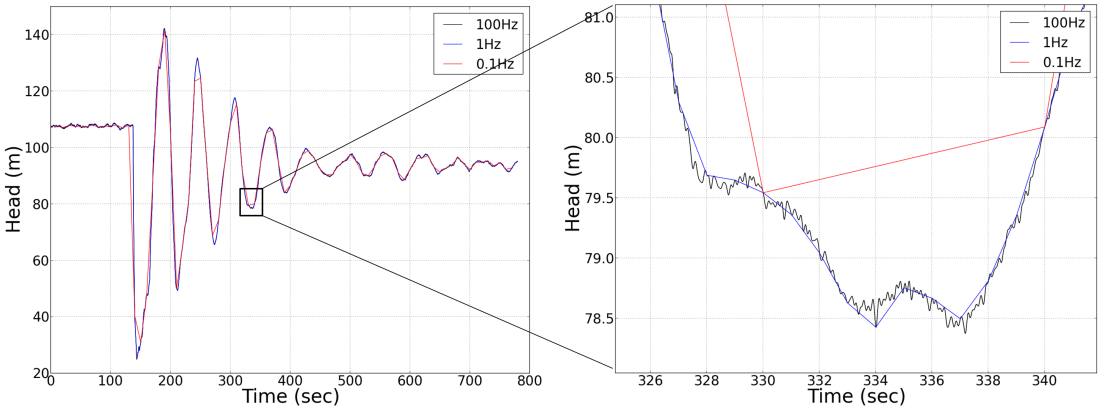
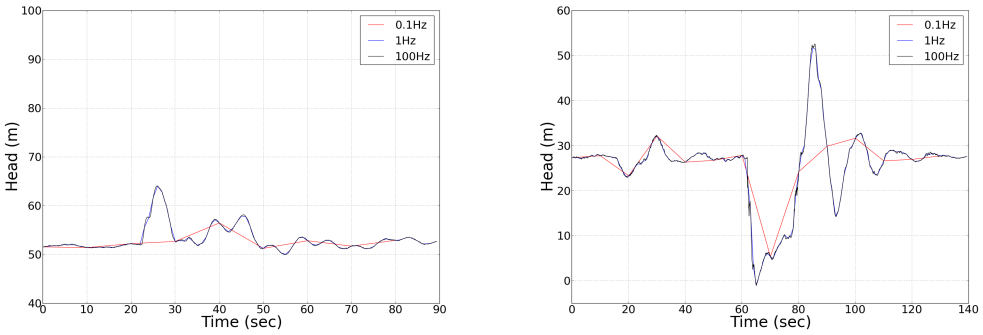
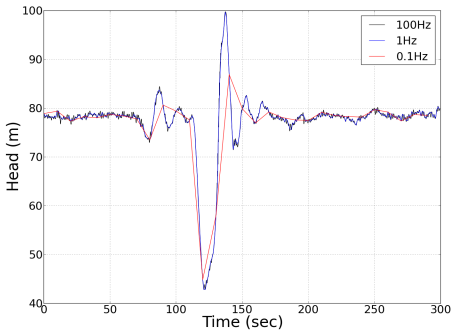


Figure 4.3: Selected pressure transient wave from Figure 4.1 recorded from an operating WDS down-sampled to 10 sec, 1 sec and 0.01 sec (0.1 Hz, 1 Hz and 100 Hz) and magnification of pressure transient wave.



(a) **(b)**



(c)

Figure 4.4: Comparison between different sampling rates show examples of different pressure transient waves recorded from diverse operating WDS sampled by 0.1 Hz, 1 Hz and 100 Hz, (a) 'small', (b) 'medium' and (c) 'large' size of pressure transient wave.

The examples demonstrate that at least 1 Hz resolution is required to visually identify a pressure transient wave from time series data. A 100 Hz resolution, which is two orders of magnitude higher, should be sufficient to detect and subsequently characterise pressure transient waves. Higher frequencies could provide better description. Although digital memory options which capture several weeks of data exist, an excessive sampling rate would limit the battery capacity and therefore the duration of the recording in the field.

Planned fieldwork aimed to cover long duration (potentially weeks) recordings of time series pressure data without regular access. This requires an instrument with a sufficient power pack. Available battery technology, suitable for deployment in fire hydrant chambers limited data collection period due to the battery size. Considering these practicalities a suitable technology combination was identified. This allowed a maximum 100 Hz to be used as it was decided that for all field monitoring covered by this research this optimum frequency should be used. The resolution of 100 Hz allowed to record 60 mln data points per week at each location.

4.4 Transient data collection

The field monitoring program was designed to cover significant time periods such that randomly and regularly occurring pressure transients could be captured. The dynamic pressure data was collected over an 11 month period from a UK water company distribution network. The monitoring program also aimed to capture a representative proportion of the network which is divided into district metered areas (DMAs, hereinafter known as sites). This also included zones that have a pressure reducing valve (PRV) installed at the inlet to control pressure (hereinafter zones).

During this large-scale and long-term pressure transient research, a vast amount of data was collected. To aid the data collection process, geographic proximity of sites was considered as an important factor during the site selection process.

4.4.1 Selection of sites

Prime criterion for site selection was a site repair histories. Anecdotal evidence from UK water utilities suggests that high frequencies of pipe repairs exist as a result of pressure transients, this was also previously suggested (Skipworth et al., 2002). Provisional assessment of sites according to their repair histories was undertaken. The average failure rate was calculated for all available DMAs and was equal to 2.5 repairs per kilometre. If a site had an average value less than the average it was classified as a low pipe repair frequency site. Sites with higher numbers of repairs were classified as high repair frequency sites. Data was collected from sites with high and low repairs frequencies; 34 sites with high and 33 sites with low numbers of pipe repairs. This criterion

was made to prevent bias analysis if collecting data only from sites with high pipe repairs.

Selection of sites aimed to include those sites with expected substantial and also low (including zeros) numbers of pressure transients. This approach would provide evidence of the diverse range of pressure transients. A review of different transient sources was undertaken to aid the selection of sites where pressure transients could be expected, i.e., traditionally associated with transients and traditionally not associated with transients.

4.4.1.1 Sites associated with transients

Pressure transients have historically been strongly associated with pumped systems, especially when switching on and off. Some of these events had severe consequences as described in section 2.1.4. Sites in close proximity to pumping stations were, therefore, targeted (5 sites, see Table 4.1).

Table 4.1: Sites with traditionally associated pressure transients sources.

Sources	Number of sites
Pump	5
Industrial user	23
Sum	28

Pressure transients also result from changes in flow velocity during any fast closing or opening of the valves. Sites with water related businesses or industrial users were therefore also identified. Industrial users are potential pressure transient sources due to opening and closing of valves related to, for instance, high water demand, drawing water from the network to fill water tanks or other water consumption activities. Industrial users are suspected to generate pressure transients which then propagate into the system. 23 sites with key industrial customers were selected and classified as 'industrial' (see Table 4.1). For pressure transient prone areas, including sites directly associated with pumps, 28 sites were monitored in total.

4.4.1.2 Sites less likely associated with transients

Along with monitoring transient prone areas, data were collected from a wide range of sites where pressure transients were expected to be less common or even not present, e.g., those which were mainly residential and these were classified as 'residential'.

Pressure reducing valves (PRVs) were considered as a potential pressure transient source. PRVs require maintenance and if not maintained may lead to unstable networks (see 2.1.1). The malfunction of these devices can possibly cause pressure transients. Overall 41 selected sites had

a PRV installed. 12 PRVs were located in sites of low average pipe repairs per kilometre and 12 in high failure sites. Eight sites were indirectly associated with pumps. In total 39 sites were monitored in areas less likely associated with pressure transients.

Sites identified for monitoring are summarised in Table 4.2. Some sites have a PRV and a pump installed, therefore numbers of sites in Table 4.2 are overlapping.

Table 4.2: Number of sites selected during site selection process for field monitoring.

Literature review expectation	Criteria: Average number of pipe repairs per km	Criteria: Possible sources	Number of sites
Associated with pressure transients: pumps and industrial users (28 sites)	high (16 sites)	one	7
		multiple	9
	low (12 sites)	one	3
		multiple	9
Less likely associated with pressure transients: residential areas and PRVs (39 sites)	high (18 sites)	one	12
		multiple	2
		none	4
	low (21 sites)	one	15
		multiple	1
		none	5
67 sites	67 sites		67 sites

4.4.2 Number of measurement points

To achieve some indication of pressure transient occurrence within a site at least one high speed monitoring instrument is required per site. Ideally this is situated centrally within the site, on the main supply pipe, to characterise the pressure transient response across the site.

To further characterise pressure transient responses in different parts of the network and other pressure changes a second monitoring point was required. This would be at another location, usually at the periphery of the site. This multiple location strategy would help to understand how pressure transients spread and change within sites. The simultaneous use of more than one pressure logger per monitored site also provided for cross verification of the results from more than one source.

4.4.3 Instrument deployment procedure

The data collection process followed a ‘milk round’ approach, where fully charged and data-free equipment was exchanged with previously deployed and data-replete ones. To facilitate the requirements for field data collection (e.g., number of sites, two measuring locations per site, instrument battery

size to fit into the standard hydrant chamber, battery lasting time and 'milk round' instruments and data collection process) 40 bespoke loggers were used. The geographic proximity of the selected sites, the requirements of a 'milk round' data collection process and required number of sites resulted in the overall network coverage as presented in Figure 4.6.

Some of the physical logger deployment and exchange was organized with a UK water company field technician who (after being trained by the author) collected the instruments measuring the pressure data from the University of Sheffield and deployed them in the given and specified locations. Records were kept in log sheets.

During the fieldwork some of the locations for the instruments were limited due to hydrant accessibility (see Figure 4.5) or due to site accessibility (security issues). In other cases the data was not collected (instruments were sometimes disconnected by a third party).

Figure 4.5: Hydrant issues: difficult to access (highways, private premises, loggers fitted), poor conditions (leaks, old, rusty), not present, waste from digging out.

4.5 Instrumentation

To facilitate frequency recording requirements for field data collection a high resolution data acquisition instrumentation was identified and supplied. Custom-built pressure data loggers manufactured by Race Technology were adapted in the University of Sheffield for use within water networks. The loggers (see Figure 4.7) were converted for deployment on hydrants and washouts using a standard hydrant cap with push fit connectors. Each logger comprised a DL1 data logging system powered by a lithium-ion battery pack enabling continuous field pressure monitoring on site. The logging system (DL1) was connected to a 0-20 bar pressure sensor of $\pm 1\%$ accuracy (as stated by supplier). The DL1 was modified to include a real time clock which was externally stamped using GPS. The DL1 was preconfigured to log a single analogue input at a rate of 100 Hz into a 4

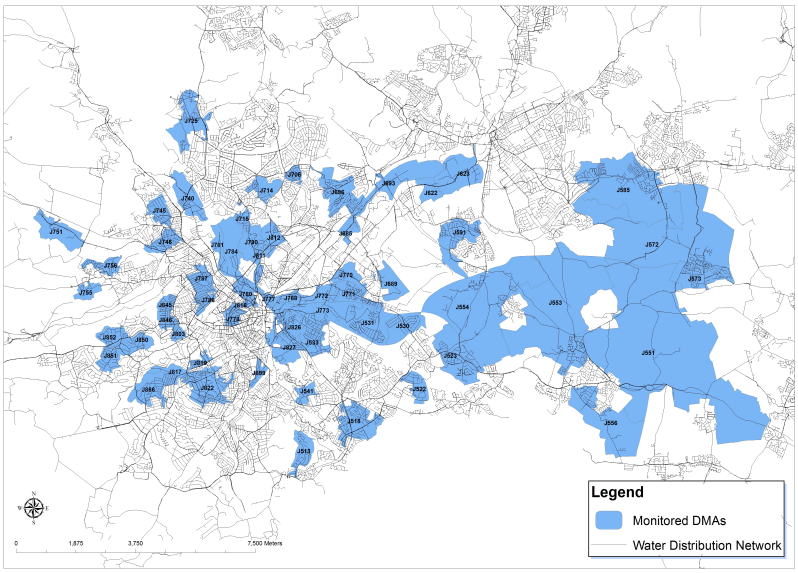


Figure 4.6: Locations of the sites that were monitored around Sheffield.

GB Compact Flash memory card. Software was supplied by Race Technology to convert logged data into a text format. For the collection of pressure transient data from multiple locations it was decided to include a GPS real time clock in the logger requirements. The use of GPS time allows correlating signals from different locations in a system. That technique was also previously successfully applied by Stoianov and Nachman (2007).

(a) **(b)**

Figure 4.7: (a) Example of the loggers developed and used during the monitoring programme and (b) logger sensor attached to the standard hydrant cap.

4.5.1 Instrument precision

To validate the supplied instrumentation and to establish a minimum value for a difference in pressure readings the precision of the instrument was determined. Equipment precision describes

how close a measurement comes to another measurement and is usually determined by a standard deviation. Below this value the difference between readings cannot be stated. High precision is indicated by low standard deviation and low precision by high standard deviation. The values of the instrument precision help to verify when the equipment provides accurate results and it was used in the analysis.

To establish the precision of the instrument the University of Sheffield laboratory facility was used. It was a standard recirculating system supplied by a variable speed pump. It comprised of a single 140 m long and 50 mm nominal internal diameter medium-density polyethylene (MDPE) pipe. Along the pipeline one test section and four butterfly valves were installed. Loggers were attached to the system at the test section.

For a range of static pressures the data was taken for approximately 2 min during which valves isolating the test sections were closed and the pump switched off. Then the pump speed was increased, valves opened and the procedure of recording data repeated. For each step change of pressure a standard deviation of a linear best fit line was recorded as presented in Table 4.3. The table shows the example of two different instruments to illustrate the calculation of the instrument precision.

Table 4.3: Examples of standard deviations taken for two instruments at different pressure steps.

Pressure head step number	Pressure head value (m)	Standard deviation (m)	
		Instrument 1	Instrument 2
step 1	3.5	0.059	0.062
step 2	7.8	0.060	0.059
step 3	14	0.061	0.063
step 4	21.5	0.061	0.057
step 5	30.5	0.058	0.054
Average		0.060	0.059

The average value for all step changes was calculated, which indicated the precision of the instrument. The obtained precisions per instrument were then averaged, which gave the overall instruments precision equal to 0.06 m.

For pressure transient events, if the measured value is lower than the precision of the instrument there is no statistical evidence that this is a valid reading. Two standard deviations represent 95% of the data, if the data is normally distributed. Anything which is beyond the limit of two standard deviations is then considered as a valid reading. Based on the obtained value of standard deviation the precision of the instrument is equal to 0.061 m, the value of two standard deviations is then 0.12 m. Values below two standard deviations should not be considered during the analysis of

the recorded pressure transients. This decision influences the assessment of the pressure transient characteristics e.g., size of pressure transient waves.

4.5.2 Instrument calibration

The loggers were validated and calibrated firstly by the supplier and secondly in the UoS facility. The calibration process was done using a pre calibrated digital transducer (reference pressure transducer) RDP Transducer Indicator (model-E308) in the laboratory prior to deployment in the field. One logger at a time was calibrated. Data was recorded at 0.01 sec intervals (maximum frequency). At approximately 1-2 min intervals the pressure in the system was increased (by increasing a pump speed) and the actual pressure was recorded from the reference digital transducer. Data from each logger was then plotted and an average voltage value was calculated for the stable period of each pressure step, after 'settling'. The average recorded voltage for the loggers was then plotted against actual pressure obtained from the reference pressure transducer, and the equation describing the straight line through the points was used to adjust the gradient and offset to the logger data, R^2 checked if close to 1 (a typical example can be seen in Figure 4.8). The original data recorded was then adjusted for gradient and offset and replotted (see Figure 4.8c). It can be noticed that each increase of a pump speed created more noise in the system.

It should be noted that pressure data loggers were calibrated in the factory, and the calibration carried out at the UoS was in effect an ongoing calibration on top of the original supplier calibration to check for any change with time. The gradient obtained in the laboratory ($y = 50.28x - 32.56$) remains very similar to the supplier settings ($y = 50.3x - 34.9$) and only the offset changes. The calibration of the data was a standard procedure prior to loggers deployment in a field.

4.5.3 Correction for instrument drift

Pressure data measurement instruments experience change over time which is manifested in drift of the data (see Figure 4.9 (a, b, c)) if assumed linear. The drift can be corrected according to the drifting rate during the deployment period. The correction of the instrument drift was done by taking mean values from the beginning of the dataset when the instrument was not connected to the network and from the end of the dataset when the device was disconnected from the network. A straight line was obtained connecting both mean values. This line represented a drifting rate during the period of deployment. The pressure data was then adjusted by fitting the line to all datasets (see Figure 4.9 (d,e)). All time series pressure data locations were corrected for instrument drift.

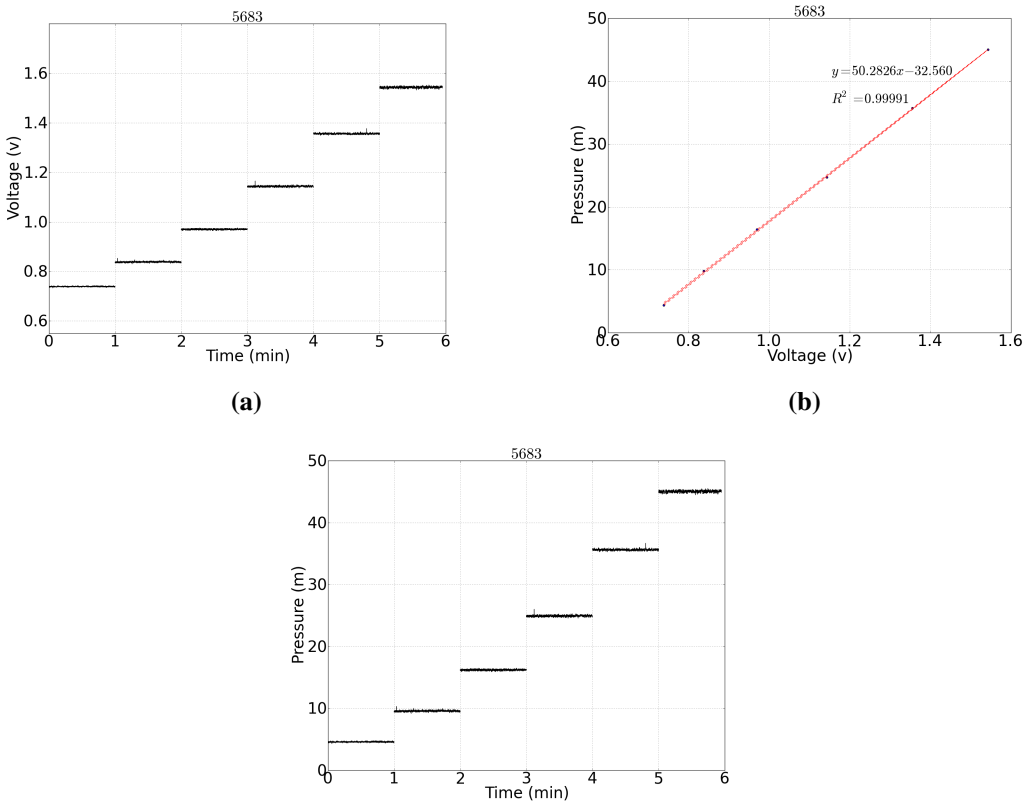


Figure 4.8: Example graphs of pressure loggers calibration process; the increase in noise is due to the increasing pump speed, (a) raw data, (b) calibration coefficients, (c) calibrated pressure data.

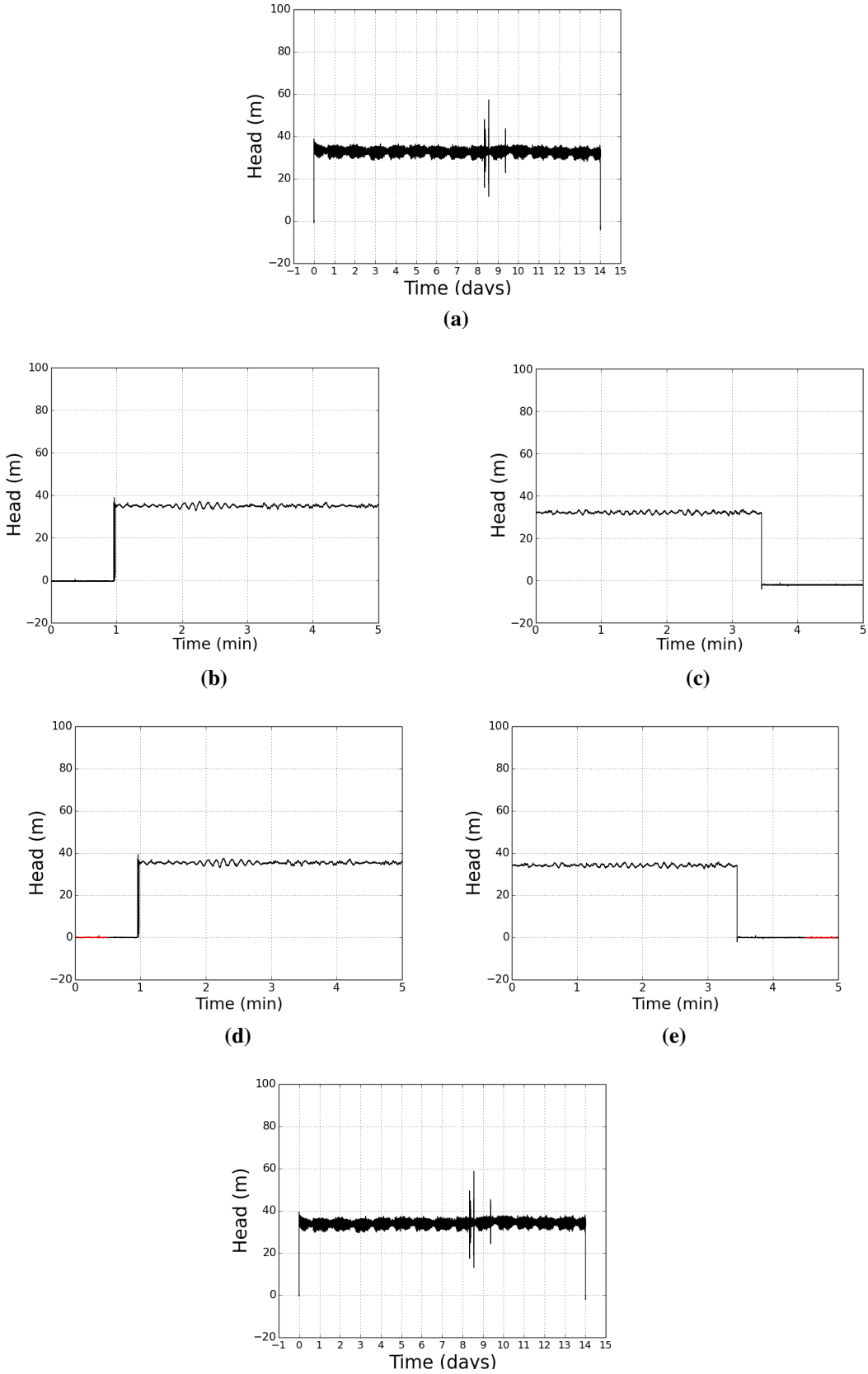


Figure 4.9: Example of pressure instrument drifting correction process, (a) shows raw data, (b) and (c) magnification of the beginning and end of the data respectively from which mean values were taken to estimate a drifting rate, (d) and (e) show data corrected according to mean values (in red), (f) shows corrected data. This was than trimmed to to remove start and end atmospheric pressure data.

4.6 Evaluation of noise from a measuring device

The aim of this section is to explore noise in the collected high-speed data. Noise represents unpredictable variations in the signal which come from different sources, i.e., instrument, vibrations or interference. Such variations are observed over the entire recording period and have an impact on the quality of the data. Estimation and elimination of the instrument noise was undertaken to improve quality of the data and to aid the detection and characterisation of pressure transients. For a very noisy signal it is difficult to extract and then measure the properties of pressure transient waves in the time series data. It is particularly challenging to accurately measure its magnitude and duration. Such characteristic of pressure transients are an important part of the analysis carried out in this thesis.

The characterisation of noise was achieved by the determination of frequencies which could not be a part of a signal from a network and therefore should be removed. During this process a physical network characteristics were investigated. The analysis of power spectral density (PSD) of the signal helped the decision process. Removal of frequencies of the noise 'smooth' representation of a signal (see Figure 2.3 (b) in subsection 2.2.1) suitable for subsequent analysis.

4.6.1 Determination of cut off frequencies

This section presents initial evaluation of possible frequencies which can be removed from the signal based on a physical network characteristic of the water network and transient pressure wave oscillations. The behaviour of the pressure transient wave in a real system was investigated and the duration of one cycle/oscillation along the pipeline was determined.

A minimum length of the pipe, i.e. minimum distance from a source to primary reflection, was assumed to be 50 m. The length of pipes depends on a reflection point and not on the distance between subsequent joints; for metal pipes the distance between subsequent joints is usually about 1 to 2 m. The assumption of 50 m was based on examples of the real network taken from five different sites (Starczewska et al., 2015) for which the average minimum pipe length was calculated. This was also further verified from network maps and models.

For the determination of frequency of one complete oscillation for the wave to travel in a pipe twice to the point of reflection and back, the Δt was calculated. The calculation of Δt was done for two pipe materials, i.e. steel and plastic. Plastic (MDPE) wave speed was estimated as 295 m/s, time 0.68 sec. Steel wave speed was 1400 m/s, time 0.14 sec. In the case of a steel pipe, where *4L*

is equal to 200 m and wave speed c is equal to 1400 m/s (Thorley, 1991), section 2.1.2:

$$t = \frac{200}{1400} = 0.143(sec) = 6.9(Hz) \tag{4.1}$$

In a case of plastic:

$$t = \frac{200}{295} = 0.68(sec) = 1.47(Hz) \tag{4.2}$$

Based on these calculations 6.9 Hz is the fastest frequency of pressure transient that is likely to be of interest. This cut-off was further investigated on real pressure data.

4.6.2 Power spectral density (PSD)

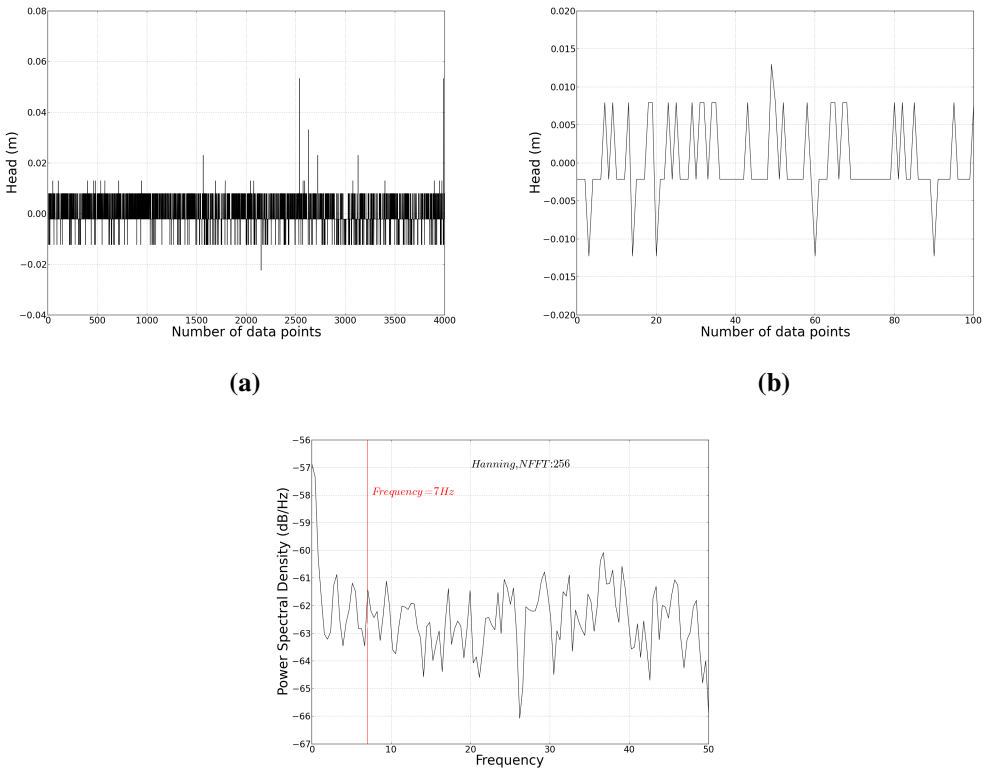
This section investigates if less than 7 Hz frequency cut-off (see section 4.6.1) can possibly remove some frequencies of noise from a measuring device. Power spectral density (PSD) was used to check a frequency response of the signals and to cross check a cut-off frequency of less than 7 Hz.

Figure 4.10 (a) shows data from the beginning of the file (see section 4.5.3) when the measuring device was not connected to the water network, e.g. recorded atmospheric pressure and noise. Figure 4.10 (b) evidences that the instrument has an analogue to digital conversion accuracy (the smallest change it can detect in the quantity that is measured) equal to 0.005 m. These random changes/variations in the signal are considered to be a 'white noise' or stationary signal (as described in subsection 2.2.1). If it is assumed that noise originate only from a measuring device and if it is ascertained that this is a 'white noise' then it has equal 'power' (i.e. flat spectral power density) over the range of frequencies. Figure 4.10 (c) shows the PSD of the recorded data. It shows that use of a frequency cut off of 7 Hz would allow to remove some parts of 'white noise' from the signal, however it could not verify whether lower frequency cut-off can be applied.

Figure 4.11 and Figure 4.12 show PSD applied to the data recorded when instrument was connected to the water network. It shows that use of a frequency cut off of 7 Hz allows the signal to have all the important information retained. It also shows that lower frequency cut-off could possibly be applied.

During the power spectral density analysis ranges of windows were investigated (Hanning, Bartlett, Blackman, Kaiser) from which a Hanning window was selected as this window type gave very similar results when compared to the others (e.g., Blackman and Kaiser). The width of the window of 256 gave the optimal results when tested on the data and was used for subsequent analysis on the data measured when instrument was measuring water network pressure.

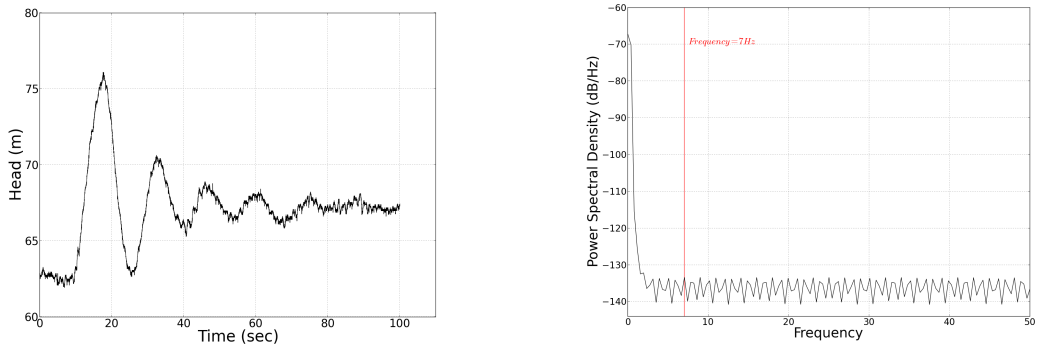
It has been confirmed that a frequency of 7 Hz allows retaining a portion of 'flat/even' signal power. This ensures that some 'white noise', related to the instrumentation, is removed but also



(a) **(b)**

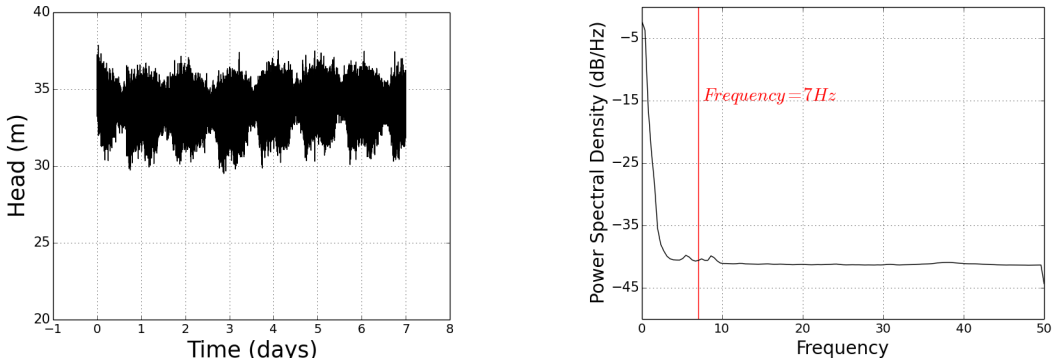
(c)

Figure 4.10: The sample of noise from the instrument, (a) data from the beginning of the dataset, (b) magnification of first 100 data points showing variations in pressure recordings; the minimum difference in high between points is 0.005 m, (c) application of Hanning window size of 256 to the data.



(a) **(b)**

Figure 4.11: PSD of real pressure transient wave, (a) pressure transient wave, (b) PSD of pressure transient wave.



(a) **(b)**

Figure 4.12: PSD of real pressure transient recorded for 7 days at 100 Hz, (a) pressure transient data, (b) PSD of (a).

that the important part of the signal is retained. Filter is, therefore required to apply the cut off of 7 Hz to the signal.

4.6.3 Filters

This section describes the application of a filter to verify the choice of 7 Hz frequency cut-off by comparing raw and denoised datasets. The aim of this analysis was to obtain a 'smooth' signal, without noise resulting from the sensitivity of measuring devices.

After determining which frequencies should be removed from the signal, a filter which effectively removes undesired frequencies was chosen. The Finite Impulse Response (FIR) filter was used with a cut off frequency of 7 Hz and 5 Hz transition width. The application of the filter to data can be seen in Figure 4.13 where raw and denoised signal are presented. The figure shows that instrumentation noise was removed and that care was taken to avoid impacting the small scale pressure fluctuations.

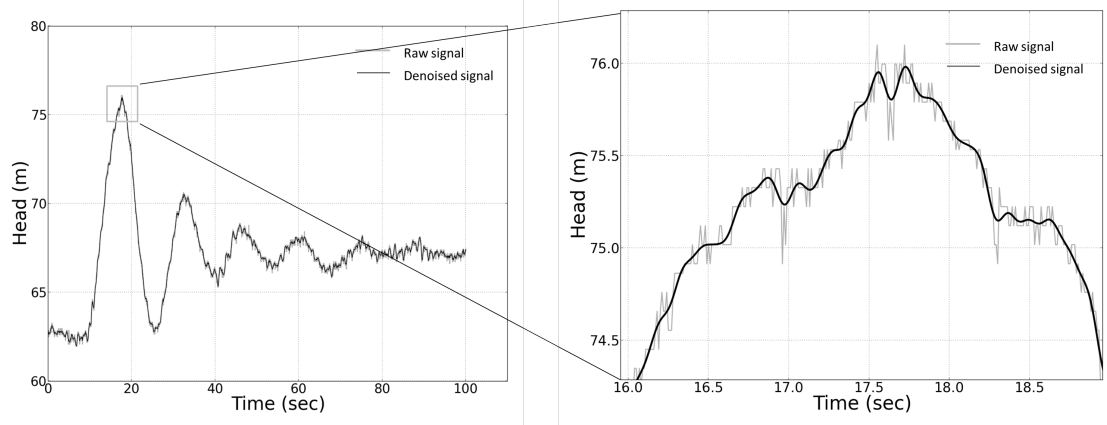


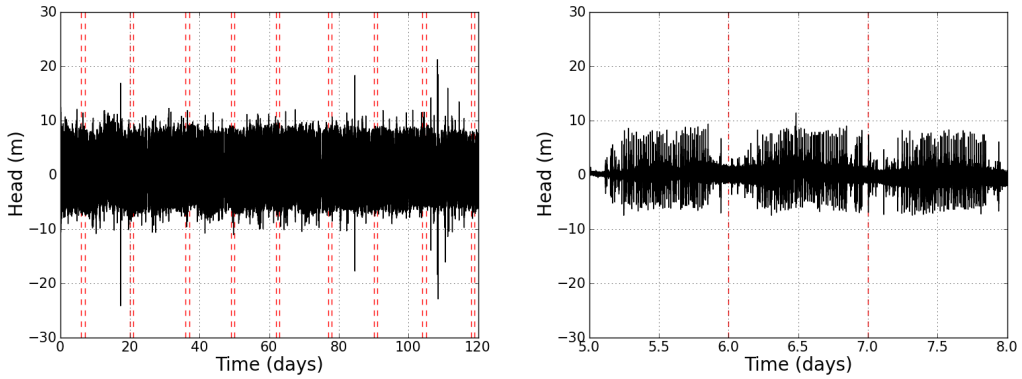
Figure 4.13: Example of a denoised pressure transient wave and magnification of its part.

4.7 Minimum required length of the data

The aim of this section is to determine the shortest possible recording time period that provides sufficient representation of pressure transient behaviour evident from a single point of measurements (i.e. one location). This was identified using a stationarity check as defined in subsection 2.2.2 on the data collected from one location over a long time period. This was done to enable down-sampling to identify minimum period to achieve a representative sample of what a pipe experienced.

The location for long-term monitoring was selected to ensure day-night pressure transient variations. Pressure data was recorded near the industrial user to ascertain that information about pressure transient activity would be captured. This decision was based on initial visual assessment of the pressure traces from diverse sites which revealed repeatable pressure transient patterns (diurnal patterns) visible during weekdays working hours and being less present during weekends.

Data was recorded continuously for a duration of 4 months, an equivalent to 1 x 10^9 (1036800000 data points) measured using a 100 Hz resolution. This data comprised of 10 files from 10 different instruments assembled into one 40GB file. To each dataset the calibration and correction of data for logger drift was applied. The compiled data can be seen in Figure 4.14. This represents detrended data due to the means per logger being removed.



(a) **(b)**

Figure 4.14: Dataset (down-sampled to 1 Hz for plotting purposes) with means per logger substracted prior to assembly (a) four month long dataset, the red dashed vertical lines indicate a day (midnight to midnight) of logger exchange, (b) zoom to a typical pressure transients experienced in a day.

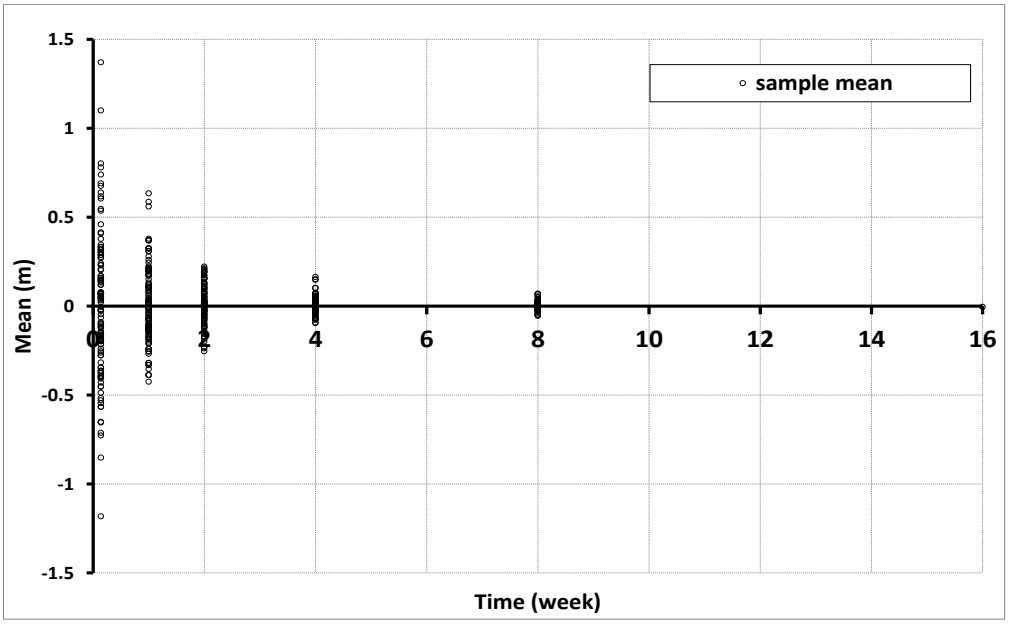
4.7.1 Stationarity of the data

To check the stationary properties of 4 months data recording time lengths the data was divided into sections (representing different span lengths) with the longest being a full dataset (4 months period = 16 weeks). The longest section was divided into two months period = 8 weeks, then into one month = 4 weeks, two weeks = 2 weeks, one week long set = 1 week, and one day = 0.14

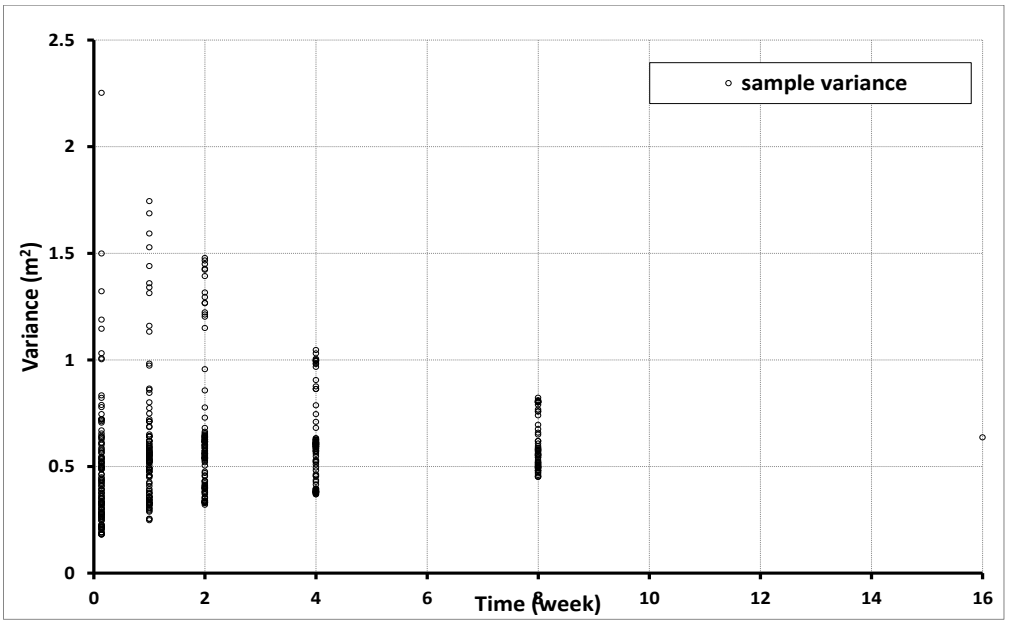
week. For each of these windows (i.e. 8 week, 4 week, 2 week, 1 week and one day) estimates of means and variances were calculated by moving each window by one day across the data. The results can be seen in Figure 4.15 (a) and (b).

Figure 4.15 (a) shows that estimated means of the data stabilised to the single value indicating a stationarity of the data. The Figure 4.15 (b) shows that the spread of the estimated variances also decreased. As expected, a single day shows the highest variance (Figure 4.15 (b)) in comparison with longer spans. It indicates that continuous recording for one day is not long enough to accurately capture repeating pressure transient events. Each day is slightly dissimilar as can be seen in Figure 4.14 (b). Increasing the recording period to 1 week caused the differences between days to be less apparent. This also suggests that pressure transient variability can possibly be lost in a volume of data. Two and four weeks recording period were considered to provide enough pressure transient variabilities.

It has been observed that in the majority of sites pressure transients occur regularly indicating some degree of stationarity and regular behaviour. Therefore, it was decided that a minimum data collection period should not be less than one week to capture all pressure transient behaviour as standard weekdays are expected to be different from weekends. The instrument battery pack allowed the collection of the data for a maximum period of two weeks. All locations were, therefore, monitored for this maximum possible time length (two weeks).



(a)



(b)

Figure 4.15: The comparison of the estimated means and variances obtained for one day, one week, two weeks, one month, two months and fourth months based on the data compiled with subtracted means, (a) estimated means for different lengths of data, (b) estimated variances for different lengths of data.

4.8 Preliminary results

The collected data provided unique details of pressure transient activity and behaviour in real WDS in terms of the duration it run for and the areas over which the pressure transient data was collected. Based on the size and duration of the dataset and the number of sample points this is a valuable data set showing, for the first time, the wide scale occurrence of pressure transients in water systems. This can also be seen as a proof of concept for the technology and techniques employed in the field across such a wide area and for the analysis of the large quantities of data produced by the monitoring devices. Because such extensive monitoring in a real network has not been undertaken previously this data novelty was utilized in this research.

Throughout the data collection programme, transient activity was found in the majority of the sites monitored. It has been noticed that pressure transients events regularly occurre in water distribution systems. Four types of pressure transient occurrences were identified:

- random,
- regularly occurring during weekdays but not on weekends,
- regularly occurring during daytime but not night-time,
- regularly occurring irrespective of day- or night-time.

In residential sites there was evidence of random, infrequent and occasionally occurring pressure transients (typically one significant transient event of a magnitude above 5 m per week) and the source of those are not currently known. In these sites, diurnal pressure responses and daily patterns due to domestic consumption were commonly seen. The typical pressure trace from the residential site can be seen in Figure 4.16 (a).

The industrial sites had regular pressure transients seen during weekdays but not on weekends, see Figure 4.16 (b) and (c). There is also a clear break in pressure transient activity observed during a national holiday (days 11,12,13) in Figure 4.16 (b). The regular occurrence of pressure transient in industrial sites were also observed during the daytime but not at night-time, see Figure 4.16 (d). In this site transient activities were observed typically from 7:00 until 22:00 and then significantly decreased from 22:00 to 7:00. The regular occurrence of these pressure transients in industrial sites confirms their association with commercial users activities taking their water supply from the network at nearby locations.

Sites associated with pumps exhibited regularly occurring pressure transient unrelated to weekdays or weekends as can be seen in Figure 4.16 (e) and (f). Sometimes those were observed to occur

from up to 20 times per day Figure 4.16 (e) to one per day (as presented in Figure 4.16 (f)), possibly due to the opening and closing of valves related to a pump.

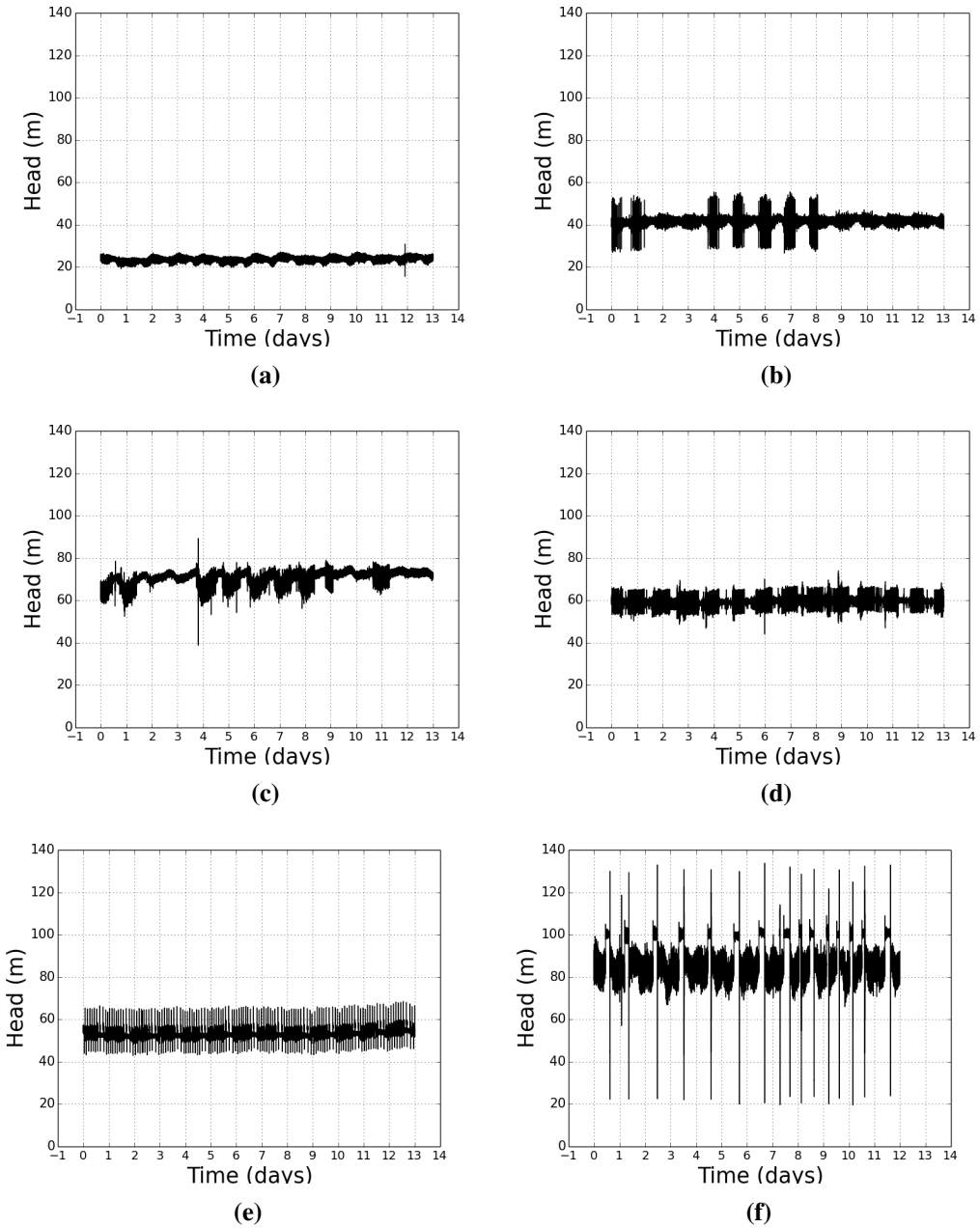


Figure 4.16: Example pressure traces recorded in (a) residential site, (b) industrial site, (c) industrial site, (d) industrial site, (e) pump site, (f) pump site.

Shortly after the preliminary monitoring program was accomplished it became evident that pressure transients can be also found in the areas where their presence was not previously expected. The scale of the activity which was seen varied with location and ranged from, low magnitude (10 m) oscillations to large scale (100 m) transients. In almost all sites, a low magnitude, but high-frequency oscillation was found (5-10 m, 100 times an hour). This oscillation was typically superimposed on top of the normal diurnal pressure variations due to domestic consumption.

The successful data collection program provided the hard evidence of pressure transients collected from 67 sites. The data set is unique providing extensive detail of transient activity and behaviour in real water networks. This can also be seen as a proof of concept for the technology and techniques employed in the field across such a wide area and for the analysis of the large quantities of data produced by the monitoring devices (60 x 10^6 data points/week). Such extensive monitoring in a real network has not been undertaken previously, therefore, this data is valuable and its novelty is utilised in this research to investigate what a given pipe experiences at a specific location.

4.9 Summary

The survey of WDS networks was undertaken for the collection of pressure transient time series data. A high-speed (100 Hz) continuous data collection for a duration of two weeks from at least two points per each site was identified as a minimum requirement and undertaken.

To ensure the collection of a wide range of pressure transients, high-speed monitoring successfully covered sources historically associated with pressure transients and sites not historically associated with transients. These were in relatively close geographic locations to facilitate 'milk round' data collection process. Two monitoring locations per site required the supply of 40 high-speed pressure data acquisition instruments. The instrument precision was determined and the process of reduction of the instrument noise was conducted.

The assessment of pressure transient occurrence in monitored sites showed a strong evidence confirming their wide spread existence in such systems. Some pressure transient activity was seen in all sites studied. This formed the basis for investigations on the contribution of pressure transients to the observed pipe repair data. Monitoring of water networks showed strong evidence of regular pressure transient behaviour. It was seen that occurrences of pressure transients were highly dependant on the character of the monitored site. Preliminary results confirmed that some of these events are specific to industrial (majority of regularly occurring pressure transients) or residential sites. Investigation of the collected data showed that pressure transients were seen in almost every monitored location (see Appendix A). Furthermore, differences in pressure transients sizes and shapes were observed. Analysis of the pressure transient characteristics will investigate the contribution of regular and irregular pressure transients on the observed pipe repairs data.

Chapter 5

Transient data analysis

5.1 Introduction

The field methodology developed during this research allowed the collection of time series pressure transient data from a diverse range of operational WDS. Interpreting this data into a meaningful understanding with respect to the contribution to the network repair records, however, presented a challenge. One approach to evaluate the contribution of pressure transients is through a descriptive modelling. To achieve this a method is required to 'convert' the time series data into a form which characterises pressure transient events captured during the recording period.

This chapter presents a method which characterises and quantifies pressure transients selected from time series data in terms of discrete events. Detection techniques were developed for the selection of these discrete events and subsequent pressure transient event characterisation by magnitude, duration and number of occurrences. Goodness of fit measures were explored to reduce the number of unsatisfactory events and enhance the characterisation of pressure transient events. The result is a method which yields a transient fingerprint for each time series recorded in WDS.

5.2 Problem definition and conceptual design

To assess the contribution of pressure transients to pipe repairs a method to characterise pressure transient events is required. It should reflect characteristics which may subsequently be correlated with a contribution of pressure transients to the observed repair data; however it is currently unclear which characteristics are important. The method of characterisation should also be able to represent the sufficient range of oscillations in pressure that may contribute to pipe repairs.

Figure 5.1 captures some previous measures of pressure transients (see section 2.2.3) represented on a real world isolated example of pressure transient wave. These focus on values of extreme

pressures, pressure range (maximum observed pressure - minimum observed pressure) or pressure variation ranges (e.g., 0-5m, 5-10m, 10-15m) and also rate of change of pressure.

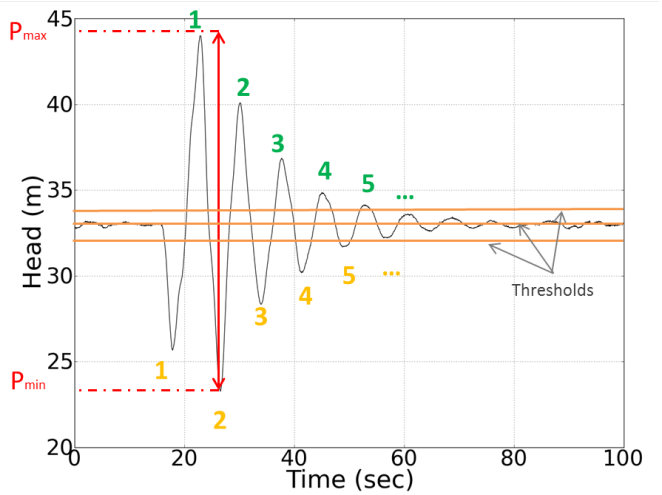


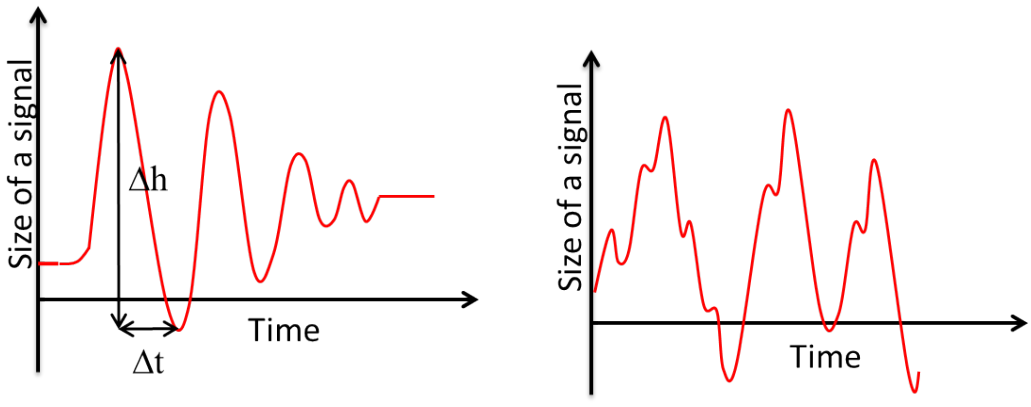
Figure 5.1: What is it about pressure transient that contributes to the observed pipe repairs? Plot showing some of possible interpretations related to e.g., extreme pressure, pressure range ($P_{max} - P_{min}$), pressure variation ranges and number of positive versus negative events with applied thresholds.

These methods, however, are not able to sufficiently describe the contribution of pressure transients to the observed pipe repairs as there is still unknown which characteristics of pressure transients are important and dominant. For example, it is rare to capture one extreme pressure transient which causes a pipe to fail. As none of such events occurred over two weeks of continuous recording utilised in this research (see section 4.8), therefore it may be argued whether a single maximum (or minimum) value would show any correlation with pipe repairs. Pressure range or pressure variation ranges omit the fact that typical transient waves are comprised of a number of oscillations following an initial pressure surge. In addition, pressure variation ranges do not characterise pressure transients accurately e.g., do not precisely identify discrete pressure transient events. In addition, a number of pressure transient oscillations (number of upsurges versus number of downsurges) may also have an effect on the network if their cumulative effect on the pipe is considered. This is significant when considering that it is currently unknown whether it is extreme transient events that effect complex networks and cause pipe failures or if long term, fatigue loading (see section 2.2.3) of repeatedly occurring pressure transients induces a greater risk. Therefore, magnitudes of pressure transients and the number of their occurrences are important factors in the assessment of their contribution to observed repair records. An important characteristic which has previously been omitted in pressure transient characterisations is the duration, i.e. length of an entire pressure transient wave. An example can be seen in Figure 5.2 (a) where the duration (Δt) is between peak and trough. This characteristic has not been considered despite the fact that greater risk of damage to a pipe may be associated with more sudden (i.e. short duration events

of large magnitudes) pressure transient events. Large events happening over a shorter time may have different impacts on, and pose different risks to, the network than the same magnitude events happening over a longer period. For these reasons time may be an essential factor to consider when analysing the contribution of pressure transients to the observed repair records. Consequently, a high potential for damage to a pipe would be a combination of pressure transient magnitude, and duration, with more significant damage coming from high magnitude and short duration events (not only to events of large magnitudes).

Consequently, it has been decided that to improve the inference of possible contribution to observed repair records associated with pressure transients, any method developed should consider the following three components: the magnitude of pressure transient events, their duration and the number of occurrences. In this research a method is presented to characterise pressure transients in terms of these criteria based on samples of time series data previously collected.

Identification of these characteristics can be problematic, especially when real pressure data is considered. An idealistic example of a transient wave comprising several oscillations can be seen in Figure 5.2 (a). For this kind of pressure transient wave it may be relatively straight forward to capture the magnitude, duration and the number of their occurrences. In real networks, however, such idealistic transients are rarely observed. Due to reflections from water network features (e.g., valves, changes in pipe diameters), the shape of transient waves often take the form as presented in Figure 5.2 (b). In this figure a number of small magnitude oscillations impact on the transient wave. Such oscillations cannot be disregarded, therefore, in this research, the small magnitude changes are considered as of equal potential interest as the events of large magnitudes in terms of their contribution to the pipe repair records. This illustrates that some form of enhanced methodology (technique) is needed to describe the range of pressure transient events in terms of the three characteristics identified.



(a) **(b)**

Figure 5.2: Conceptual examples of pressure transients (a) ideal, (b) real network pressure transient.

5.3 Evaluation of time series characterisation methods

There is currently no established method to characterise different pressure transient events from recorded time series signals, as highlighted in section 2.2.3. While it is relatively easy to visually identify pressure transients in a small time series dataset, it becomes increasingly difficult and subjective with data comprising millions of data points. For instance, one hour of data recorded at 100 Hz resolution has 360,000 data points, whereas a one day time series is equivalent to 8,640,000 data points. With the typical recording of pressure data over a period of two weeks ($1.2x10^8$ data points) used in this research (section 4.7.1), visual assessment of this large data does not show the fine detail required to observe pressure transient waves. Visual assessment of pressure traces also does not allow for quantitative analysis of pressure waves seen in the pressure trace. Consequently an important criterion for a characterisation method is the ability to effectively process extremely large datasets. To aid the pressure transient characterisation process, criteria for assessing the possible methods were developed. These are shown in Table 5.1.

Table 5.1: Criteria identified for evaluating time series characterisation methods.

Label	Criteria
a	The method should reduce time series data into a number of 'metrics' /characteristics that can be an input into a descriptive model.
b	The method should characterise the pressure transient regime by identification of magnitude, duration and number of occurrences.
c	The method should capture 'all' range of changes in pressure ('all' possible magnitudes and their durations) to show what a pipe actually experiences. This information will allow the assessment of a contribution to observed repair records and will help determine transient characteristics associated with such contribution.
d	The method should represent the original signal (e.g., pressure transients) as accurately as possible.
e	The method should work effectively on large datasets.

5.3.1 Frequency domain methods

In principle, pressure transient events are discontinuous series. For this reason the majority of frequency domain methods are unsuitable to characterise pressure transients from time series data. They cannot be used to obtain the three characteristics of pressure transient events identified previously (Table 5.1 (b) and (c)). Therefore, the Fourier Transform (FT) method is not appropriate for a series of discontinuous events and does not provide actual magnitudes or duration of pressure transients. As a result, the method is not able to capture the useful characteristics of the pressures experienced by the pipe despite the fact that it reduces time series data (Table 5.1 (a)) and works

effectively on large datasets (Table 5.1 (e)). Magnitudes and durations can provide additional information when drawing conclusions about the contribution to the observed pipe repair from the descriptive model.

The use of frequency domain methods to characterise pressure transients in a different form than magnitudes and durations does exist e.g., power spectrum density (PSD). It however, do not convey identified physical meaning and interpretations. This is because there is a limited association between pressure transient characteristics and network repair records.

Some other frequency domain methods, such as wavelet analyses, are capable of dealing with pressure transient waves as a series of discontinuous events. Wavelet analysis has been successfully applied in transient analysis for peak, leak and burst detection as highlighted in section 2.2.4. However, these methods require high recording frequency (Srirangarajan et al., 2012). This method may not be effective for a large databases. Furthermore, the wavelets do not reduce time series data (Table 5.1 (a)) and also do not provide clear counts or magnitudes of pressure transient events (Table 5.1 (b) and (c)).

5.3.2 Time domain methods

Continuously recorded time series can be assessed and characterised using time domain methods. Different forms of representations of measured signal were investigated to accurately characterise the time series pressure data and aid their assessment.

5.3.2.1 Histogram of pressure

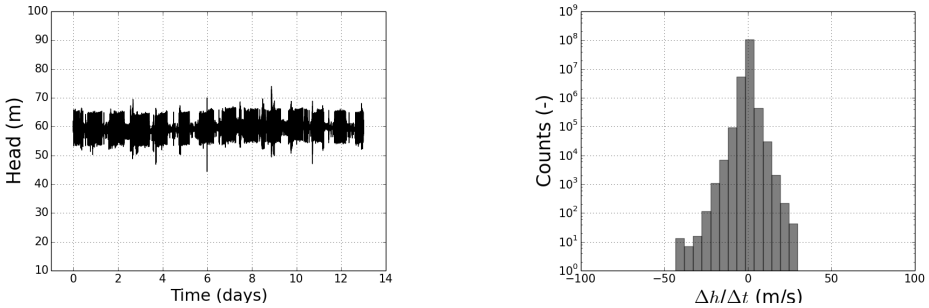
The histogram of continuously recorded time series pressure data reduces the original time series to a series of bins (Table 5.1 (a)). The histogram of raw time-series pressure data, however, does not provide counts of magnitudes nor counts of durations and therefore does not meet the criterion (b) and (c) in Table 5.1. The use of a histogram on continuously recorded time series pressure data is limited, and would require preprocessing of data, as it has been found that diurnal patterns mask some pressure transients. Pressure transients are more sudden (Δ h/Δ t) than the much slower changes of pressure observed through the day or between night and day.

5.3.2.2 Histogram of the rate of change of pressure

Time series data can be represented as the rate of change of pressure with respect to time, i.e. $\Delta h/\Delta t$ between sequential data points. This method reveals how fast pressure changes occur in time series data and, therefore, may provide some insights into the characteristics of pressure transients. To provide a reduction in data dimensionality (Table 5.1 (a)) data can be represented and assessed in

a histogram.

Figure 5.3 shows time series data and a histogram of the rate of change of pressure as a method to characterise pressure transients. Figure 5.3 (b) has a logarithmic y-axis due to the large number of small scale fluctuations.



(a) **(b)**

Figure 5.3: Application of the rate of change of pressure to pressure data, (a) time series pressure trace; (b) histogram of the rate of change of pressure in a logarithmic scale due to the large number of small fluctuations dominating larger changes in pressure.

The histogram of the rate of change of pressure does not, however, distinguish between some events as shown in Figure 5.4. The examples in Figure 5.4 show that the rate of change of pressure of various signals (Figure 5.4 (a), (c), (e), (g)) return the same number of events (Figure 5.4 (b), (d), (f), (h)). In this example two sizes of events are always obtained (three counts of each size) despite different samples of pressure traces being analysed. As a result, this method is insufficient to show the number of transient events that are present in the recorded signal and their real magnitudes or durations (Table 5.1 (b)). It therefore does not represent the signal as accurately as possible, (Table 5.1 (d)). However, it works effectively on large datasets (Table 5.1 (e)) is simple and fast and meet the most of the other criteria. It also provide a description of a variations in a signal, which may be ultimately useful to explore pressure transients contribution to observed pipe repair data.

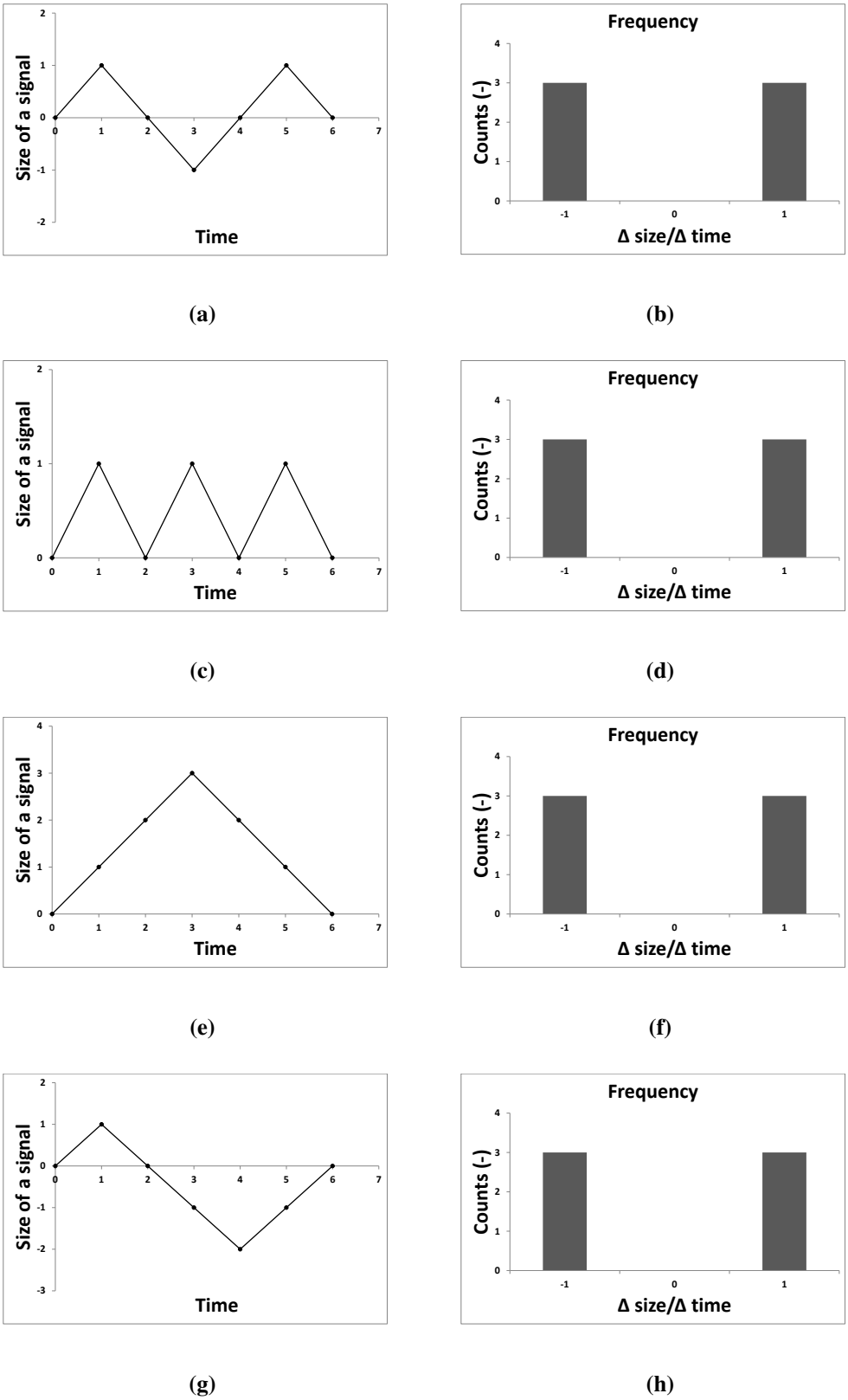


Figure 5.4: Example of a rate of change of pressure with respect to time of simple signals; (a) (c) (e) (g) examples of a signal, (b) (d) (f) (h) number of occurrences of the events observed after applying the method.

5.3.2.3 S-N curves from cycle counting in fatigue analysis

Irregular and quite often long load parameters (e.g. force, stress, strain, acceleration) are often characterised by providing a number of cycles of various sizes occurring in the form of a histogram of cycles known as S-N curve (section 2.3.2). This method can reduce data to counts of cycles thus meeting criterion (a) (Table 5.1). The cycle can be used to describe stress from pressure fluctuations which may induce or aggregate the development of cracks and failures. To form a histogram, the cycles are firstly identified using different methods, e.g. level-crossing, simple-range counting, range-pair counting, peak counting and rainflow counting (section 2.3.2).

The commonly used method in fatigue analysis is a level-crossing counting. The use of this method to obtain the three characteristics of pressure transients has some limitations. The method does not identify all peaks and troughs and does not allow for the identification of the unique peak-trough pairs (cycles). In the process of the identification of peaks and troughs the method relies on the estimation of the 'levels' of the load. Therefore, it does not account for 'all' magnitudes and durations of transient pressure changes (Table 5.1 (c)). It would be challenging to extract the high magnitude events if there were lots of small events superimposed on the pressure change.

Simple-range counting or range-pair counting (section 2.3.2) allows for identification of the unique peak-trough points. This method, however, relies on the estimation of the ranges of the load and does not account for all possible pressure changes. It would be difficult to extract the high magnitude events if there were small events between a peak and trough.

Peak counting cycle counting method (section 2.3.2), can identify these unique points. In this method peaks above and troughs below the reference load level (usually the mean) are counted. Sometimes a variation of this method is used, which is based on counting all peaks and valleys without regard to the reference load. This method does not account for a full range of oscillations in pressure. Therefore, it is not suitable for the characterisation of pressure transients from time series data.

A method in fatigue cycle counting (applied on already detected peaks and troughs) is rainflow counting (2.3.2). For use in pressure transient characterisation this method shows some benefits as it captures range of different magnitudes of pressure transients. However, rainflow calculations requires long computational time. Furthermore, in its pure form the method does not directly account for a time component. For this, a modification of the algorithm (or development of a new one) and further testing would be required.

5.3.2.4 Histogram of sequential peak-trough pairs

The use of algorithms for peak-trough detection allows for automatized detection of these points in any given time series data. These points can then be subsequently paired and represented in a histogram. The magnitude and duration can easily be acquired from the pair (Table 5.1 (b)) and histograms provides further characterisation (Table 5.1 (a)). However, a solution is required to account for 'all' possible magnitudes and durations (Table 5.1 (c)). Histogram of sequential peak-trough pairs are the simplest method of discrete input to the histograms, easy and fast to implement on large datasets (Table 5.1 (e)) and are easily applicable in standard programming languages.This method is also able to accurately represent the signal (Table 5.1 (d)). The preferred method should be able to identify peaks and troughs to successively acquire 'all' magnitudes and durations for discrete pressure transient events.

5.3.3 Summary

The assessment of the possible methods for characterisation of pressure time series data is presented in Table 5.2. It shows that none of the available methods outlined above satisfy the defined criteria. However, counting sequential events by identification their peaks and troughs has the potential to identify discrete events and to capture not only magnitudes, but also durations and number of their occurrences. A solution is, therefore, required to account for 'all' magnitudes, durations and counts of discrete pressure transient events. Possible methods for peak and trough detection were, therefore, investigated further.

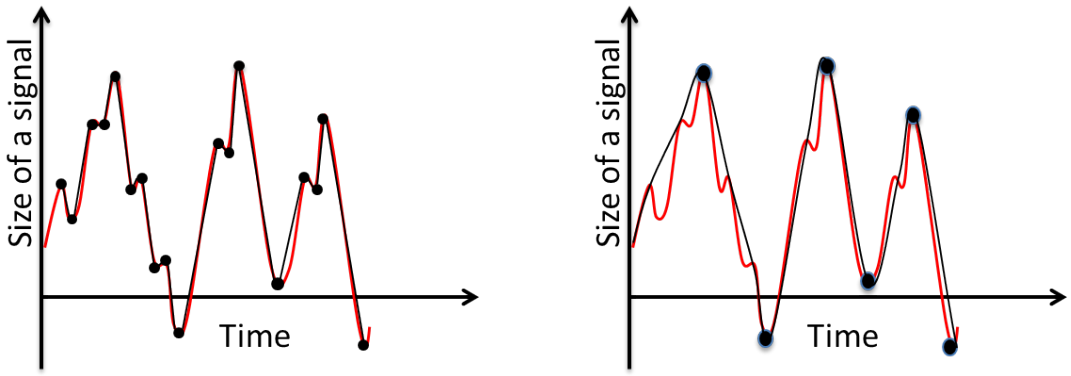
Table 5.2: Evaluation of the pressure transient characterisation methods against criteria.

Criterion	**Methods**					
	Frequency domain		**Time domain**			
	FT	Wavelets	Histogram of pressure	Histogram of $\Delta h/\Delta t$	S-N curves	Histogram of succesive events
(a)	√	–	√	√	√	√
(b)	–	–	–	–	–	√
(c)	–	–	–	–	–	√ (?)
(d)	√	√	–	–	√	√
(e)	√	–	√	√	√	√

5.4 Methods for peak and trough detection

The correct identification of peaks and troughs is essential to form unique peak-trough pairs as presented in Figure 5.5. The selection of these pairs allows the capture of not only magnitudes of pressure transients but also their duration and subsequently the number of occurrences. However a potential method needs to be evaluated whether it can get beyond successive pair of peaks and

troughs. The evaluation of potential methods for peak and trough detection is presented in a following section.



(a) **(b)**

Figure 5.5: Conceptual identification of a real/large pressure transients; (a) issues due to small magnitude transients, (b) a solution for the capture of large transients.

5.4.1 Zero-crossing (first derivative)

This method detects all peaks and troughs based on differences in the change of the slope sign (see section 2.2.4 (a)). It detects local maxima and minima in the time series data, as described in Figure 5.5 (a). The sequential paring of the detected peaks and troughs, however, is a limitation. After pairing points, this method would not detect large magnitude pairs as seen in Figure 5.5 (b). Therefore, it is not suitable method.

5.4.2 Peak and trough detection algorithm

The peak and trough detection algorithm (MATLAB script: `http://billauer.co.il/peakdet.html`) was found to be suitable to detect peaks and troughs in the time series data. The algorithm detects peaks and troughs using incremented values in a 'delta h' threshold. To detect a peak the algorithm looks for the highest point around which there are points lower by the size of a 'delta h' threshold on both sides. What makes a peak is the fact that there are lower (by a size of a 'delta h' threshold) points around it. The use of a 'delta h' threshold by the algorithm is an advantage which captures large pairs as identified in Figure 5.5 (b).

5.4.3 Summary

Two methods were evaluated for the detection of peaks and troughs: zero-crossing and the peak and trough detection algorithm. The peak and trough detection algorithm was identified as a potential method to successively pair peaks and troughs. As coded, the algorithm is only able to identify peaks and troughs, likewise zero-crossing method. Therefore, both methods could be used

to successively pair peaks and troughs. However, the algorithm uses the size of a 'delta h' threshold to successively increase the threshold in defined steps. For this reason, it has a potential to apply different sizes of 'delta h' thresholds to find 'all' magnitudes and durations of peak-trough pairs. This was identified as an advantage to go beyond sequential peak-trough pairs only, to achieve the result as identified in Figure 5.5 (b). An enhancement of this method is required to firstly successively pair peaks and troughs into unique pair of points and the evaluation of a 'delta h' threshold.

5.5 Enhancement of the preferred method

This section details how the standard peak-trough detection algorithm (identified in section 5.4.2) works. An enhancement of the algorithm was developed and applied to successively connect the identified peaks and troughs into unique pairs of points to go beyond successive pairs of peaks and troughs.

To demonstrate the standard peak-trough detection algorithm, simulated data comprising of six points was used. The process of identification of 'all' possible peaks and troughs based on the initial size of the 'delta h' threshold is shown in Figure 5.6 (a). Figure 5.7 (a) shows the effect of increasing the size of the 'delta h' threshold on the detection of peaks and troughs.

The standard peak and trough detection algorithm was enhanced to include the following steps:

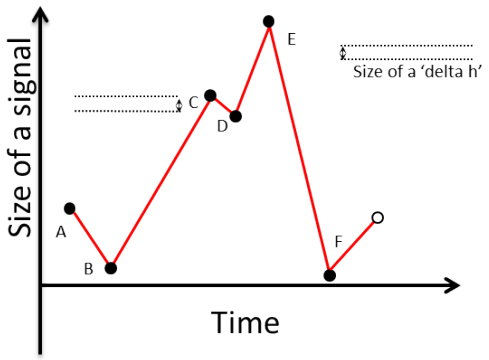
- peaks and troughs are subsequently connected to form peak-trough pairs. Their characteristics (i.e., magnitude and duration) are calculated and saved in the outcome list produced at the end of the algorithm run, as described in Figure 5.6 (b).

- repeat of the process with increased a 'delta h' threshold; each increment in a 'delta h' threshold produces the outcome list, as described in Figure 5.7 (b),

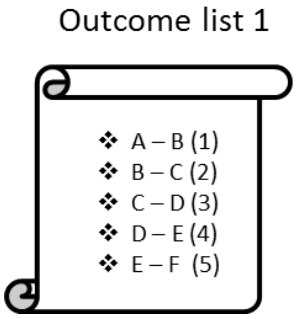
- each outcome list (see Figure 5.6 (b) and Figure 5.7 (b)) is compared to remove duplicates and subsequently form the final outcome list, as described in Figure 5.8.

The enhanced peak and trough detection algorithm therefore includes the implementation of the outcome list consisting of peak-trough pairs with magnitudes and durations, repeat with increasing threshold and defining unique list.

Figure 5.8 shows that some pairs may appear to be counted twice, e.g., pairs B-C (2) and D-E (4) can be considered as part of B-E (6). For this research it is proposed to consider all of these events as equally valid. Accounting for all possible pairs allows for exploration of the contribution of all and any sizes of pairs on the pipe repairs.

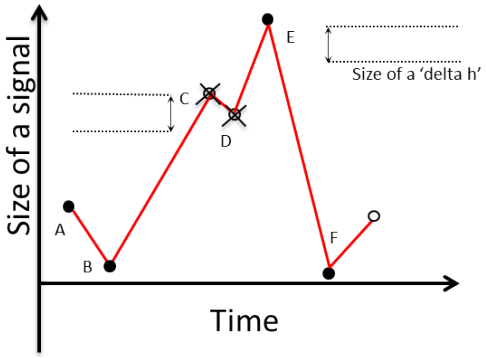


(a) Standard algorithm: the black circles show points which were identified and returned by the algorithm according to the initial size of a 'delta h' threshold. Identified peaks are A, C, E and troughs B, D, F respectively.

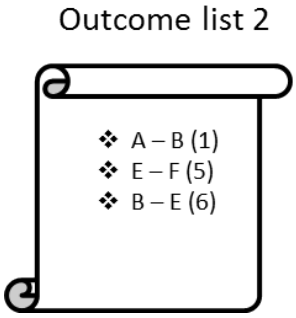


(b) Enhancement of the algorithm: the outcome list 1 is compiled at the end of the algorithm run which comprises successively connected pairs of points: peak – trough, trough – peak. Five pairs can be identified in this example.

Figure 5.6: Visual description of the application of the standard algorithm and its enhancement for the initial pressure transient detection.



(a) Standard algorithm: the size of the 'delta h' threshold is increased. The algorithm identified peaks: A, E and troughs: B, F respectively.



(b) Enhancement of the algorithm: The outcome list 2 is compiled and comprises pairs of points: peak – trough, trough – peak saved during the second run of the algorithm. These pairs can be identified in this example.

Figure 5.7: Visual description of the application of the standard algorithm and its enhancement after increasing size of 'delta h' threshold.

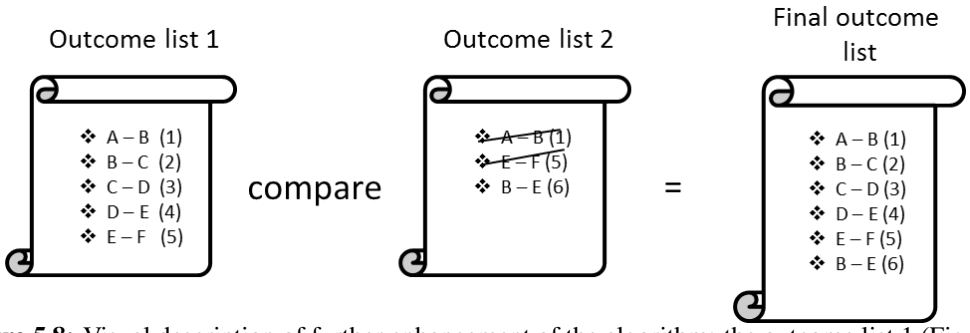


Figure 5.8: Visual description of further enhancement of the algorithm: the outcome list 1 (Figure 5.6 (b)) and the outcome list 2 (Figure 5.7 (b)) are compared. The unique results from both lists are retained, duplicates are removed. The final outcome list is compiled. It comprises of unique pair of points.

5.5.1 Test application of the algorithm

Figure 5.9 shows a simple data sample that will be used to evidence the application of the enhanced algorithm. In this figure black dots represent peaks and red dots troughs.

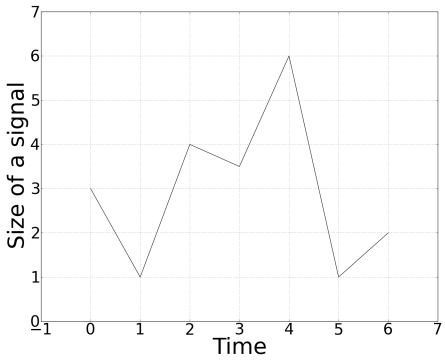
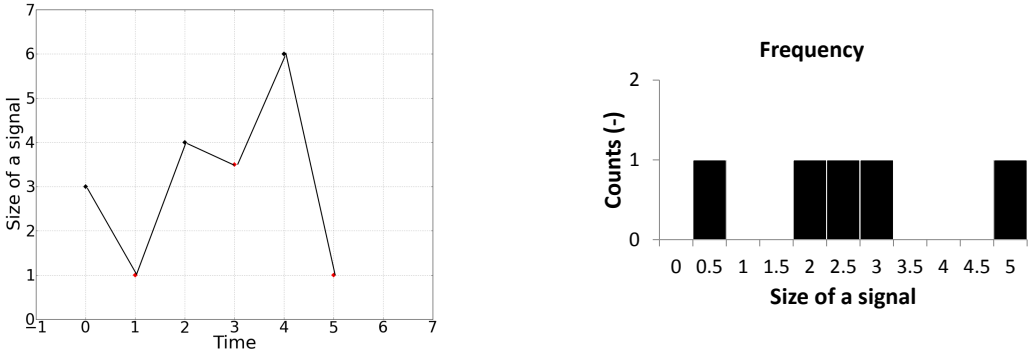


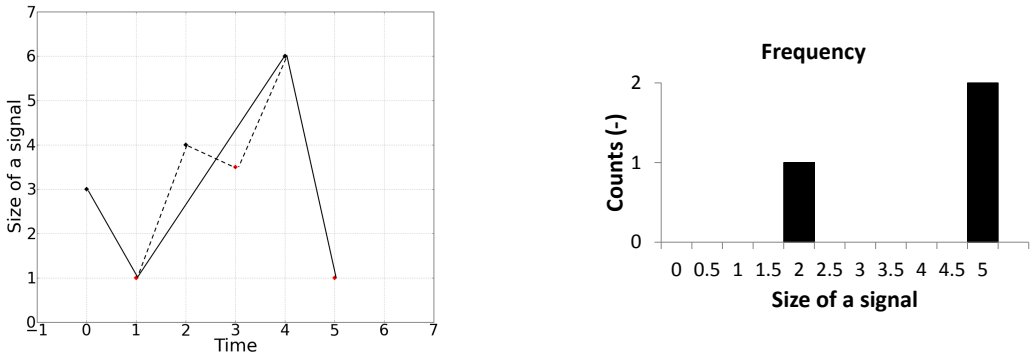
Figure 5.9: A signal consisting of seven points.

The application of increments in a size of 'delta h' threshold are presented in Figures 5.10 and Figure 5.11. The histogram shows sizes of peak-trough pairs created at the end of each run of the algorithm (see Figures 5.10 (b) and Figure 5.11 (b)).



(a) **(b)**

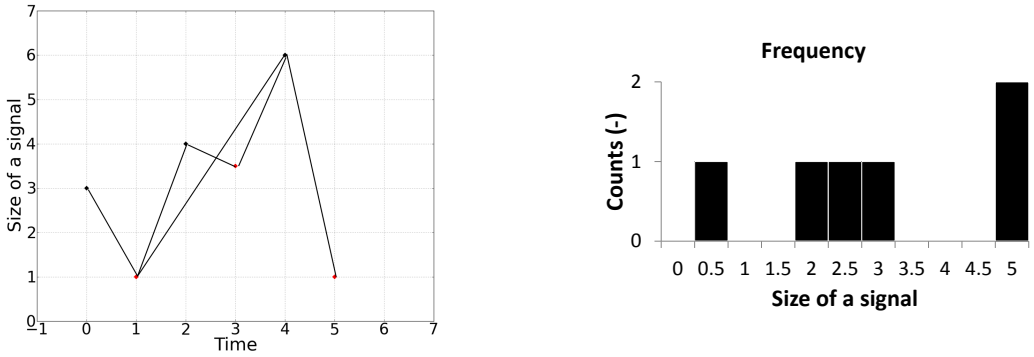
Figure 5.10: Initial identification of peaks and troughs whose magnitudes are greater than the size of the 'delta h' threshold value = 0.1 size of a signal. All pairs are detected and compiled in the output.



(a) **(b)**

Figure 5.11: Increment in the size of a 'delta h' threshold value = 1 size of a signal results in one magnitude being removed from the signal (value lower than the threshold). The new magnitude can now be detected. The final outcome shows one additional/new event and two events previously detected (see Figure 5.10). The new event will be added to the outcome list.

The final outcome, after cleaning for unique pairs only, of the enhanced algorithm is presented in Figure 5.12. In total, six peak-trough pairs of different sizes were identified in this example.



(a) **(b)**

Figure 5.12: A final result shows six peak-trough pairs in the final outcome. To achieve this two databases with pairs detected at each stage of the algorithm run (Figures 5.10 and Figure 5.11) were compared to remove repeating pairs and add any new ones.

5.5.2 Practical application of the algorithm

To evidence practical application of the advanced algorithm, two real sets of pressure data are presented. Care was taken to choose examples from real water distribution networks representing different shapes of pressure transients and, therefore, sufficient complexity to evidence practical application. Before the application of the algorithm the initial size of 'delta h' threshold and the subsequent increment threshold size were evaluated.

5.5.2.1 Initial threshold size

The initial size of 'delta h' threshold for the detection of all peaks and troughs was chosen to be 0.005 m. This value is the analogue to digital conversion accuracy, and also the measurement resolution of the instruments (see section 4.6). The choice was then verified by applying various initial threshold values to different datasets. It was concluded that to detect large magnitude pressure transients from complex, real signals the choice of the smallest possible value was necessary to identify 'all' possible peaks and troughs and successively pair them. As an example, one dataset shows the impact of different initial sizes of 'delta h' thresholds on the peaks and trough detection process (see Figure 5.13).

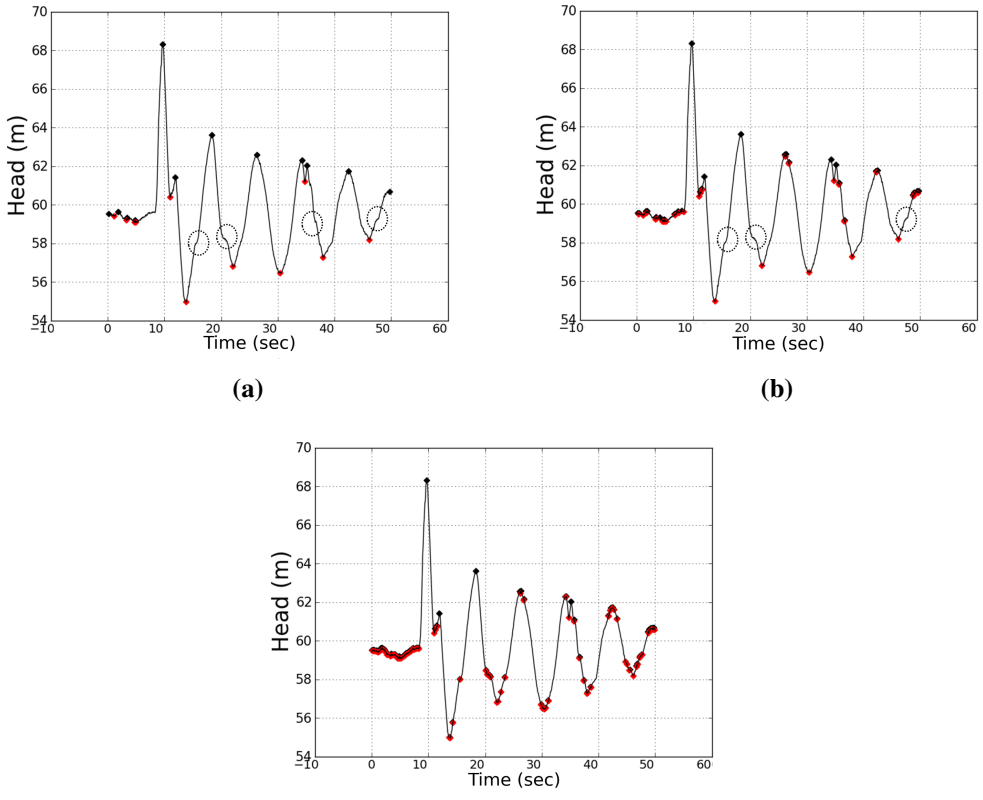


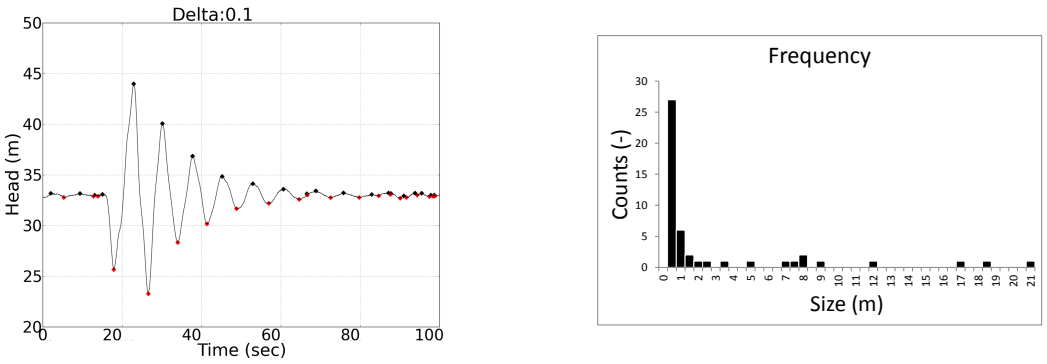
Figure 5.13: Example of application of different sizes of the initial 'delta h' threshold, peaks (black dots) and troughs (red dots) identified, (a) size of 'delta h' threshold = 0.01 m (b) size of 'delta h' threshold = 0.05 m (c) size of 'delta h' threshold = 0.005 m. Circles highlight points of potential interest that are not detected.

After the detection of all possible peak and trough pairs (using 0.005 m size of the 'delta h' threshold), only the pairs which had a magnitude equal or greater than 0.1 m were retained. This size represented two standard deviations of the instrument precision (determined in section 4.5.1). As described in that section everything beyond two standard deviations was considered as a valid reading (previously described as 'all' possible peak-trough pairs).

5.5.2.2 Increment threshold size

The subsequent increment in the size of the 'delta h' threshold was chosen to be two standard deviations of the instrument precision. The size of the 'delta h' threshold was increased by 0.1 m up to the maximum magnitude detected in the initial threshold size process. This was done to save computational time as any other high value would require longer data processing.

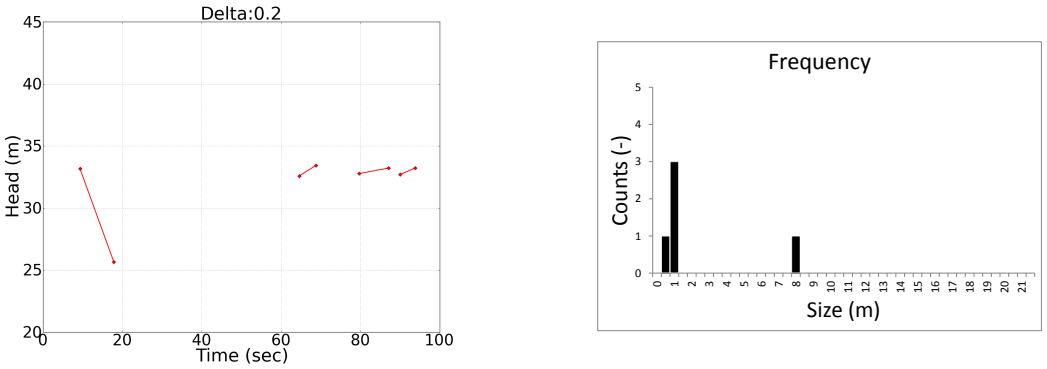
Figure 5.14 shows the first example of the practical application of the enhanced algorithm to the real pressure data previously seen in Figure 5.1. Figure 5.14 (a) shows peak-trough points after first application of the smallest size of the 'delta h' threshold and only pairs with magnitudes equal or greater than 0.1 were retained. As can be observed the algorithm found a large number of points.



(a) **(b)**

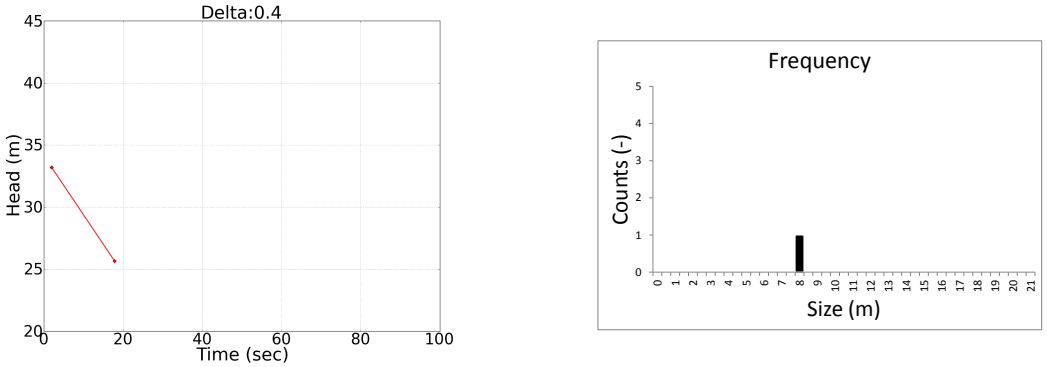
Figure 5.14: Initial identification of peaks with magnitudes greater than the size of the 'delta' threshold = 0.1 m; (a) peaks (black dots) and troughs (red dots) identified, (b) histogram of peak-trough pairs.

The subsequent increments (which returned new pairs) in the size of the 'delta h' threshold by 0.1 m are presented in Figure 5.15, Figure 5.16 and Figure 5.17.



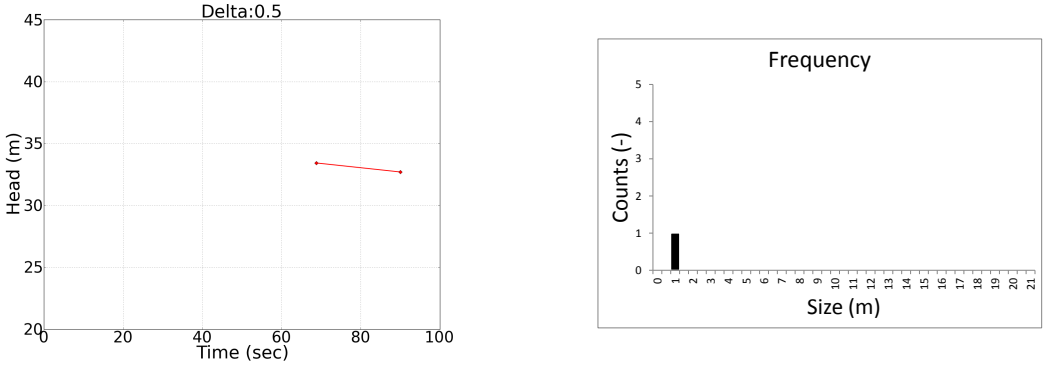
(a) **(b)**

Figure 5.15: Increment in the size of the 'delta' threshold=0.2 m of a threshold value results in four new pairs being detected; (a) new peak-trough pairs, (b) histogram of new peak-trough pairs.



(a) **(b)**

Figure 5.16: Increment in the size of the 'delta' threshold=0.4 m of a threshold value results in one new event being detected; (a) new peak-trough pair, (b) histogram of new peak-trough pair.



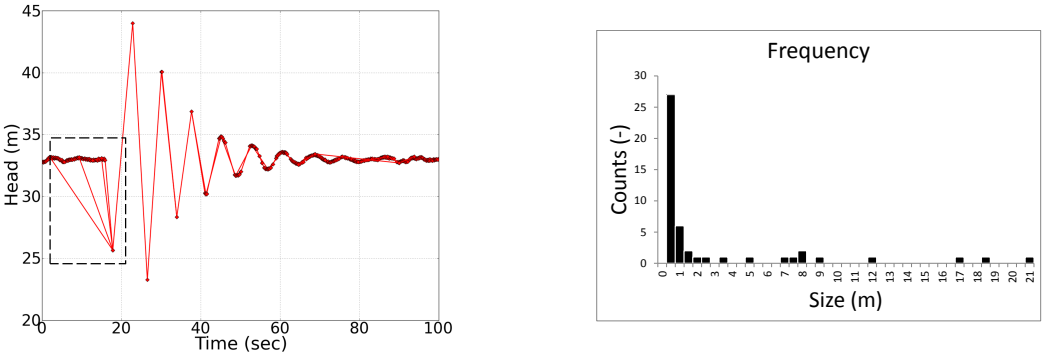
(a) **(b)**

Figure 5.17: Increment in the size of the 'delta' threshold=0.5 m of a threshold value results in one new pair being detected; (a) new peak-trough pair, (b) histogram of new peak-trough pair.

Figure 5.18 shows the final outcome of the enhanced algorithm after the removal of duplicates. The peak-trough pairs shown in Figure 5.18 are listed as a series of pairs of different magnitudes and durations and provide a characterisation of the pressure regime experienced by a pipe. The figure shows also pairs which do not represent accurately the original signal.

Figure 5.19 presents a second example of pressure transients. It is a 'slow', repeating pressure transient event recorded in the real WDS. The final outcome of the application of the enhanced algorithm is presented in Figure 5.20. The peak-trough pairs identified by the enhanced algorithm are listed as a series of peak-trough pairs of different magnitudes and durations and provided a characterisation of the pressure regime experienced by a pipe.

Detection of peaks and troughs allowed the characterization of 'all' (possible) peak-trough pairs starting from the minimum possible value, e.g. analogue to digital conversion accuracy. Pairing the subsequent pairs, however, resulted in the detection and inclusion of peak-trough pairs inaccurately representing the measured data in the final outcome. Therefore, a technique which can assess the pairs and identify the ones which accurately represent the measured signal must be



(a) **(b)**

Figure 5.18: A final result shows how many pairs are detected in the final outcome. To achieve it databases with pairs detected at each stage of the algorithm run were compared to remove repeating pairs and add any new pairs. (a) Signal with peaks (black dots) and troughs (red dots) identified and paired (red lines). The pairs highlighted in a rectangle inaccurately represent measured time series data. (b) Histogram of peak-trough pairs.

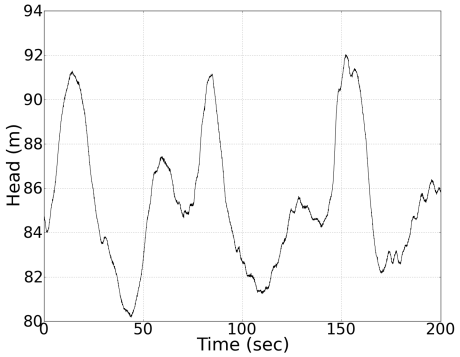
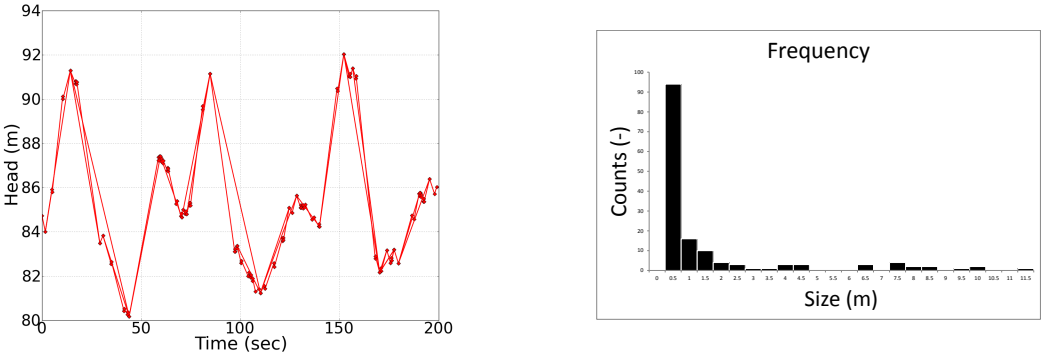


Figure 5.19: Original signal from a water network consists of 20,000 data points.



(a) **(b)**

Figure 5.20: The final result shows peak-trough pairs identified in the final outcome. To achieve it databases with pairs detected at each stage of the algorithm run were compared to remove repeating pairs and add any new ones. (a) Signal with peaks (black dots) and troughs (red dots) identified and paired (red lines). The figure shows also some pairs which inaccurately represent measured time series data. (b) Histogram of peak-trough pairs.

developed and applied. The peak-trough pairs which accurately represent measured data are then considered as pressure transients events.

5.6 Transient event selection

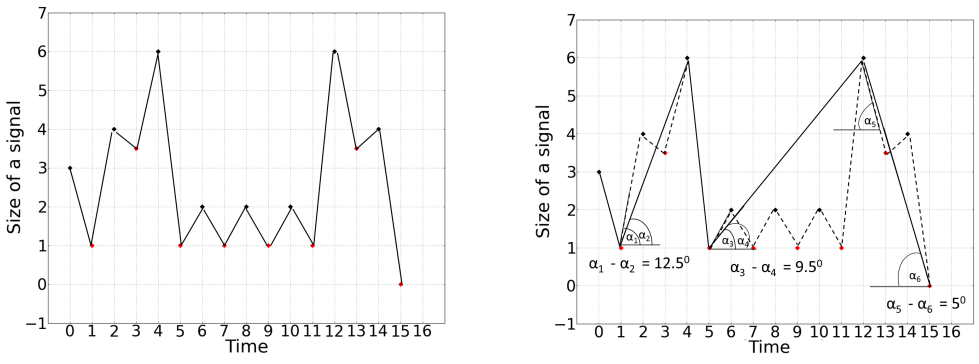
This section describes the development and application of potential methods to remove possible peak-trough pairs which are poor/inaccurate representation of measured signal identified during the application of the enhanced algorithm to time series data. This assessment was undertaken to then successively remove the inaccurate representations from the final characterisation, allowing only pairs which accurately represent measured data, i.e., pressure transient events, to be retained.

5.6.1 Comparison of angles between pairs

To eliminate some of the pairs which were connected during the analysis, a comparison of the angles between the pairs originating at the same point was undertaken. The conceptual application of this method to the data sample representing pairs which correspond to the original data and unsatisfactory pairs can be seen in Figure 5.21. The successive connections of pairs between each increments of the size of the 'delta h' threshold resulted in the inclusion of some pairs (e.g., α_4) in the final output. The method should show that the greatest difference between the angles should be for pairs α_3 and α_4. This would allow the pair α_4 to be deselected from the final outcome. To find angles between pairs originated at the same point, the slope for each pair was calculated using the equation:

$$m = \frac{y_2 - y_1}{x_2 - x_1} \tag{5.1}$$

where m - slope, y_1, y_2 - time values, x_1, x_2 - signal size values.



(a) **(b)**

Figure 5.21: The comparison of the angles from two processes of point detection, (a) shows the detected points greater than the initial (first) peak threshold (size of a 'delta h' threshold), (b) shows the detected points greater than the second peak threshold.

In the example in Figure 5.21 (b) the α_1 and α_2 were calculated as follow:

$$m_1 = \frac{4-1}{3-2} = 3 \tag{5.2}$$

$$arctan_1(3) = 1.249(rad) = 71.56^\circ \tag{5.3}$$

$$\alpha_1 = 71.56^\circ \tag{5.4}$$

$$m_2 = \frac{6-1}{5-2} = 1.66 \tag{5.5}$$

$$arctan_2(1.66) = 1.03(rad) = 59.01^\circ \tag{5.6}$$

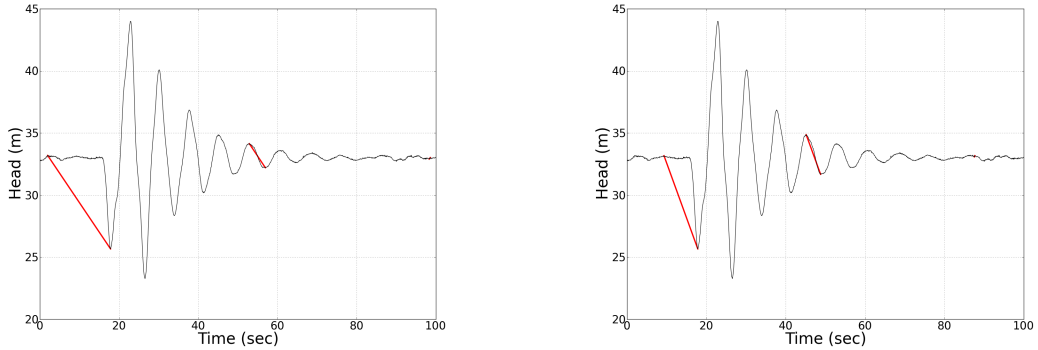
$$\alpha_2 = 59.01^\circ \tag{5.7}$$

$$\alpha_1 - \alpha_2 = 71.56^\circ - 59.01^\circ = 12.55^\circ \tag{5.8}$$

Figure 5.21 shows that the comparison of angles between pairs originating at the same point is not an appropriate method to discard the unsatisfactory pairs. It shows that difference between angles α_3 and α_4 is less than between α_1 and α_2. Therefore, this method is conceptually not able to distinguish between 'true' and 'false' pairs as both 'true' pairs can have a difference between angles greater and less than the 'false' pair.

5.6.2 Ratio of change of head with respect to time

The ratio of Δh/Δt values for each pair was explored to reveal whether thresholds on this value would be a sufficient indicator to reject unsatisfactory pairs. Figure 5.22 (a) and Figure 5.22 (b) show two pairs (in red) in each figure for which a ratio Δh/Δt was calculated. One of these pairs is unsatisfactory (first pair in Figure 5.22 (a) and (b) respectively). The second pair is a satisfactory representation of the measured data. For both pairs in each figure the calculated ratios Δh/Δt value were the same, 0.4 m/s for the pairs highlighted in Figure 5.22 (a) and 0.8 m/s in Figure 5.22 (b) respectively. It shows that a single threshold cannot distinguish between 'true' and 'false' pairs. It was, therefore, not used for elimination of the unsatisfactory pairs.



(a) **(b)**

Figure 5.22: Plot shows two examples with pairs (in red) of equal ratio values $\Delta h/\Delta t$. First event does not match a real data; second event matches the data; (a) events with $\Delta h/\Delta t = 0.4m/s$, (b) events with $\Delta h/\Delta t = 0.8m/s$.

5.6.3 Comparison of goodness of fit measures

To evaluate which pairs are realistic representations of pressure transient events experienced by the pipe (i.e pairs which are close to the original signal) and which are not, the original signal was compared to the linear fit between peak–trough pairs. During the algorithm development process a number of goodness of fit coefficients were tested. These ranged from R^2, MAE and coefficients based on the RMSE. To ensure that the process removed unsatisfactory pairs, regardless of overall magnitude or duration, it is required that the chosen correlation measure should be sensitive to a transformation of shape, scale of events and their durations (number of points).

This section presents a number of correlation measures which were investigated in the transient events selection. A range of conceptual plots were designed to compare peak-trough pairs to various shapes of a conceptual signals (see Figure 5.3). These signal samples represented typical challenges which may be encountered while comparing pressure transient events to a real signal. The samples were used to test different correlation measures. The best correlation measure, which was able to distinguish between unsatisfactory pairs was then tested on real data to check its performance. The investigated measures of goodness of fit included:

- sum of squares of residuals ($S_r{}^2$)

$$S_r^2 = \sum e^2 \tag{5.9}$$

Where: e^2 represents the fitting error, the deviation of a measured value of an element from its estimated function value.

- Young's Coefficient of Determination (R^2) (Young et al., 1980)

$$R_t^2 = 1 - \frac{\sum e^2}{\sum y^2} \tag{5.10}$$

Where: y^2 represents measured data values

- mean square error MSE

$$MSE = mean(e^2) \tag{5.11}$$

- mean absolute error MAE

$$MAE = mean(|e|) \tag{5.12}$$

- median absolute error (MdAE)

$$MdAE = median(|e|) \tag{5.13}$$

- root mean square error RMSE (Anderson and Woessner, 1992)

$$RMSE = \sqrt{MSE} \tag{5.14}$$

- normalized root mean square error NRMSE

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{5.15}$$

Where: y_{max} and y_{min} represents the range of observed values of a variable being analysed.

Table 5.3 shows seven Events from which Events 1 and 2 (in red) are visually satisfactory, i.e. they represent pairs that are close to the original signal (in black). Events 1 and 2 show the events that agree visually. Events 3, 4, 5 and 6 represent visually unsatisfactory events that are not close to the original signal. These events fall into two types. Events 3 and 5 represent the first type and Events 4 and 6 - the second type. Seven separate goodness of fit measures were used to calculate how well the events agree with the original signal (see Table 5.3). The obtained results were then evaluated against the following criteria:

(a) the measure should identify Events 1 and 2 as satisfactory and Events 3, 4, 5 and 6 as unsatisfactory,

(b) the difference between events visually satisfactory and visually unsatisfactory should be easily identified,

(c) the measure should identify the same type of events irrespective of their sizes, i.e., Event 3 and Event 4 are not the same, Event 5 and Event 6 are not the same.

Table 5.4 shows evaluation of the goodness of fit measures. Measure $S_r{}^2$ correctly identified visually satisfactory events but did not identify events of the same type. $R_t{}^2$ correctly identified satisfactory and unsatisfactory events but did not characterise events well. MSE did not correctly identify visually satisfactory events, i.e. showed that Event 2 is very similar to Events 3 and 4. Similarly MAE showed that Events 2, 3, 4 are almost the same. Median Absolute Error (MdAE) showed that Event 2 is a poorer representation than Event 4 and likewise MAE did not correctly distinguished between satisfactory and unsatisfactory events. RMSE showed similar results as MSE. RMSE itself was unsuitable for use, as it did not perform accurately with transients of different sizes. However, it showed a sufficient sensitivity to the shape of pressure transient events. NRMSE, which is RMSE normalized by the range of observed values, has a high sensitivity to the shape of transient events and is capable of dealing with events of different sizes. According to the evaluation only the NRMSE measure fulfilled all designed criteria. This measure appeared as potentially suitable for pressure transient event selection with a threshold of around *0.2* provided a separation between 'good'(visually satisfactory) and 'poor' (visually unsatisfactory) events for the examples in Table 5.3.

Table 5.3: Comparison of correlation measures on conceptual plots. Values highlighted in red show when the measure did not meet the criteria.

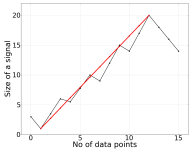
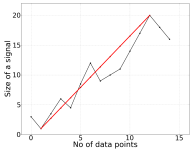
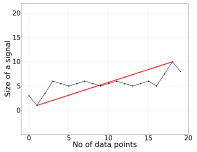
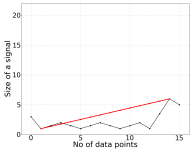
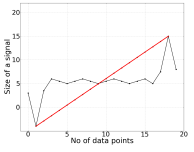
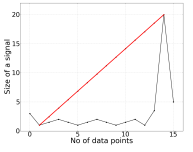
	Visually satisfactory events		**Visually unsatisfactory events**			
Correlation measures	**Event 1**	**Event 2**	**Event 3**	**Event 4**	**Event 5**	**Event 6**
						
S_r^2	18.51	49.57	77.23	60.11	375.47	1222.12
R_t^2	0.98	0.97	0.87	0.20	0.48	-1.78
MSE	1.54	4.13	4.29	4.29	20.86	87.29
MAE	0.91	1.67	1.68	1.69	3.75	7.57
MdAE	0.73	1.61	1.87	1.48	4.16	7.29
RMSE	1.24	2.03	2.07	2.07	4.57	9.34
NRMSE	0.06	0.11	0.23	0.41	0.24	0.49

Table 5.4: Evaluation of the correlation measures against criteria.

	Correlation measures						
Criterion	S_r^2	R_t^2	MSE	MAE	MdAE	RMSE	NRMSE
(a)	√	√	√	√	–	√	√
(b)	√	–	–	–	√	–	√
(c)	–	√	–	–	√	–	√

5.6.4 Evaluation of the algorithm

Figure 5.23 shows a short section of time series pressure data. It shows the output of the enhanced algorithm which subsequently connect peaks and troughs into unique pairs. A large number of peak-trough pairs were found. These also included pairs which are unsatisfactory representation of the time series data.

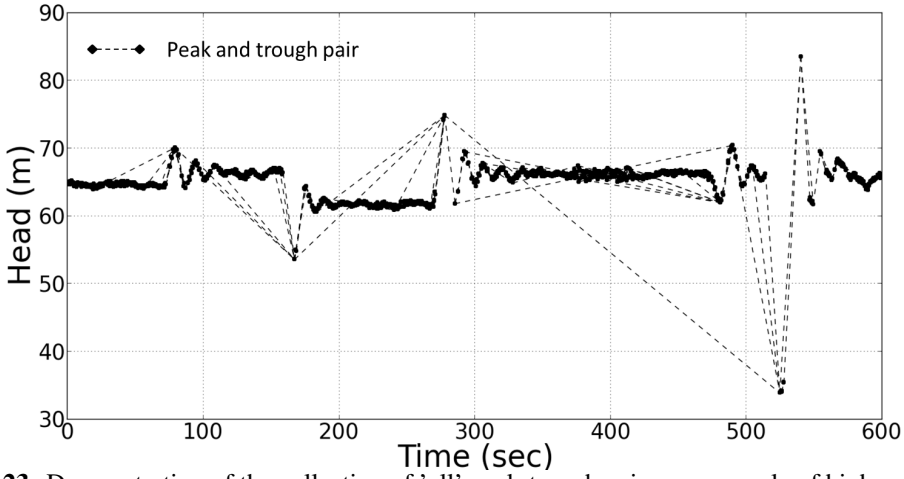


Figure 5.23: Demonstration of the collection of 'all' peak-trough pairs on a sample of high speed data.

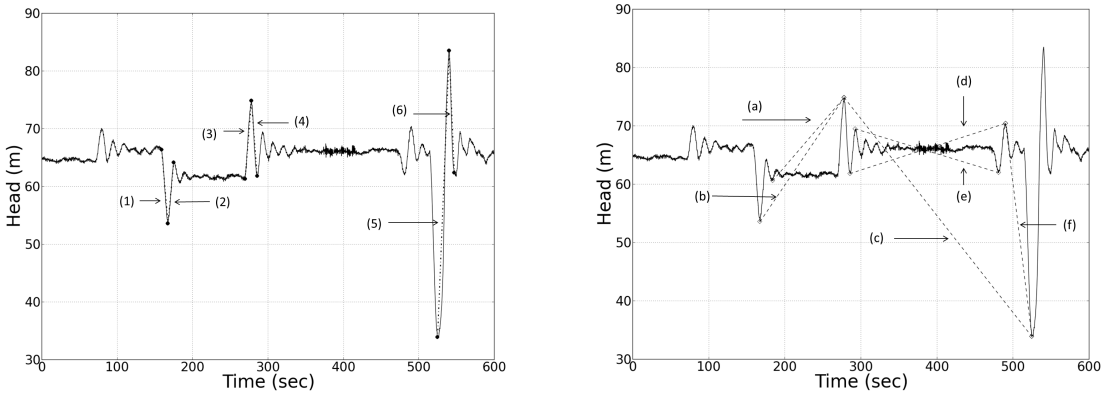
5.6.4.1 Transient event selection

After the generation of all possible peak-trough pairs has been completed (as shown in Figure 5.23) any unsatisfactory pairs have to be removed. To eliminate unsatisfactory events a goodness-of-fit measure NRMSE is used, as identified in section 5.6.3. It compares the peak-trough pairs to the original signal, and any which fall below a defined threshold of fit are discarded.

Figure 5.24 highlights a number of the peak-trough pairs detected by the enhanced algorithm developed in section 5.5. Some examples of the pairs that are a satisfactory representation of the actual pressure signal and therefore should be retained are presented in Figure 5.24 (a). Figure 5.24 (b) shows examples of the pairs which are unsatisfactory representation of original signal; they do not represent what the pipe actually experiences and should be removed from the final characterisation.

5.6.4.2 Experimental evaluation of the threshold

NRMSE correlation measure was identified as potentially suitable for pressure transient events selection. The selection of only accurate representation of an original signal (see Figure 5.24 (a)) requires a determination of a threshold value. The value of a threshold was determined based on evaluation of eight examples of real pressure transients from diverse networks. To achieve



(a) **(b)**

Figure 5.24: Example of peak-trough pairs (a) correctly identified peak-trough pairs and (b) incorrectly identified peak-trough pairs which are not reasonable representation of what the pipe experiences.

it, firstly the increment in NRMSE threshold was successively increased by 0.1 and the number of removed vs retained pairs were compared. When the number of retained events were found to remove satisfactory events the increment in NRMSE threshold was decreased to 0.01 until it removed sufficient number of unsatisfactory events. Care was taken to not remove valid pairs and each step change was visually assessed. It was found that an NRMSE threshold of 0.18 should be applied as the selection criterion to discard peak-trough pairs which should not be considered as part of the pressure transient event. As this threshold value was obtained on much larger datasets from real water network collected during the fieldwork it gives it more generic nature.

Figure 5.25 shows a scatter plot for each unique peak-trough pair characterised by $\Delta h/\Delta t$ against NRMSE values. The threshold is shown with the events from Figure 5.24 (a) and (b). It can be seen that there is no definite split in data for the threshold. However, Figure 5.26 shows the previously investigated dataset (Figure 5.23) to which the NRMSE threshold was applied. From this figure it can be seen that all previously identified 'problematic' events were removed leaving peak-trough pairs which can be seen to be a reasonable representation of the original pressure signal and what the pipe actually experiences.

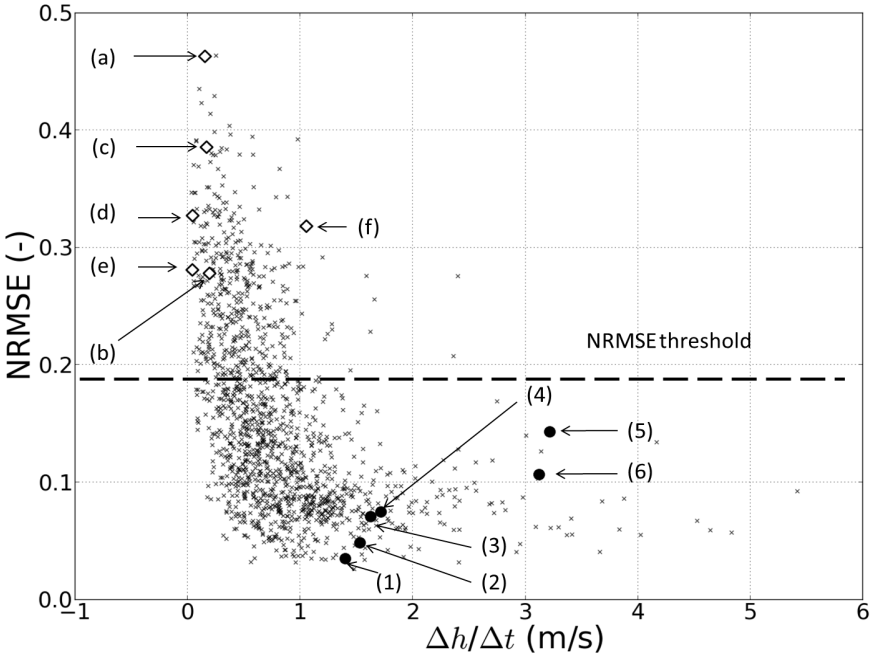


Figure 5.25: Peaks and troughs NRMSE vs $\Delta h/\Delta t$.

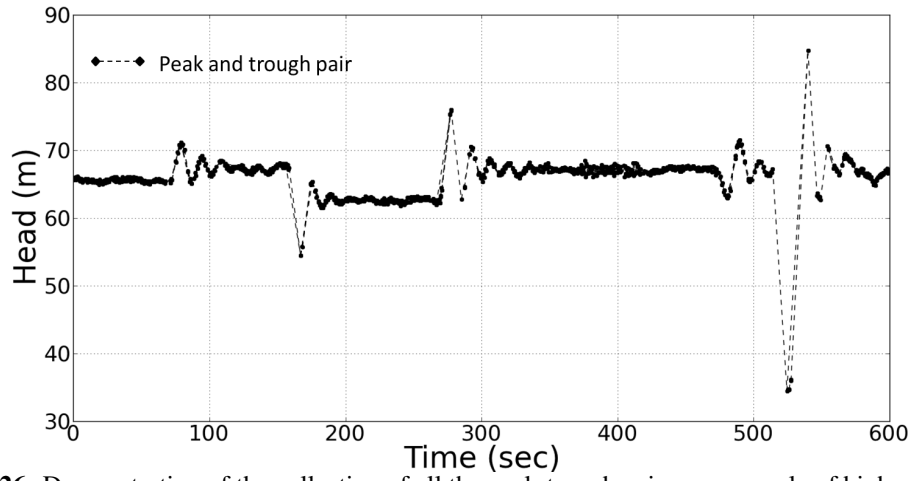


Figure 5.26: Demonstration of the collection of all the peak-trough pairs on a sample of high speed data.

The peak-trough pairs shown in Figure 5.26 are listed as a series of peak-trough events of different magnitudes and durations and provide a characterisation of the pressure regime experienced by a pipe. From this list it is possible to determine summary statistics about the magnitude and durations of the individual transient events in the system. For instance in the example above there were 965 events of magnitude less than 1 m, 8 upsurges with magnitude greater than 10 m and 6 downsurges with magnitude greater than -10 m. It is informative to visually represent the events as system specific transient fingerprint.

The NRMSE measure with a threshold of 0.18 provided distinction between 'good' and 'poor' events for the real data.

5.7 Transient event characterisation and quantification

Correctly identified pressure transient events can be counted and represented in a 2D form of a histogram termed transient fingerprint, see Figure 5.27. The transient fingerprint depends on the observation period and for this research was estimated to be two weeks (see section 4.7). It shows the number of events detected at a given magnitude or duration for the data shown in Figure 5.26. More 'energetic' transient events, those with a greater magnitude and shorter duration, will appear in the bottom left and bottom rights of the plot, depending if they are downsurges or upsurges respectively. Towards the centre of the plot are events of low magnitude and the top of the plot shows events of long duration. There are few events captured with very high magnitudes (Δh = + 50, - 32 m) highlighted in red rectangles in Figure 5.27, that do not form part of the main mass of the plot, these are associated with the large-scale transient event that occurred at the end of the recording period.

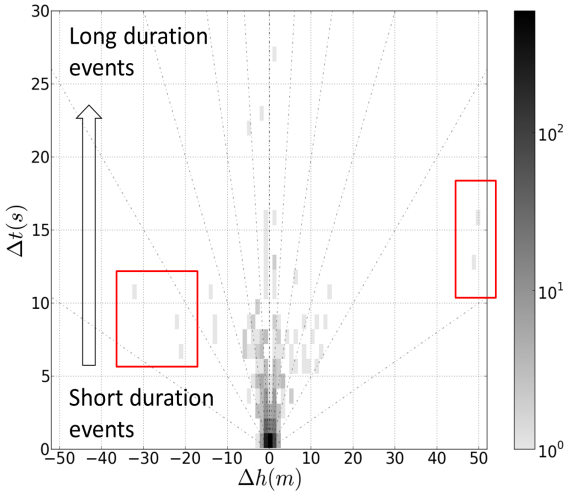
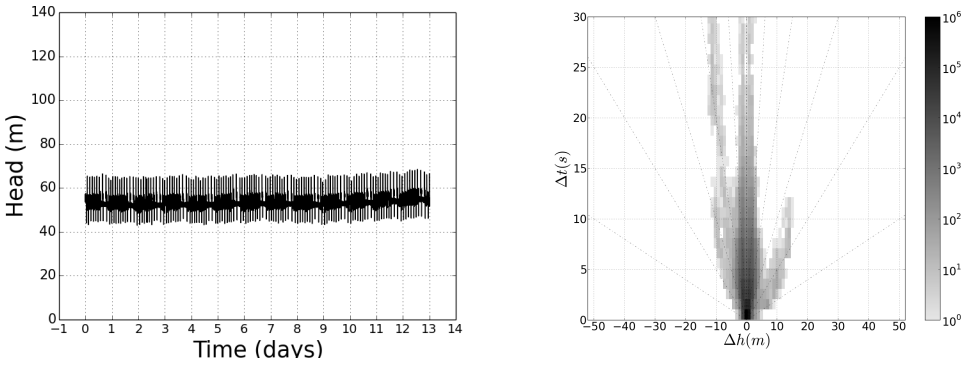


Figure 5.27: A site unique transient fingerprint for 10 minutes of pressure data, from Figure 5.26; a graphical representation of how big, how fast and how often transient events are observed in a water distribution system. High magnitude events are highlighted in red rectangles.

5.7.1 Application to a larger dataset

In this section, results are presented from applying the method to characterise and count different magnitudes of pressure transient events in the recorded time series from two weeks recording in the pump system. Figure 5.28 (a) shows the time series pressure response from the pump system of a very regular pattern of 10-15 m upsurge and downsurge transients being experienced, with periods of slightly higher or lower pressure between. This is consistent with transients due to start/stop of the nearby located pump supplying the system. Figure 5.28 (b) presents the associated transient fingerprints output of the characterisation method. Figure 5.28 (b) is generally

symmetrical (around $\Delta h = 0$), with a large cluster of events at low magnitude regardless of duration. There is a slightly larger number of downsurge events with a magnitude of 5 – 15 m that occur for a duration of 5 – 10 seconds than the equivalent magnitudes of upsurges.



(a) **(b)**

Figure 5.28: Example of two weeks of data from a pump site (a) pressure trace (b) transient fingerprint.

Figure 5.29 presents 120 sec of data, (a) typical upsurge and (b) downsurge observed in the pump site. The upsurge (a) is due to the pump switching on and takes about 8.8 sec with a magnitude of 14.11 m. This sharp pressure rise is then followed by the sharp downsurge of magnitude -10.49 m and duration 11.18 sec. The differences in magnitudes are a result of the increase in static pressure due to the pump being switched on. Conversely the downsurge due to the pump stop (b) of magnitude about -11.49 m takes about 28.73 sec and it is followed by sharp upsurge of magnitude 10.5 m with duration 4.7 sec. For these events the magnitude of the initial upsurges and downsurges are approximately the same (10 – 15 m) however the durations of the downsurge is longer than the upsurge which accounts for the asymmetry seen in Figure 5.28 (b). The ability of the transient fingerprint to distinguish the differences in the pressure traces gave confidence that it is a good method for characterising the pressure response of the system.

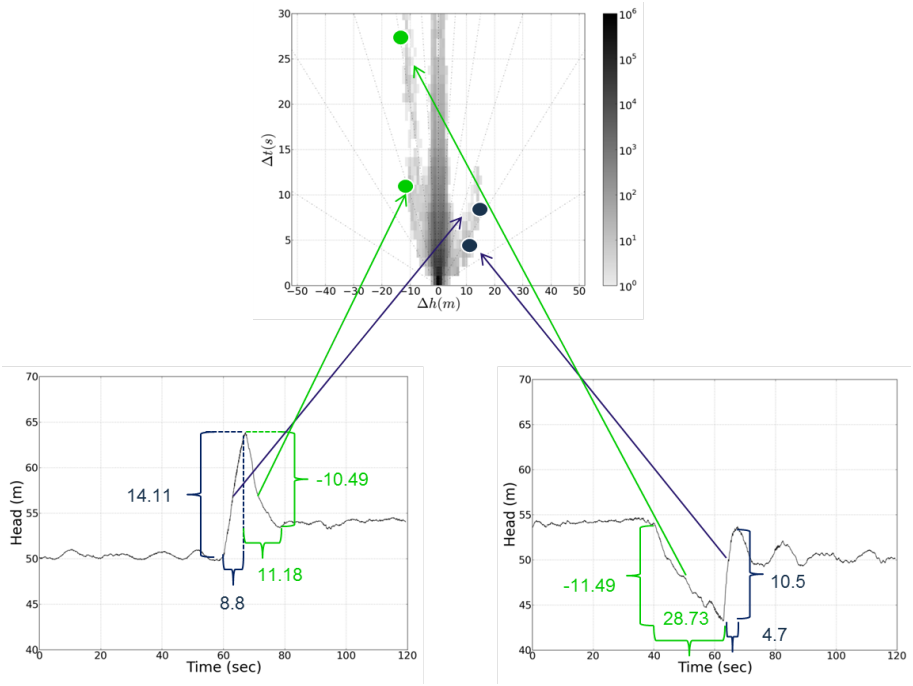


Figure 5.29: Examples of the upsurge and downsurge observed in the pump site (both plots show a duration of two minutes)

5.8 Summary

A method using an enhanced peak-trough detection algorithm was developed, which facilitates the detection of 'all' possible discrete peak-trough pairs. The process of pressure transient event selection was verified by a number of techniques for non-representative peak-trough pairs detection and elimination. A comparison of goodness of fit measures identified a NRMSE as suitable for pressure transient event identification.

The resulting output is a transient fingerprint for each signal recorded from the water network. It allows quantification of the transients experienced by a pipe in a given location. Some of the pressure transient events appear to be counted multiple times. This, however, was designed to account for 'all' possible pressure transient events to subsequently assess their possible contribution to pipe repair records.

The derived transient fingerprint characterise pressure transients in terms of discrete pressure transient events. It enables the assessment of pressure transient events based on their magnitudes, durations and number of occurrences for the first time. Through this characterisation the contribution of pressure transients to the observed pipe repair data can be evaluated.

Chapter 6

Descriptive modelling methodology

6.1 Introduction

This chapter presents the development of methodology framework to understand the contribution of pressure transients to the observed pipe repair records. The methodology was designed to develop the best approach to assess such contribution with the methodology steps driven from pressure transients data. The schematic of a conceptual methodology framework is presented in Figure 6.1.

The methodology starts by categorising pressure transient data as an input to a descriptive model. Two forms of pressure transient data are considered: continuous and discrete, i.e. the rate of change of pressure and the novel transient fingerprint developed as part of this research. Following pressure transient categorisation data scales are determined driven by the pressure transient data availability. The subsequent methodology step describes variables selection for a descriptive model because a pressure transient contribution to pipe repair rates is assessed in combination with other variables (as described in objective 2 of this research). The variables include the traditional modelling variables and categorised pressure transients. Additional soil data is included to compare with categorised pressure transient variables. All selected variables are grouped into form of cases. The form of cases are then input to the modelling method. The final step describes possible modelling methods, their assessment and final selection of a method(s) used to assess the contribution of pressure transients to the observed pipe repair data.

6.2 Pressure transients categorisation methods

This section presents a development of simple values from the rate of change of pressure and transient fingerprint by their categorisation. The categorisation process is capturing key information,

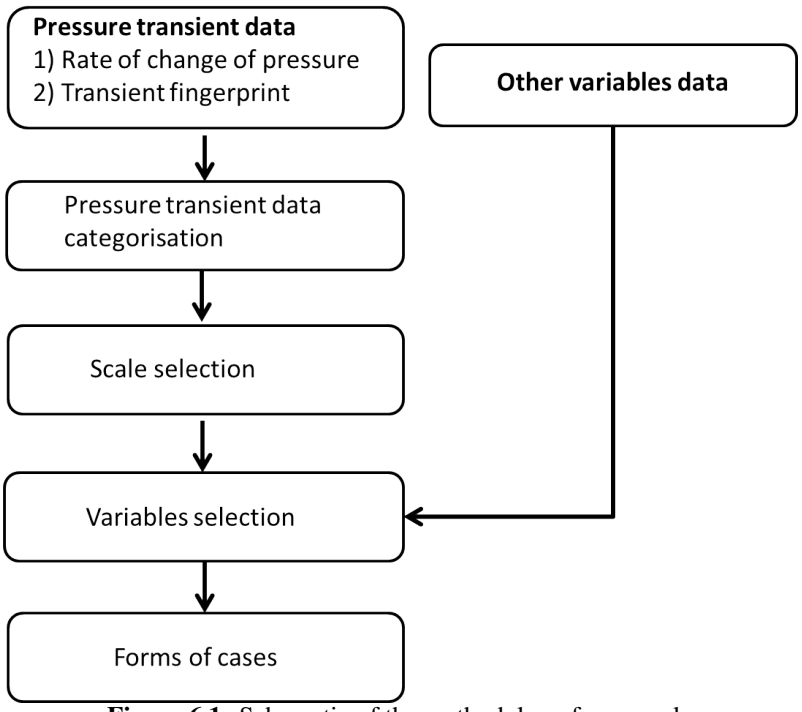


Figure 6.1: Schematic of the methodology framework.

from the rate of change of pressure and transient fingerprint, likely to contribute to pipe repairs. It ultimately forms a simple input variable to be used in a descriptive model.

Previous research suggests (section 2.1.4), that more significant pressure transient events (in terms of damage) exhibit high magnitudes. It has been reported that one extreme high magnitude pressure transient (upsurge or downsurge) caused a pipe failure (see section 2.1.4). The possible categorisation method should therefore highlight the importance of the magnitude. Pressure transients of the same magnitudes but different in terms of a sign $(+/-)$ are given the same importance (e.g., no differentiation with respect to the $(+/-)$ magnitudes) thus absolute values of magnitudes are considered for categorisation of the rate of change of pressure and transient fingerprint.

Previous studies investigating mechanical properties of materials were also investigating a number of occurrences of load changes applied over time, e.g. loads recorded per hour (Pugsley, 1966) or calculated as a strain rate (D'Souza, 1966). It was also reported that if material is affected by corrosion, lower stress or shorter duration of continuous stress application is required for a failure to occur (D'Souza, 1966). Therefore, while magnitude is considered as a primary factor for the possible categorisation, this research also includes time as an important component.

The categorisation method should focus on the extremes of the data and not on the majority (dominating events). For the simplicity, the possible number of divisions (categories) of the data should provide enough resolution of the data by small number of divisions, as higher numbers could bring additional complexity to a future modelling method. Therefore, a reasonable compromise is required between simplicity and complexity allowing capturing interesting behaviour and what

is interesting about different magnitude and duration events.

6.2.1 Rate of change of pressure

The rate of change of pressure (see section 5.3.2.2) is one way to characterise time series data by initial, quick and effective description of variabilities in the signal. This section describes how the rate of change of pressure is categorised into a simple input variable which captures key information likely to contribute to the pipe repairs.

6.2.1.1 Interpercentile ranges

To achieve a simple characterisation of the rate of change of pressure as an input to a modelling method the interpercentile ranges are used. The interpercentile ranges effectively describes variabilities in the signal which may be associated with pressure transients. This categorisation was decided to be used for the initial assessment of the contribution of pressure transients, as it is fast and reasonably advanced to provide a pressure transient descriptor.

Four interpercentile ranges were selected, i.e., 98, 99.8, 99.98 and 99.998 to allow focusing on extremes of the data. The use of high percentile ranges allows to describe and capture fast changes in pressure signal that can be associated with pressure transients. Figure 5.3 (Chapter 5) presents an example of a histogram of rate of change of pressure for which a series of interpercentiles values are calculated. For this example the 98th interpercentile range = 0.05 m/s, the 99.8th interpercentile range = 0.13 m/s, the 99.98th interpercentile range = 0.22 m/s, the 99.998th interpercentile range = 0.33 m/s.

The series of interpercentile ranges were identified as an effective method to categorise the rate of change of pressure. This categorisation provides series of singe variables which can be an input to the modelling method. Four interpercentile ranges were identified as a possible input. Therefore, a method is required to select the best one that provides description of the pipe repair data.

6.2.2 Transient fingerprint

Transient fingerprints were developed to provide a description of discrete pressure transient events in terms of their magnitude, duration and number of occurrences. In addition, this characterisation is too complex to be an input to a modelling method. It was decided that pressure transient events comprising transient fingerprint should be categorised into regions to help assess their contribution to observed pipe repairs. Transient fingerprints categorisation is designed to capture possible significance of pressure transients though both magnitude and time components.

6.2.2.1 Possible choices of categorisation

Transient fingerprint can be categorised into regions using vertical, horizontal lines, line of constant gradient (constant $\Delta h/\Delta t$) and curves. These options are conceptually presented in Figure 6.2.

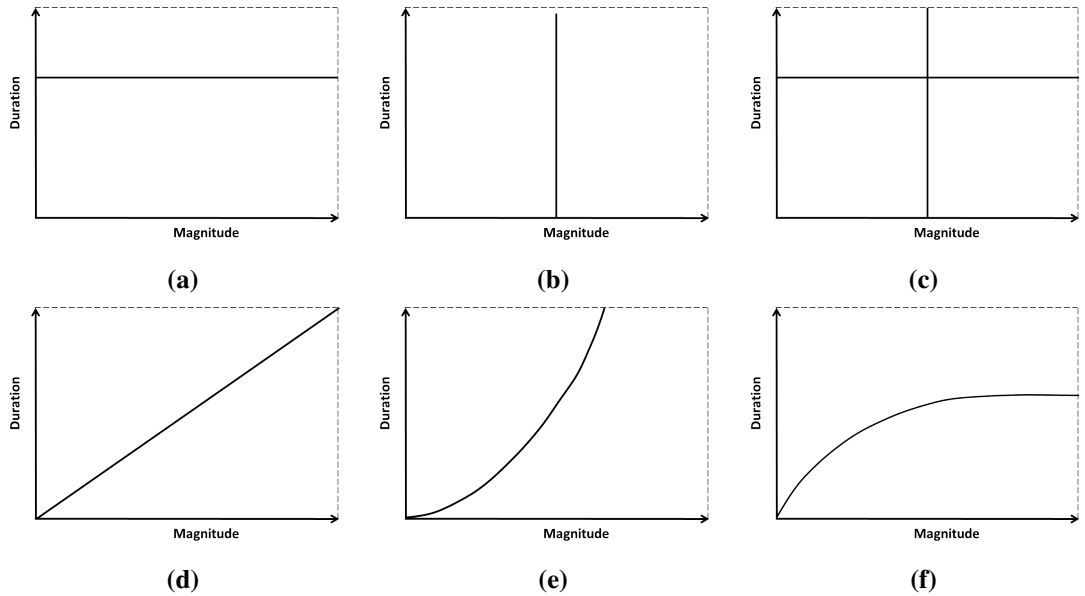


Figure 6.2: Conceptual plots showing possible choices of categorisation of a transient fingerprint into regions using (a) horizontal line, (b) vertical line, (c) horizontal and vertical lines, (d) line of constant $\Delta h/\Delta t$ (e) curve $\Delta h^2/\Delta t$, (f) curve $\Delta h/\Delta t^2$.

The options were analysed and their advantages and disadvantages are presented in Table 6.1. Options which only considered time component were excluded, i.e. option (a). Similarly, options which take into account only magnitudes were excluded, i.e. option (b). While time is an important component it is not more important than magnitude, as suggested by option (f); therefore, this option was not taken further.

Options (c), (d) and (e) were based on both components: magnitude and duration. However, it is unclear whether both are of equal importance. Option (c) required application of thresholds on magnitude and duration which are uncertain values, challenging to estimate. This was seen as a disadvantage and this option was not taken further. Option (d) considered magnitude and duration as equally important. Therefore, option (d) was considered as one of a possible choice. Option (e) ($\Delta h^2/\Delta t$) weights magnitude of the events over the events with shorter duration. This option was also identified for further consideration.

Table 6.1: Possible options for categorisation of transient fingerprint into regions.

Option number	Possible plot type	Advantages	Disadvantages
(a)	horizontal line	Only one component (Δt) to investigate	Δt component only important: Δh not considered
(b)	vertical line	Only one component (Δh) to investigate	Δh component only important: Δt not considered
(c)	vertical and horizontal lines	both components, Δh and Δt considered, a simple division of transient fingerprint into regions	application of two Δh and Δt thresholds (arbitrary numbers) to analyse two components
(d)	line of constant $\Delta h/\Delta t$	both components used in a single combination of $\Delta h/\Delta t$	Δh and Δt are equally important components
(e)	curve $\Delta h^2/\Delta t$	both components, Δh and Δt, in a singe combination, advantage of weighting Δh over Δt, i.e. magnitude more important than time	–
(f)	curve $\Delta h/\Delta t^2$	both components, Δh and Δt, in a singe combination	weighting Δt over Δh time more important than magnitude

6.2.2.2 Theoretical evaluation

Previous research found that the rate of change of energy during a transient event forms a part of the equation describing energy in the pipe (Karney and McInnis, 1990). The rate of change of energy in the pipe was then related to power (i.e. change of energy in unit time during a transient event). Therefore, a component $\Delta h^2/\Delta t$ found in option (e) (see Table 6.1) can be loosely associated with the dynamic component in a rate of change of energy in a pipe. The relation of option (e) to the rate of change of energy shows that option (e) is possibly more suitable choice for the categorisation of pressure transients than option (d).

6.2.2.3 Practical evaluation

A significant amount of small events was found in each transient fingerprint. Each transient fingerprint consisted of millions of events which had magnitudes slightly above zero (Δh) and very close to zero duration (Δt). These events were observed close to the origin of the plot

(as an example see Figure 5.27). A threshold of 0.1 m in magnitude was applied during the pressure transient event selection process. This threshold represented two standard deviations of the instrument precision as presented in section 4.5.1. Pressure transient events with magnitude above or equal this value were included in transient fingerprints. The excessive amount of small events which dominated each transient fingerprint was found as a possible limitation during the application of selected options (d) and (e). A possible advantage of option (e) over option (d) was a potential to categorise small events into one region. Correctly designed curves could 'drop' very close to horizontal axis near the origin of the plot allowing such categorisation.

6.2.2.4 Categorisation of the rest of the data

This section describes the final categorisation choices made for the transient fingerprint. The categorisations were assessed by two criteria. According to the criteria the designed categorisation should:

1. Select large magnitude and short durations events into one region,

2. Capture differences between transient fingerprints, i.e. differentiate between recorded pressure traces.

It was decided to derive the values of possible curves $\Delta h^2/\Delta t$, by dividing transient fingerprint into regions, where a data driven approach could be implemented allowing the data to dictate the divisions. The values of curves $\Delta h^2/\Delta t$ were found from the global transient fingerprint data comprising all, i.e. 144, locations (391,229,872 events in total). Merging the data from all monitored locations allows the data driven approach to derive values of curves based on all information available.

To categorise the dominant type of events (small and short) into one region, as described in criterion 1, four choices of global centiles are proposed, i.e. 85%, 90%, 95% and 99%. A first $\Delta h^2/\Delta t$ curve should therefore be designed to select each global centile of the events (i.e. 85%) into one region, see Figure 6.3 (a).

The rest of the data was decided to be further divided into more regions. Following pressure transient categorisation method criteria (see section 6.2) four regions are selected as a reasonable compromise between simplicity and complexity as presented in Figure 6.3 (b). In this case region 1 could be related to the events of high rate of change of energy and region 4 to low rate of change of energy respectively. For the simplicity the possible number of regions should be evenly distributed except region 4 which is derived based on the global centiles. Four regions should also differentiate between recorded pressure traces, as described in criterion 2.

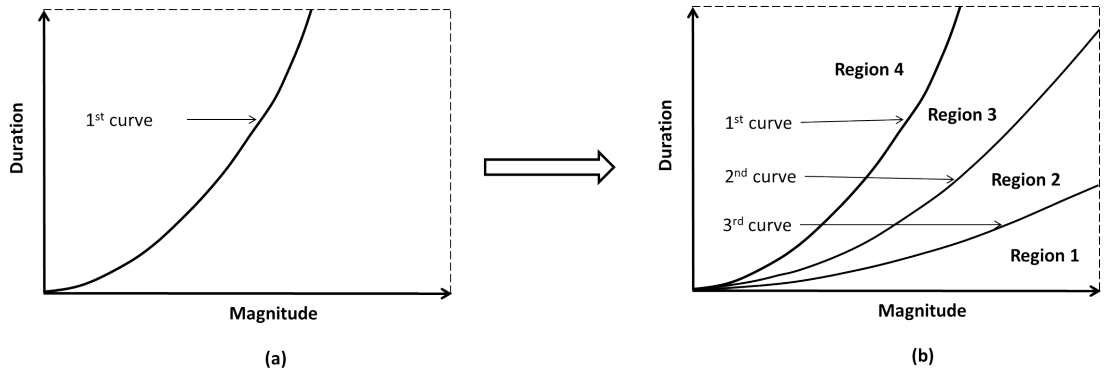


Figure 6.3: Schematic plot showing locations of transient fingerprint regions for one centile example, (a) first curve selecting centile of events, (b) division of the rest of the data into regions based on first, second and third curve values.

6.2.2.5 Application to the real data

This section shows application of previously identified centiles, i.e. 85%, 90%, 95% and 99%, to a global transient fingerprint events data with a division of the rest of the data into equal regions. Table 6.2 shows values for every curve $\Delta h^2/\Delta t$ found for each global centile. The first centile, the 85 % centile in Table 6.2, has a first curve value of 1.67. This value was obtained by:

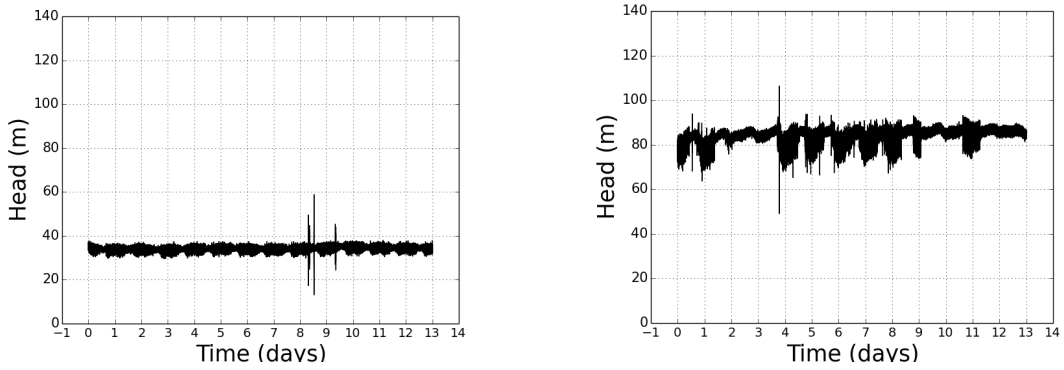
1. calculating $\Delta h^2/\Delta t$ for each point in the merged data,
2. ordering the calculated values from the smallest to the largest,
3. selecting the value representing 85% of the data from the segregated data,
4. dividing the rest of the data into three equal parts, where 1/3 of the data is represented by a second curve value and 2/3 by a third curve value.

The value representing 85% centile is the first value seen in Table 6.2 (1.67). The second value in the table (1.09) represents 1/3 of the rest of the data. The third curve value represents 2/3 of the data.

Table 6.2: Values of curves found for the division of the data.

Centiles	First curve value	Second curve value	Third curve value
85%	1.67	1.09	0.52
90%	1.09	0.71	0.31
95%	0.52	0.31	0.12
99%	0.07	0.05	0.03

The curve values are then applied to two real pressure traces recorded in different zones (see Figure 6.4 (a) and (b)) to evidence the ability of the method to differentiate between recorded signals.



(a) **(b)**

Figure 6.4: Examples of pressure traces used for the application of the categorisation method, (a) location 1 (b) location 2.

The global curve values (Table 6.2) are applied to location 1 and location 2 respectively to illustrate the categorisation of the events from different locations into four regions. Figure 6.5 presents the application of the method to pressure traces recorded at location 1 (Figure 6.4 (a)) whereas Figure 6.6 to location 2 (Figure 6.4 (b)) respectively.

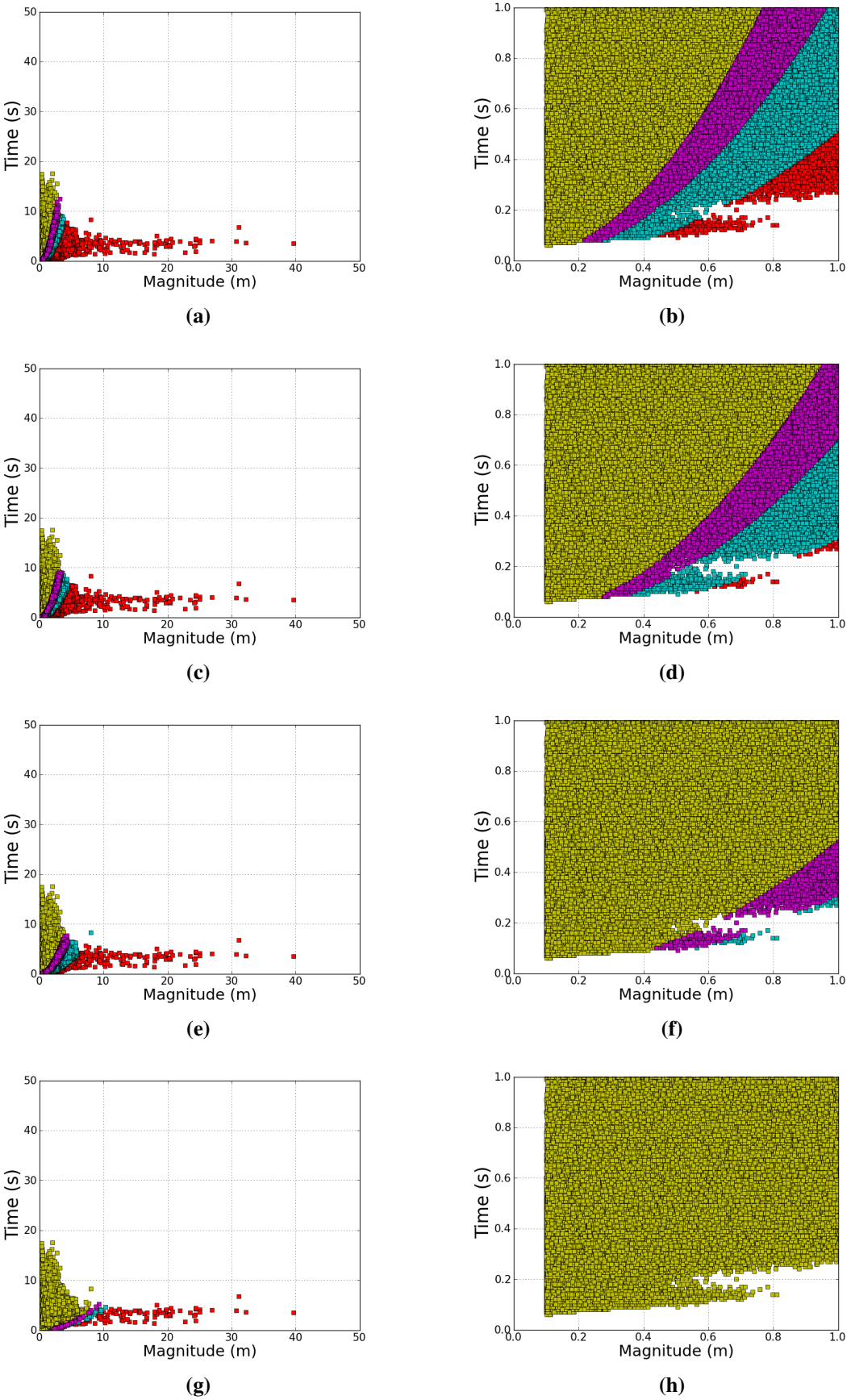


Figure 6.5: Example of the application of three global curves for transient events categorisation (a) 85% of the data and (b) zoom into the data, (c) 90% of the data and (d) zoom into the data, (e) 95% of the data and (f) zoom into the data, (g) 99% of the data and (h) zoom into the data.

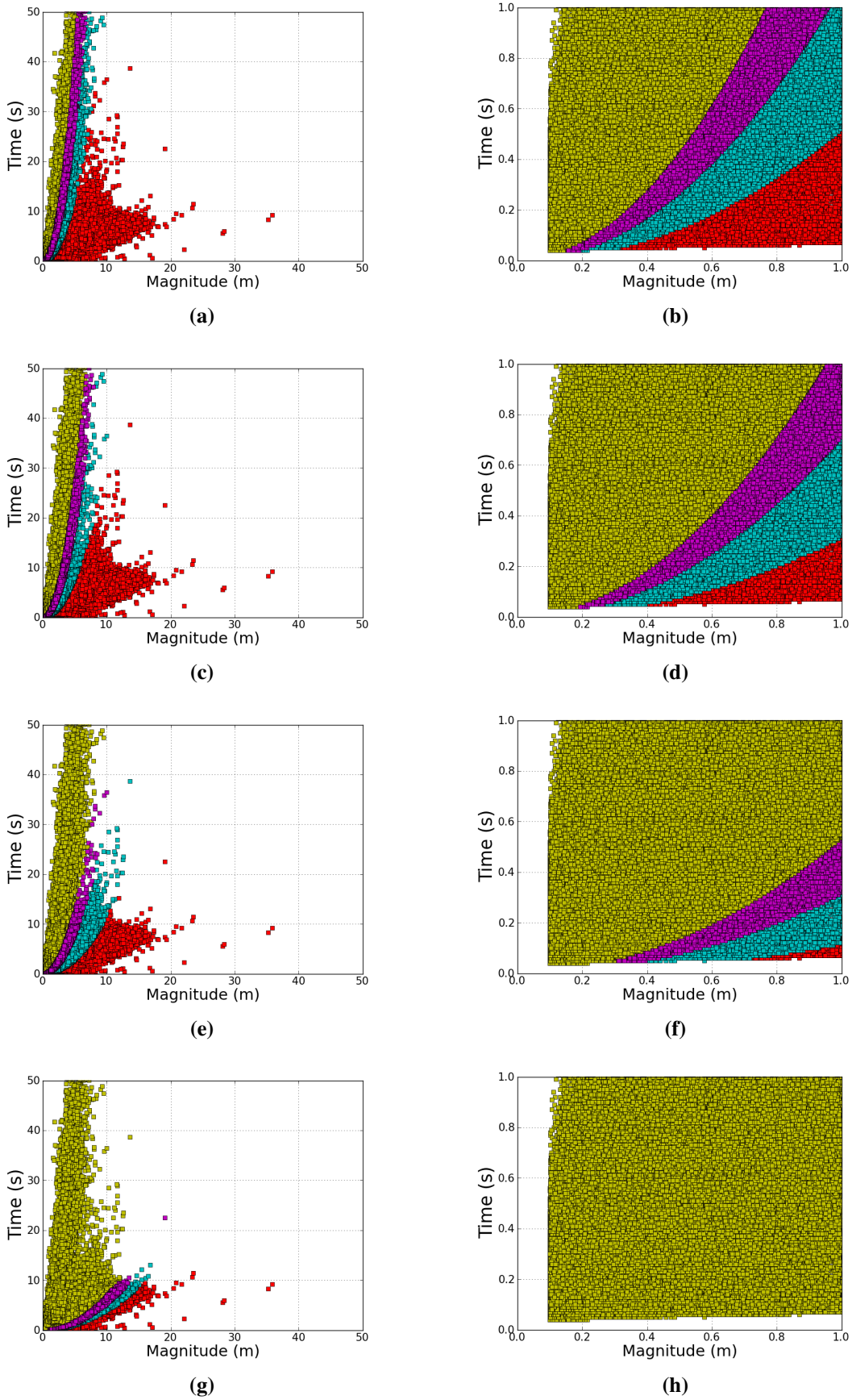


Figure 6.6: Example of the application of three global curves for transient events categorisation (a) 85% of the data and (b) zoom into the data, (c) 90% of the data and (d) zoom into the data, (e) 95% of the data and (f) zoom into the data, (g) 99% of the data and (h) zoom into the data.

As can be seen in Figure 6.5 and Figure 6.6 the application of $\Delta h^2/\Delta t$ curves allowed segregating small magnitude and duration events into region 4. Large magnitude events were categorised into region 1 satisfying the requirements of criterion 1. The figures showed that sites from two different locations are dissimilar. Table 6.3, Table 6.4, Table 6.5 and Table 6.6 present the number of counts in each region for four centiles presented in Figure 6.5 and Figure 6.6. The tables evidence differences between location 1 and location 2 satisfying criterion 2.

Table 6.3: Count of events for a 85% centile.

Location number	Region 1 counts	Region 2 counts	Region 3 counts	Region 4 counts	Total counts
1	111,781	231,420	182,408	1,641,052	2,166,661
2	682,781	623,269	372,162	1,450,134	3,128,346

Table 6.4: Count of events for a 90% centile.

Location number	Region 1 counts	Region 2 counts	Region 3 counts	Region 4 counts	Total counts
1	34,078	160,014	149,109	1,823,460	2,166,661
2	391,883	539,858	374,310	1,822,295	3,128,346

Table 6.5: Count of events for a 95% centile.

Location number	Region 1 counts	Region 2 counts	Region 3 counts	Region 4 counts	Total counts
1	851	33,228	77,703	2,054,879	2,166,661
2	207,310	184,576	290,895	2,445,565	3,128,346

Table 6.6: Count of events for a 99% centile.

Location number	Region 1 counts	Region 2 counts	Region 3 counts	Region 4 counts	Total counts
1	92	22	46	2,054,879	2,166,661
2	125,295	31,541	20,306	2,951,204	3,128,346

Theoretical and practical evaluations showed that a curve $\Delta h^2/\Delta t$ is a more suitable choice for transient fingerprint categorisation than a line $\Delta h/\Delta t$. The option $\Delta h^2/\Delta t$ allows selecting events of small magnitudes and durations into one region and can be related to the rate of change of energy in a pipe during a pressure transient event.

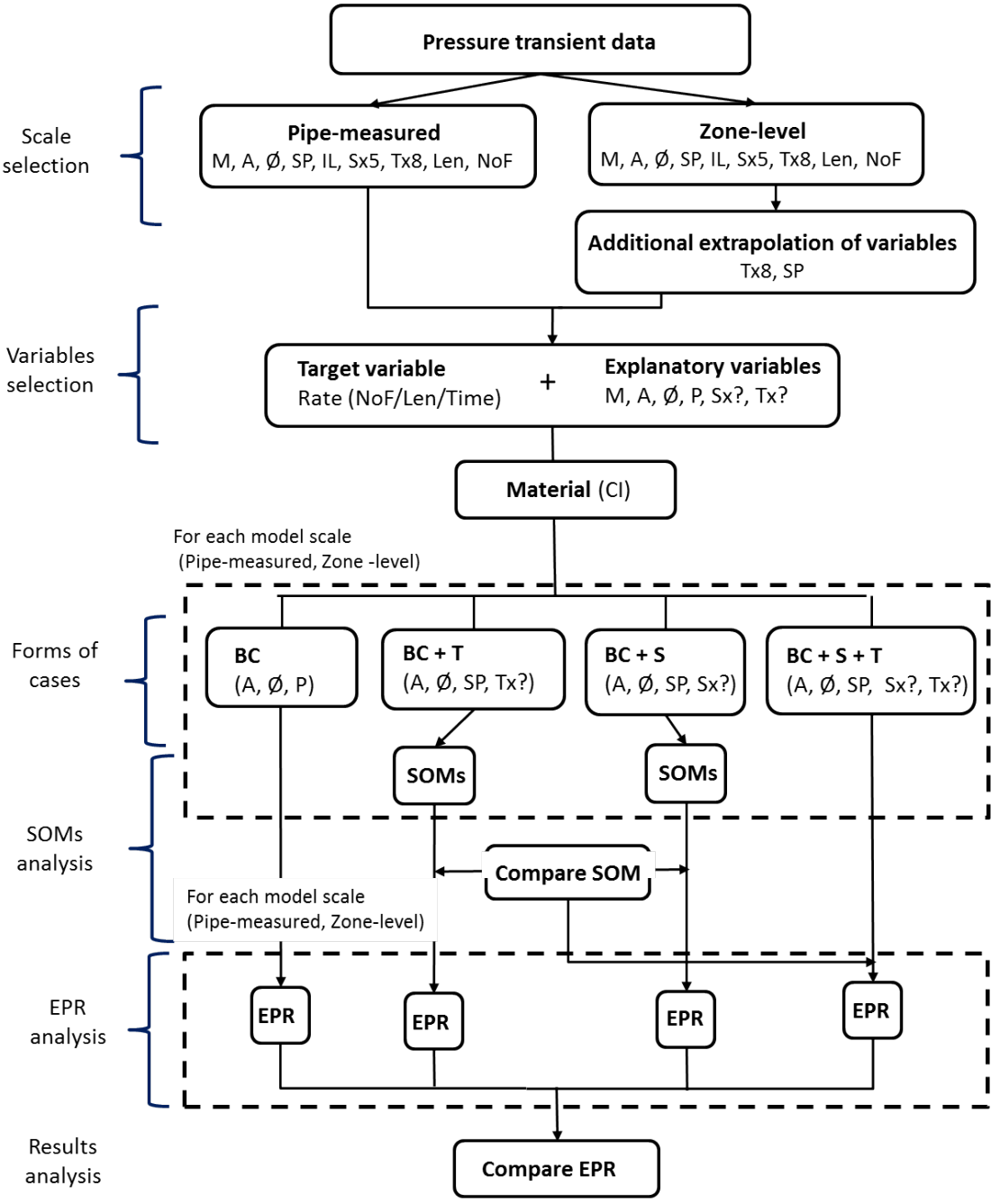
The designed method of categorisation of a transient fingerprint into four regions provided differentiation between recorded pressure data and identified pressure transients of high magnitudes and short durations into one region. The categorisation outcome, as presented in Figures 6.5 and 6.6 provides a simple input to a descriptive model. The method is required, however, to select which of the global centiles presented in these figures best describe the variability in the data, this is presented in a separate section.

6.3 Sub-setting the data/ scale selection

The aim of the following sections is to provide the detailed method which allows understanding of the contribution of pressure transients to the observed pipe repair rates through a descriptive model. Figure 6.7 describes the methodology applied to achieve it. It is divided into five sections: scale selection, variables selection, form of cases, SOMs analysis and EPR analysis. Each of these sections is described in details in a separate section. First section describes possible model scales based on the pressure transient data available. The information about pressure transients was collected from different measurement locations (pipes) within DMAs (as described in section 4.4.2). Two possible options of the input data to the descriptive model were identified: pipe-measured and zone-level datasets. These are presented in the next two sections. Following is a section describing model variables selection, which includes defining target and explanatory variables. This section includes the decision of which material type should be further investigated. Forms of cases are presented in the next section. This section describes grouping of all the available variables into four forms of cases (BC, BC+T, BC+S and BC+S+T). This process is described in details in a separate section. Two final sections assess possible modelling methods and select the two which best describe the contribution of pressure transients to pipe repair rates. This process is presented in details in final sections.

6.3.1 Pipe-measured

The pipe-measured dataset comprises of only the pipes at which the actual measurements of pressure were taken and consists of 144 pipes. As the data is collected from this number of pipes it may be less representative of the whole system. The advantage of utilising the pipe-measured data is that it utilises the high quality pressure transient characteristic for a specific pipe and represents what that pipe actually experienced over the duration of monitoring.



Legend:

NoF – number of failures
M – material
A – age
Ø - diameter
SP – static pressure
Len – length
IL – internal lining
Sx5 – soil variables
Tx8 – pressure transients

BC – base case
BC+T – base case and transients
BC+S – base case and soil
BC+S+T – base case and soil and transient

Figure 6.7: Decision process investigating pipe repair rates from the contribution of pressure transients and other variables at the pipe-measured and the zone-level using forms of cases.

6.3.2 Zone-level

To address the size of the pipe-measured data the assumption of pressure transients being representative across the network (across the areas at which the points of measurement were placed) is applied to obtain a large, zone-level scale data. This key assumption limits the pressure transient variability within monitored sites and requires extrapolation of pressure transient data. This implies that to some extent pipes within the same area experience similar pressure transients. This assumption, however, needs to be made to help categorise the important contribution the pressure transient data inputs into a descriptive model.

Undertaking a zone-level scale modelling will make use of more data available for each pipe in the monitored sites utilising a high quality data and providing maximum available information available for other variables. During the selection of model scale it was decided to use PRVs to scale-down DMAs into zones to provide smaller areas for which possible variables extrapolation can be applied and account for possible impact of PRVs on pressure transient propagation. Number of DMAs where a PRV was installed can be seen Table 6.7. The process of scaling down 67 DMAs into zones increased the number of sites from 67 to 79.

Table 6.7: Presence of PRVs within monitored DMAs.

PRV Presence	No of DMAs
Yes	41
No	26

As can be seen in the Table 6.7, if the measured location was within a PRV zone each pipe within this zone was included in the data. If there was no measured point within a PRV zone then the pipes within the zone were excluded from the analysis. Table 6.8 presents the outcome of the process.

Table 6.8: Summary of data loss during the increase of site numbers in a process of conversion from DMA scale to the zone-level.

Site type	Number of sites	Length (km)	Number of pipes	Number of repairs
DMAs	67	662	19844	1718
Zones	79	604	18181	1546
Removed	–	58	1663	172
% Removed	–	8.8%	8.4%	10%

6.4 Other modelling variables

This research aim to determine the relationship between the number of repairs over the 11-year study period per length of the pipe (pipe repair rate) and a set of variables related to individual pipes previously identified in the literature as important.

The following sections describe the variables which were chosen as an input to a descriptive model. Many variables have been previously considered as having an effect on pipe repairs (see section 2.3.2). Some factors have direct impact on pipe failures, others have more complex or secondary associations. Here the selection of the most significant ones (and therefore suitable for the subsequent modelling) is presented.

Data on pipes was acquired from a UK service provider based on GIS database asset information. The water asset data comprised pipeline data with following structural characteristics: pipe installation date, diameter, material, length and internal lining. Pipe repairs occurrences were supplied in a separate database. Field data collection (see Chapter 4) provided unique information about pressure transients and static pressures for each pipe measured. Table 6.9 list all variables directly and indirectly available in the study datasets. This section investigates which variables should be included as explanatory variables in the assessment of pipe repair rates, see Figure 6.8.

Table 6.9: Characteristic of data available for modelling of pipe repairs.

Variable	Symbol	Method applied
Pipe repair history	NoF	proximity search and association
Pipe length	Len	directly available
Material	M	directly available
Age	A	directly available
Diameter	⌀	directly available
Static hydraulic pressure	SP	field measured
Soil Class	SClass	proximity search
Soil Type	ST	proximity search
Soil Corrosive	SCorr	proximity search
Soil Fracturing	SFrac	proximity search
Soil Workability	SWork	proximity search
Internal lining	IL	directly available

6.4.1 Pipe repairs history

The UK water service provider supplied the digitalised data of historical pipe repair records, over a time span of 11 years. The pipe repair data comprised the repairs service, e.g., repair main using dowel piece less than 200 mm, repair main dowel piece>=200 mm, repair main other methods (not dowel piece) less than 200 mm, and repair main (not dowel) >=200 mm. Other types of

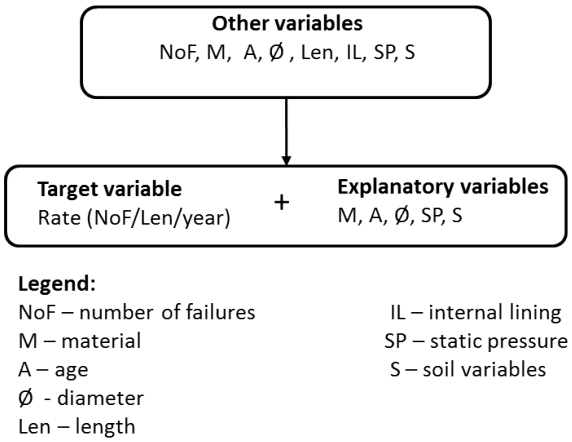


Figure 6.8: Variables selected for modelling of pipe repairs.

repair or maintenance were not available, therefore the analysis conducted in this research does not categorise between types of repairs.

Pipe repairs history was obtained by association of individual pipe repair records to specific pipes from separate (unrelated) databases. Data examination and cross relation techniques allowed to obtain a unique number of repairs per each pipe. The method applied here utilized not only proximity search techniques with a specified radius but also cross referenced to all available information from previously unrelated databases.

Individual pipe lengths available through digitalized asset records from the UK water company were used to form a target variable expressed in units of number of pipe repairs per unit length per unit time referred to as pipe repair rate (NoF/Len/year), see Figure 6.8. One of the reasons to express pipe length as a part of a target variable was a concern about the quality of information for old records, which were subjected to digitalization process. This was evidenced in very small pipe lengths functioning as a valid record in the asset database. Furthermore, small pipe lengths tend to experience more failures. This was historically associated to the number of connections between them rather than to the actual lengths (see 2.3.2). The cleaning of the data was therefore undertaken to remove all small pipe lengths, i.e. less than 1 m, from the dataset as presented in Table 6.10.

Table 6.10: Summary of pipes removed during the small pipe lengths removal process.

Variable	Applied threshold value	% of pipes removed	Number of repairs	Length of pipes (km)
Length	< 1 m	9.79	2	0.67

6.4.2 Pipe material

This section investigates the effect of pipe material types on pressure transients and available repair rates to decide which material type can be chosen for further analysis. The literature review showed that researchers who investigated historical pipe breaks typically segregated the data by pipe material, e.g., cast iron, plastic (see 2.3.2). According to this principle, different models should describe pipe repair rates in contrasting pipe materials (e.g., plastic and cast iron). These materials also age differently and are differently laid in soil, e.g. plastic pipes are usually laid in sand beds whereas cast iron pipes, especially when older than 40-50 years, were usually buried directly in soils. The separation of data by material would capture all these differences. This is an approach that is also followed in this research, mainly because of a different pressure transient response seen in diverse materials (see section 2.1.2). For instance, cast iron material tends to be more brittle in comparison to plastic which is more elastic and able to absorb more energy during a pressure transient event. Different pressure wave speeds are recorded in rigid and plastic pipes. Therefore, the segregation of the data by material can be justified for the assessment of pressure transient contribution to the pipe repair rates.

Field data collection covered 144 locations from diverse materials where measuring devices were directly attached to each of these pipes. The separation of the collected data into material types is presented in Table 6.11. The number of pipes in each material group and corresponding number of pipes with repairs are demonstrated. The largest database with respect to the number of observations (size) and the number of repairs was identified as cast iron with 64 pipes (see Table 6.11). Although this data may be small to be assumed representative for larger networks, the research question does not require to undertake a comparison between the pipe-measured and the zone-level data.

Table 6.11: Characteristics of different pipe material in the pipe-measured dataset.

Material	Symbol	Total length (km)	Number of pipes measured (locations)	Number of pipes with repair records
Cast iron	CI	7.96	64	22
Asbestos cement	AC	3.49	25	5
Ductile iron	DI	2.01	19	2
Plastic	PL	3.68	36	1

Table 6.11 shows that there is not enough pipe measured and repairs for asbestos cement, ductile iron and plastic subsets of data to undertake meaningful statistical analysis and therefore

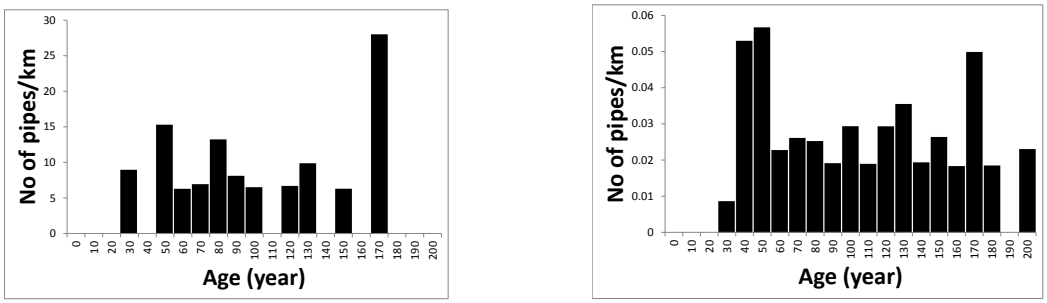
to be considered further. Only cast iron data was therefore decided to be used for the subsequent analysis on the pipe-measured and the zone-level. For the pipe-measured it consisted of 64 cast iron pipes, for the zone-level 7978 cast iron pipes, see Table 6.12.

Table 6.12: Characteristic of the zone-level cast iron data.

Material symbol	Length (km)	Number of pipes	Number of pipes with repair records
CI	290.6	7978	655

6.4.3 Age

Age was identified as one of the important explanatory variable based on other research findings (see section 2.3.2). The age of a pipe was calculated based on the information when the pipe was laid in the ground (available from digitalized water company data). This data was further cleaned, where pipes with age 'Null' or missing values were removed from the data. The distributions of this variable for cast iron pipes in the pipe-measured and the zone-level datasets are presented in Figure 6.9.



(a) **(b)**

Figure 6.9: Distribution of cast iron pipe ages for (a) pipe-measured (b) zone-level data.

6.4.4 Diameter

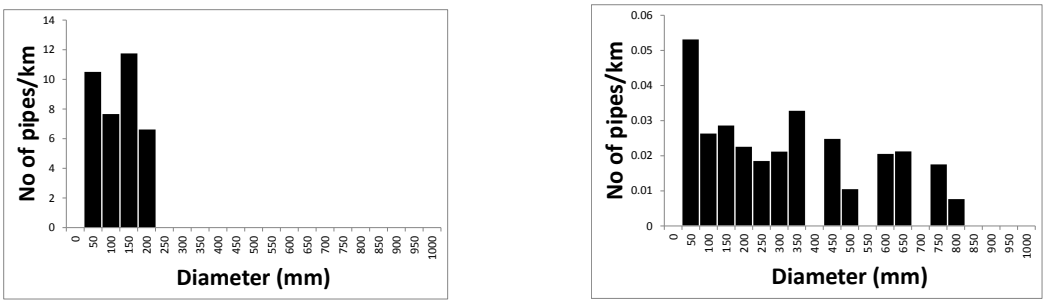
Diameter was directly available from the UK water company database. This variable was identified as important parameter by many researchers (see 2.3.2). Therefore it is included in the analysis.

A process of outlier removal was applied to the cast iron data. Small diameter pipes and extremely large were removed. This was because typical diameter values for distribution pipes were considered to be greater than 50 mm and less than 1000 mm. Smaller diameter pipes were likely to be within private premises and less likely to have pipe repair records held by a water

company. The outcome of outlier removal is presented in Table 6.13. The distribution of cast iron pipes diameter with respect to a number of pipes per length (km) is presented in Figure 6.10.

Table 6.13: Summary of pipes removed during the process of outlier removal for diameter variable.

Variable	Applied threshold value	% of pipes removed	Number of repairs	Total length of removed pipes (km)
Diameter	< 50 mm and > 1000 mm	1.04	0	0.78



(a) **(b)**

Figure 6.10: Distribution of cast iron pipe diameters (a) pipe-measured (b) zone-level data.

6.4.5 Static pressure

Static pressure measurements obtained during the fieldwork are mainly a descriptor of what has been recently introduced in water networks (i.e. pressure management schemes, see (2.1.1)). This variable is identified as important and therefore was included as an explanatory variable. Other source of static pressure data (from DG2 reporting) from the water company was also considered and compared to the field measured data. The DG2 data was of poor quality with gaps in measurements. Field measurements were much more accurate, therefore it was decided to be used as more representative. It has been decided that for the purpose of this research the static pressure recorded from the field as part of the field monitoring programme, represented by the mean value, is a representative measure. The mean value was therefore obtained for each measured location over a typical recording period of two weeks. This data was directly taken to the pipe-measured subset.

The zone-level dataset did not have static pressure values available for each pipe (as pressure transient data). The pressure values are therefore extrapolated based on redefined zones. Static pressure values are obtained from a single known pressure measurement within a zone. In addition,

the extrapolation of static pressure considered local topography (elevation). The values of static pressure for every pipe were therefore adjusted by the relative height difference from the known pressure monitoring point.

The detailed extrapolation process was as follows: For each pipe an output point was created in the middle of the pipe length. The 'X' and 'Y' coordinates were then added to the output point by a mapping function in ArcMAP. A 'Z' coordinate (above ordnance datum, AOD) was then obtained from a digital terrain map (acquired from an open source) by associating with the output point. For each pipe within the zone, the difference in elevation between the determined 'Z' coordinate and the known reference point was calculated. The static pressure was altered accordingly to the relative height difference from known reference point. The example of 10 zones is presented to evidence to process, see Table 6.14

The difference between measured and calculated static pressure values for the zones where checked to determine if accuracy was within $\pm$ 2m. This check was done because the mean static pressure was only 'truly' representative for the point of measurement. As can be seen in Table 6.15 some locations showed differences between the calculated and measured values above $\pm$ 2m, however the overall average difference was 1.6 m.

Static pressure values below 10 m were removed from the data following the current pressure regulatory standards (see 2.1.1). The 0.04 % of the pipes were removed, these had 1 repair record as summarised in Table 6.16. The distribution of static pressures after removing these records, in cast iron pipes, is presented in Figure 6.11.

6.4.6 Soil

Soil data is one of the variables used by other researchers to evaluate pipe repairs (see 2.3.2). The literature, however, presents inconclusive findings when assessing whether soil data is actually a useful predictor or not. Therefore, soil variables seem to be important secondary factor contributing to pipe repairs. This is mainly due to the fact that soil parameters seem to have complex association with pipe repairs.

Soil data was included in the analysis for comparison to pressure transient data. This research decided to compare soil variables to pressure transient data to show whether soil variables have

Table 6.14: Example of comparison of calculated and measured static pressure values for 10 sites.

Zone	Location number	Measured static pressure (m)	Calculated static pressure (m)	Difference (m)
2	1	44	–	
	2	56.7	56.6	0.1
	3	41.2	41.8	0.6
	4	45.1	46.2	1.1
52	1	58.9	–	
	2	60	41.3	9.7
6	1	26	–	
	2	37.8	39.6	1.8
	3	41.6	42.3	0.7
13	1	44.1	–	
	2	28.9	30.7	1.8
16	1	45.9	–	
	2	46.1	46	0.1
46	1	52.3	–	
	2	48	43.2	4.8
31	1	70.7	–	
	2	83.8	84.1	0.3
36	1	38.1	–	
	2	39.5	39.4	0.1
	3	39.7	39.8	0.1
38	1	40.7	–	
	2	44.5	45.4	0.9
	2	38	40.7	2.7
	3	39.4	39.4	0.0
71	1	22	–	
	2	22.3	20.1	2.2
	3	19.5	17.7	1.8
Min		19.5	17.7	0.1
Max		83.8	84.1	9.7
Average		43.5	42	1.7

Table 6.15: Overall comparison of static pressure values for 79 zones.

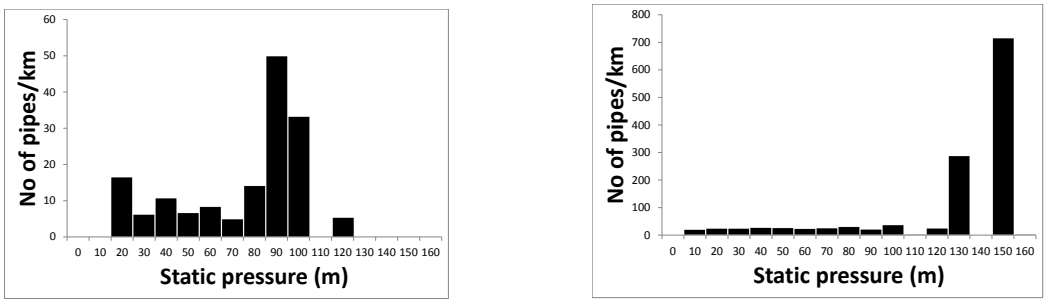
	Measured static pressure (m)	Calculated static pressure (m)	Difference (m)
Min	17.9	17.7	0.1
Max	132.5	132.5	36.5
Average	46.8	46.9	1.6

Table 6.16: Summary of pipes removed during the process of outlier removal by each variable and overall characteristic.

Variable	Applied threshold value	% of pipes removed	Number of repairs	Length of pipes (km)
Static pressure	$<$ 10 head (m)	0.04	1	0.44

more contribution to pipe repair rates than pressure transients. In addition, soil data allow to explore what relationships can be observed when both variables are simultaneously assessed.

Soil data under licence from the Soil Survey and Land Research Centre, had the following characteristics: soil class, type, corrosivity, fracturing and workability. The data had a resolution



(a) **(b)**

Figure 6.11: Distribution of cast iron pipe static hydraulic pressures in (a) pipe-measured (b) zone-level data.

100 m. Soil data was associated to individual pipes through proximity search in ArcMap which choose one soil type when a pipe crosses two soil types. 10 soil classes and 18 soil types were identified in the available data, nine keys were available for soil workability. As an example, classifications of soil fracturing and soil corrosivity are shown in Table 6.17 and Table 6.18 respectively.

Table 6.17: Description of soil fracturing keys (National Soil Resources Institute, 2010).

Ground movement key	Fracturing key (description)
1	Very low
2	Low
3	Moderate
4	High
5	Very high

Table 6.18: Description of soil corrosivity keys (National Soil Resources Institute, 2010).

Risk of corrosion to ferrous iron key	Corrosivity risk key (description)
1	Non-aggressive
2	Slightly aggressive
3	Moderately aggressive
4	Highly aggressive
5	Very highly aggressive
6	Impermeable rock

The available soil data contains multiple variables with potentially different relations to pipe repairs. Therefore, the method is required to select which soil variables are valuable and unique and should be proceed further to a descriptive model.

6.4.7 Internal lining

For this research the internal lining data was available for each pipe (if lined). Based on the engineering practice there are two types of internal lining: structural or water quality (see section 2.3.2). This was not clear in the data source. The presence of internal lining for some pipes suggests that repairs have already been occurring and pipes have been lined to enhance their performance. The correlation between pipe repairs and pipes being lined would therefore be expected. There are also other issues associated with lining, which relate not only to when the pipe was lined, such as before or after the pipe failed, but also to the strength of the lining e.g., whether the lining is still intact. This could not be verified and raised questions to the actual data quality. The internal lining is therefore excluded from the analysis.

The data available for this research was assessed and variables suitable for describing pipe repair rates were identified. Top level split of the data by material was implemented and the decision that only cast iron data should be analysed due to dataset size. Target variable included the number of repairs and pipe lengths and was expressed as rate (NoF/len/time). The selected explanatory variables were age, diameter, static pressure and soil variables. Thresholds were applied on pipe length, diameter and static pressure values and units of variables were unified (age converted to days to be consisted with a transient metric, diameter converted to meters to be consistent with rate).

6.5 Forms of cases

This section explores the role of pressure transients with other variables to improve description of observed repair data. Four case types are identified as suitable for the assessment of pressure transient contribution to repair data, as presented in Figure 6.12.

A first case type, i.e. base case (BC) is a comparison/ control type. It contains the historically identified, well established and historically accepted variables: pipe age (A), diameter ($\oslash$) and static pressure (SP). Following case type is a base case + transient (BC+T) that consists of the base case variables and pressure transient data. This case type can be compared with the base case to assess the contribution of pressure transients to observed repair rates. These two case types help to explore and understand pressure transients contribution to the description of piper repair rates.

Soil data was identified as one of the secondary contributing factors (see section 6.4.6) and therefore is not a part of the base case. Soil data formed third case type, i.e. base case + soil (BC+S). This case type allowed comparing with pressure transients to assess whether an improvement can be observed by modelling pipe repairs with soil or transient variables (how much

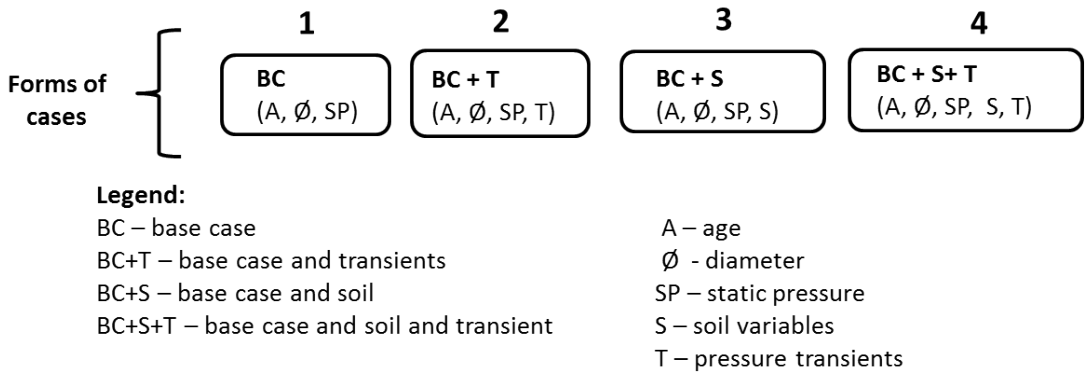


Figure 6.12: Forms of cases designed to explore the role of pressure transients with other variables to improve description of pipe repair data; base case (BC), base case + transient (BC+T), base case + soil (BC+S) and base case + soil + transient (BC+S+T).

better pressure transients would describe pipe repairs when compared to soil data). The final case type, i.e. base case + soil + transient (BC+S+T) allows to evaluate simultaneously the impact of soil and pressure transients data on pipe repairs, explaining if pressure transients and soil interact to cause pipe repairs.

6.6 Modelling methods assessment

This section assesses the modelling approaches describing the relationship between historical pipe repair data and contributing variables. The complex process of determining causes of pipe repairs has been previously investigated through physical (see section 2.4.1), data driven (machine learning, see section 2.4.3) and statistical (see section 2.4.2) modelling. To assess the suitability of the possible approaches/ descriptive models two criteria have been defined:

1. The descriptive model quantifies the historical pipe repair data using the traditionally accepted data, for example static pressure, age, diameter (i.e., base case), and provides a measure of fit.

2. The descriptive model gives the opportunity to add other variables/parameters (e.g., pressure transient categorisations) and helps to understand the contribution they make to pipe repair records (e.g., allow various input data, excess of zeros (non-failed pipes) and discrete data).

In addition, it would be useful if the descriptive model of choice would be easy to apply, interpret and simple to understand (e.g., provide easily readable symbolic expression). This would

provide physical understanding of the outcome and greater confidence with the final interpretation of the results. Other desirable characteristic of the descriptive model is its robustness and practical application (e.g., data can be challenging to collect, which may limit practicality of the potential model).

6.6.1 Physical modelling

One approach to describe the relationship between historical pipe repair records and pressure transients could be through a physical modelling. This approach has been typically based on describing the physical phenomenon/mechanisms that lead to the occurrence of a pipe failure. Physical model (see section 2.4.1) requires knowledge of complete pipe load history which was not available; therefore, this approach could not be used in this research.

6.6.2 Statistical - Regression models

The most established and used approach for the explanation of pipe repairs is through statistical modelling (see section 2.4.2). This modelling approach relies on regression based analysis, e.g., linear regression and non-linear regression. The majority of statistical prediction models are built on multivariate regression or regression-like approaches that evaluate pipe repairs due do various input factors (e.g., pipe age, material, diameter), which do not require a complete history of pipe repairs.

6.6.2.1 Life time models

One family of statistical modelling is life time modelling (see section 2.4.2.1). The suitability of this kind of modelling to assess the contribution of pressure transients to repair data would require a full load history of a pipe and its complete history of repairs. The lack of both, especially the lack of a full load history of the pipe, excludes life time modelling as a possible approach and is therefore not considered further.

6.6.2.2 Generalized linear models (GLMs)

Statistical modelling techniques have been previously applied with variable input data quality and quantity. The most promising groups to assess the possible contribution of pressure transients to the observed repair data were found to be generalized linear models (GLMs). These, however, are limited due to zero-inflation problem as described in section 2.4.2.2.

6.6.3 Bayesian modelling

Bayesian modelling was identified as one of possible approach to evaluate pipe repairs (see section 2.4.2.3). For this approach a prior assumptions are required. However, there is not enough data to derive prior assumption, which could be tested on the independent data, i.e. separated from the data used for the derivation of the prior assumption.

In terms of other variables there is possibly enough data to derive a prior assumptions, but the aim of this work is to estimate pressure transient contribution (with other variables) to the description of pipe repairs, therefore this approach was not considered further.

6.6.4 Artificial neural network (ANN)

The advantage of applying the artificial neural network (ANN), a 'black box' technique, to model pipe repair rates, comes from its ability to model complex non-linear relationships. It has been reported (see section 2.4.3) that these models have a better performance in predicting certain problems in comparison with traditional statistical approaches. However, in the process of understanding what contributes to the pipe repairs ANN may not appear as best choice as they do not readily provide a clearly readable symbolic expression. In addition, the existence of the 'black box' behind the process leads to less comprehensive understanding of the processes behind the analysis. For these reasons ANN were not proceeded further.

6.6.5 Self Organising Maps (SOMs)

SOMs are multidimensional clustering techniques, in the family of machine learning. SOMs can provide visual exploration and insight into correlations between many parameters simultaneously. But the approach provides only a visual output in a form of a map and does not have the ability to form any symbolic expression. SOMs may be useful in a first stage of assessment of the pipe repair data for rapid and visual intuitive investigation of several variables simultaneously.

6.6.6 Evolutionary polynomial regression (EPR)

EPR is a hybrid of genetic programming and statistical regression approaches (see section 2.4.5), often preferable option over other modelling techniques. EPR has all advantages of ANN and in addition provides evolutionary based structured polynomial expressions capable to capture complex and not linear behaviour between variables.

6.6.7 Modelling methods selection

The assessment has shown that for rapid first assessment of interrelationship between variables SOMs are useful. EPR seem to be a good modelling method, which allows further assessment of the contribution of pressure transients to the observed repair rates and their quantification.

Data driven tools, i.e. SOMs can identify which parameters provide additional, unique information and therefore have a greater potential to better describe pipe failure rates. SOMs were identified as a valuable tool, which from the large amount of complex variables can identify which ones do not provide additional information and these can be discarded. The important variables can be therefore identified which can be then taken forward to EPR.

The ability of SOMs for data mining multi-dimensional data sets was used to initially evaluate available variables. Visualisation of interrelationship between various types of variables allowed to identify, which ones provide unique information and therefore are strong explanatory variables. Pressure transient categorisations and soil data need to be refined by SOMs to select best variable choice to describe variability in the data and remove variables that possibly provide repeated information. The selection of pressure transient categorisations is also based on a SOMs outcome and on the assessment of variables to provide differentiation between studied clusters in SOMs. The differences between patterns in corresponding component planes for each variable for the pipe-measured and the zone-level datasets are evaluated to select best pressure transient variable sets.

EPR was identified (see section 6.6.6) as a method, which would provide an equation and quantification of the contribution of pressure transient to repair rates by assessing its goodness of fit by relative measure, coefficient of determination (see section 2.4.5). The EPR MOGA-XL tool version 1.0 was used for the static regression modelling of pipe repair rates. It utilised Multi-Objective Genetic Algorithm (MOGA) optimisation strategy based on the Pareto dominance criteria (Giustolisi and Savic, 2009).

6.7 Summary

This chapter presented the methodology to understand contribution of pressure transients to the observed pipe repair records. The selection of variables into four forms of cases was undertaken based on historically accepted variables and novel pressure transients data. Pressure transient data were categorised based on the rate of change of pressure and transient fingerprints to provide a simple variable input to a modelling method. The rate of change of pressure was categorised by the interpercentile ranges, and transient fingerprints categorised into four regions. The implemented

categorisation focused on highlighting, which characteristics of pressure transients may be important for the assessment of their contribution to the observed repair records.

Two modelling techniques, SOMs and EPR, were identified to assess the contribution of pressure transient categorisations to the observed pipe repair records. SOMs were chosen as the first assessment method for soil and pressure transient categorisations to select best set of variables. EPR fulfilled all the designed criteria and is chosen to provide readable mathematical equations and measure of model fit to quantify the contribution of pressure transients to the observed pipe repair records.

Chapter 7

Results

7.1 Introduction

The aim of this chapter is to present the results from the application of the transient fingerprint categorisation to time series data collected during the fieldwork. The assessment of the contribution of pressure transient variables to the observed pipe repair records is undertaken through the methodology developed in section 6.3. This includes the evaluation of the pressure transient modelling variables through SOMs (to select best set of variables) and then through EPR.

7.2 Occurrence of pressure transients

The field data collection as described in Chapter 4 resulted in unique, high-resolution data (total of 87,360 hours), collected from 67 DMAs (79 PRV zones). During the monitoring program, 144 locations were monitored. The dynamic pressure data was collected from diverse sites to gather sufficient numbers of pressure transient examples which would then allow assessing which characteristic of pressure transients contribute to the observed repair records as there is still unknown what is about pressure transient that causes such contribution.

In this section, the results are presented from applying the method to characterise and count different magnitudes and durations of pressure transient events in the recorded time series over a typical two weeks recording period (see section 4.7). Figure 7.1 and Figure 7.2 show recorded pressure traces and associated transient fingerprint outputs of the characterisation method. The value and ability of the method of characterisation have been demonstrated by application to different pressure traces from diverse zones. Time series pressure traces and associated transient fingerprint outputs of the characterisation method from 144 locations monitored during the fieldwork can be found in Appendix A.

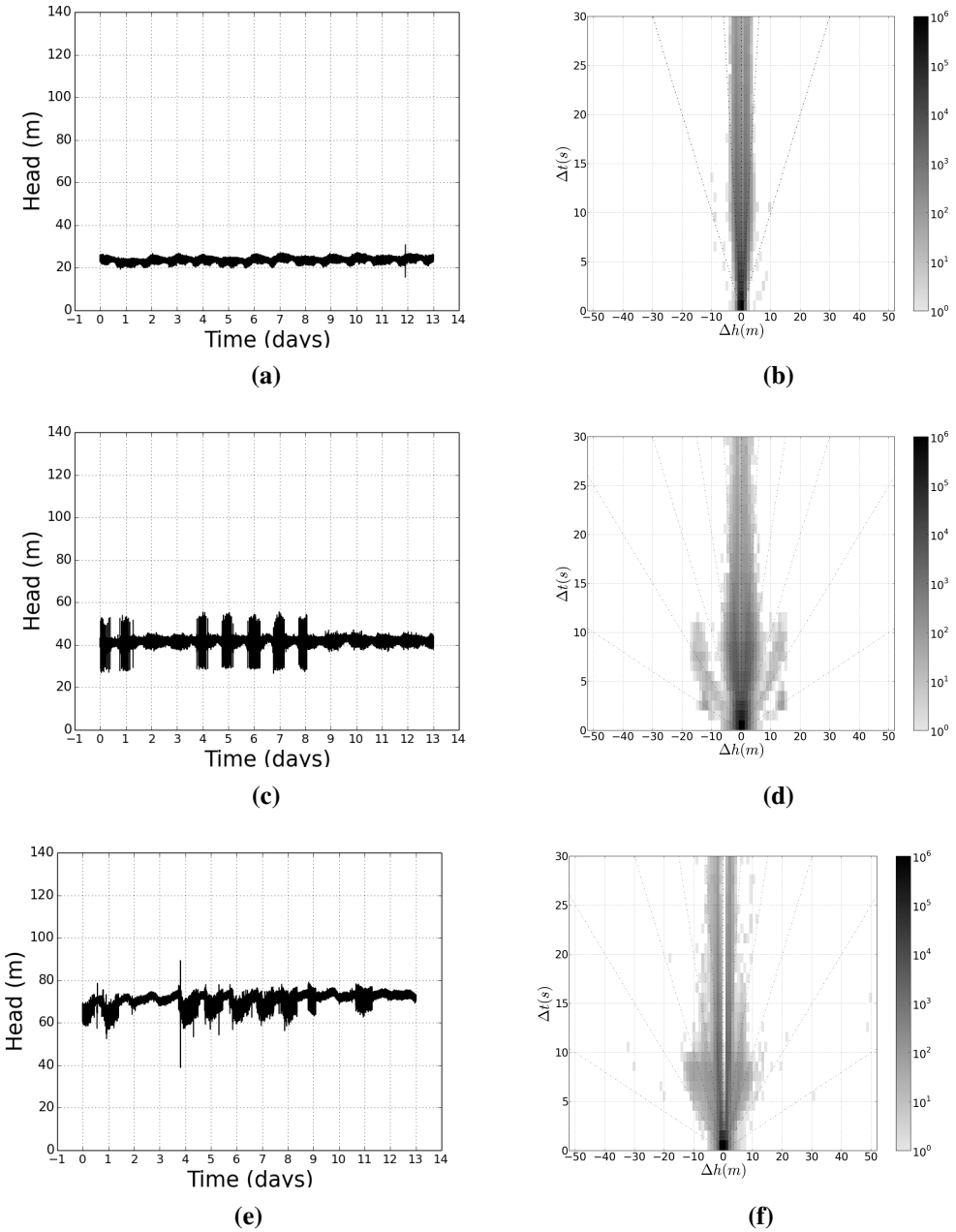


Figure 7.1: Examples of pressure traces and corresponding transient fingerprints recorded during the fieldwork in different zones, (a) pressure trace from the residential zone, (b) its transient fingerprint, (c) pressure trace from the industrial zone, (d) its transient fingerprint, (e) pressure trace from the industrial zone, (f) its transient fingerprint.

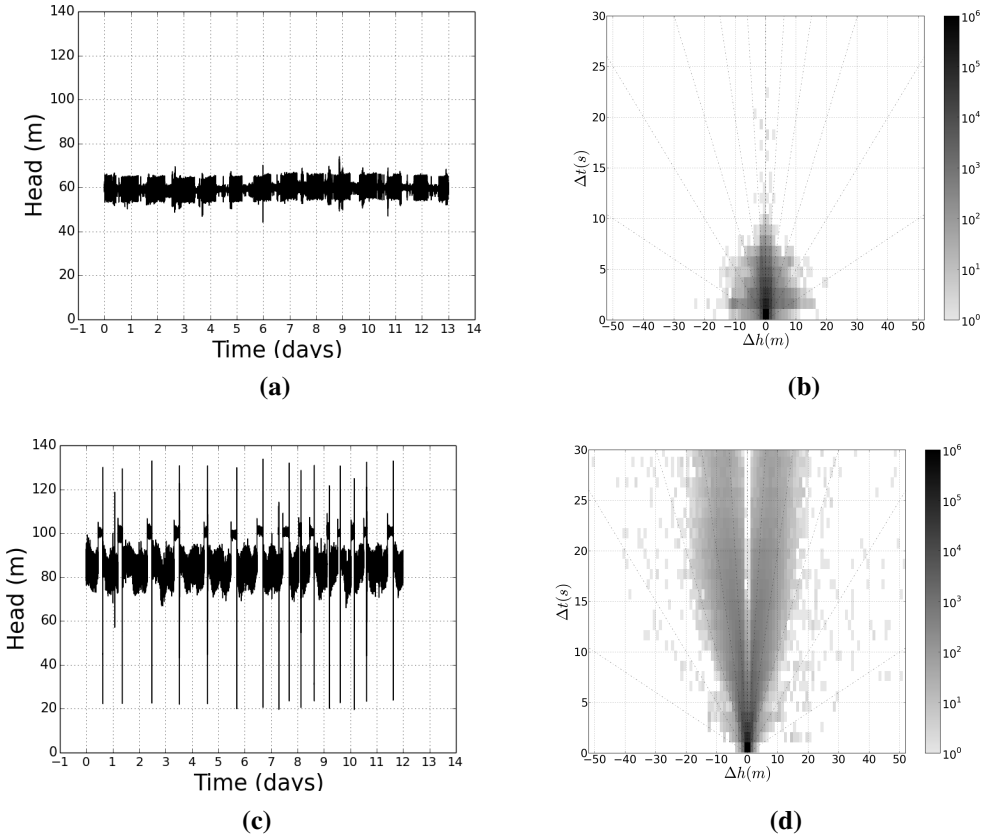


Figure 7.2: Examples of pressure traces and corresponding transient fingerprints recorded during the fieldwork in different zones, (a) pressure trace from the industrial zone, (b) its transient fingerprint, (c) pressure trace from the pump zone, (d) its transient fingerprint.

Overall, the transient fingerprints from 144 locations showed that pressure transients are observed within entire monitored networks. It was not previously expected how widespread this existence can be. The common existence of pressure transients in WDS shows that these should not be considered as static or pseudo-static systems despite the fact that the diurnal patters are seen. In the residential zones the random and occasionally occurring pressure transients were seen, as evidenced in Figure 7.1 (a). The regular pressure transient occurrence was observed in the industrial zones, see Figure 7.1 (c), (e) and Figure 7.2 (a), with large number of low magnitude events. The corresponding plot for the pumped system, see Figure 7.2 (c), is slightly different in shape. Similar to the industrial zone, a large number of events of low magnitude are seen with the apparent regular pressure transients of large magnitude (approximately 40-50m) and durations from 7 to 30 sec. Other pumped systems, see Chapter 5 Figure 5.28 and Figure 5.29, showed an obvious asymmetry to the transient fingerprint with magnitude of downsurge events having a much greater duration than the corresponding magnitude of upsurge events. Transient propagation from diverse sources (i.e. pumps) was observed much further than previously indicated by the literature and through more complex networks.

A high-speed monitoring implemented by this research in WDS has given new insights into

the behaviour of pressure transients. The field monitoring study has shown that transients are, by nature, zone specific and can vary greatly across networks on a zone-level. In all of the zones monitored, transient activity was found. In almost all zones a low magnitude but high-frequency oscillations were found (5-10 m, up to 100 times an hour). Extreme infrequent and small regular scale pressure transients were recorded. Table 7.1 presents an overview of pressure transients recorded during the fieldwork. In 78 zones pressure transients of magnitudes between 5-10 m were recorded, these were followed by 10-20 m events detected in 56 zones and 20-50 m events in 26 zones. Large magnitude events, 50-100 m, were recorded in 5 zones and only one zone had pressure transients of magnitudes $>$ 100 m.

Table 7.1: Number of zones experienced pressure transients of specified ranges.

	0-1 m	1-5 m	5-10 m	10-20 m	20-50 m	50-100 m	>100m
Number of zones	79	79	78	56	26	5	1

Table 7.2 presents a summary of average pressure transient events counts per zone. It can be noticed that some zones had large number of pressure transients, 3.5 times more than the overall average. There were also zones which experienced 6.5 less pressure transient events than the overall average.

Table 7.2: Summary of monitored locations based on transient fingerprints.

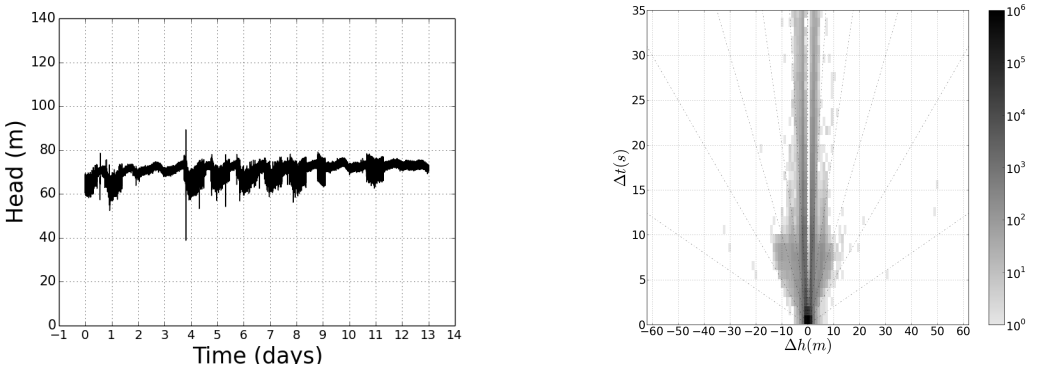
	Sum of all events	Average per zone	Minimum per zone	Maximum per zone
Number of events	391,229,872	2,716,874	416,156	9,380,523

7.3 Locations unique pressure transients

The ability of the transient fingerprint characterisation to distinguish between pressure traces form different locations recorded at the same time within the same zone is demonstrated on the example from the industrial zone where two instruments were installed within the zone.

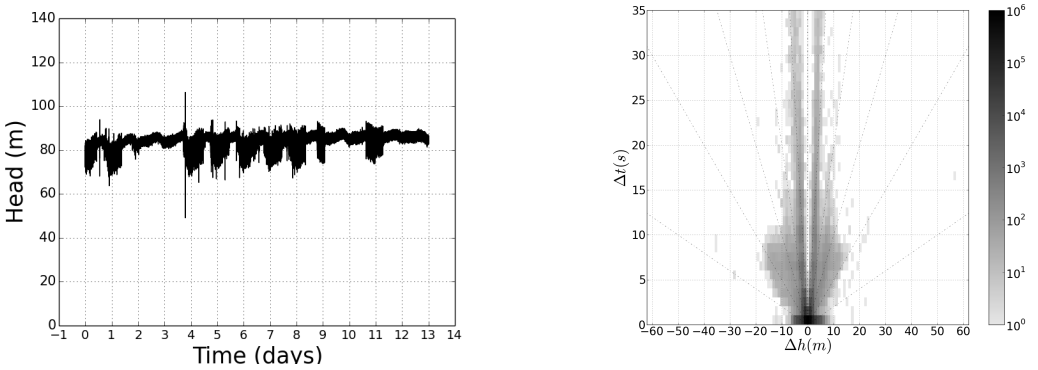
In both plots the dynamics of the pressure response is dominated by pressure transients generated by the industrial user drawing water from the system. Figure 7.3 (a) and Figure 7.4 (a) show pressure traces from this zone. In both plots a diurnal cycle through the 6 days of the working week (Monday to Saturday, days 4 - 9) can be seen. During typical working days the instruments measured a series of repeated pressure transients with magnitudes of approximately $\pm$10 to 15 m

occurring during lower pressure day time hours. In addition, a large magnitude ($\sim 50m$) pressure transients were recorded at the start of the week (end of day 3).



(a) **(b)**

Figure 7.3: Location 1 from the industrial zone (a) the pressure trace, (b) the transient fingerprint.



(a) **(b)**

Figure 7.4: Location 2 from the industrial zone (a) the pressure trace, (b) the transient fingerprint.

The differences between pressure traces recorded in these two locations can be clearly visible in the corresponding transient fingerprints (see Figure 7.3 (b) and Figure 7.4 (b)). Table 7.3 quantify the observed differences by providing number of counts of pressure transient events of different magnitudes.

Table 7.3: Comparison between two locations within the same zone.

	0-1 m	1-5 m	5-10 m	10-20 m	20-100 m	Total number of events
Counts of events in location 1	4,236,421	119,895	4,544	353	8	4,361,221
Counts of events in location 2	2,184,642	930,013	12,354	1326	20	3,128,346

Table 7.3 shows that location 2 experiences slightly larger number of pressure transient events

than location 1. Location 1 experienced twice as many pressure transients of magnitude 0-1 m whereas location 2 eight times more events of magnitude 1-5 m and almost three times more events of magnitudes 5-10 m than location 1. It can also be seen that location 1 is dominated by events of magnitudes 0-1 m and location 2 by events of magnitudes 1-5 m and 5-10 m. The study of two pressure traces recorded at the same time in the same zone showed, that pressure transients are distinct and that different pressure transients can be seen in diverse locations within the same zone.

7.4 Contribution of pressure transients to pipe repairs

This section presents the results from the application of the data modelling methodology described in section 6.3, Figure 6.7. Two sets of data, described as the pipe-measured and the zone-level (see section 6.3) were used as an inputs to SOMs and EPR.

The available variables were grouped into forms of cases, as described in section 6.5. The commonly accepted variables formed a base case to compare with other case types. The base case EPR results are presented first. Case types were then extended to consider the effect of pressure transients and/or soil variables.

The assessment of pressure transients and soil variables through SOMs is presented to select best variable choice. For the selection of variables through SOMs the following assessment criteria were developed:

1. retain variables showing correlation with pipe failure rates,
2. identify and retain variables with unique SOMs patterns,
3. discard variables with the same patterns,
4. retain the lowest value of variable for which the variability in SOMs is observed.

The selected variables are then used as an input to EPR. EPR provided equations quantifying the relationship between the variables and a model fit measure (CoD), as described in section 2.4.5. The contribution of variables to the observed repair rates is assessed by the change in the CoD value with respect to the base case at the same level. In this research an observed change in the value for a sensical relationship is important.

7.5 EPR base case

This section presents variables traditionally associated with causing pipe repairs evaluated through EPR as described in the modelling methodology (see section 6.3). These are undertaken for the

pipe-measured and the zone-level data respectively.

The EPR MOGA-XL tool version 1.0 (Laucelli et al., 2005) was used for the static regression modelling. The intercept was decided to be included in all equations to show whether all data was explained by the variables in the equations by not forcing the equation to go through origin of the plot. Number of terms was set to three. The MOGA process ran for varied number of generations (from 3690 to 11070 depending on the form of cases) as the number of runs was related to the number inputs and outputs and to the structure of the expression (Laucelli et al., 2005). No scaling of the data was used and the regression method for model parameters estimation was non-negative least squares, as recommended for static regression use. The details of graphical user interface are provided in Laucelli et al. (2005).

The chosen expression structure was (Giustolisi and Savic, 2006):

$$Y = a_0 + \sum_{j=1}^{m} a_j \cdot f((X_1)^{ES(j,1)} \cdot \ldots \cdot (X_k)^{ES(j,k)}) \tag{7.1}$$

where Y estimated output of the system, a_0 is an intercept, a_j are polynomial coefficients (i.e. model parameters), m is the number of polynomial coefficients (excluding a_0), ES is the matrix of exponents, X_k is the kth input variable, f is a function selected by the user.

The choice was also made on the the exponents. The choice was to use 0.1 separation between successive values within bounds [-2,2]. The use of such discrete exponent choice was chosen to allow EPR a greater 'freedom' to produce a better model fit. The investigation of the possible bounds showed that there is no obvious limit to restrain the exponent's limits. Therefore, the decision was made, based on the engineering judgement, to bound the exponents within limits from -2 to 2 which was historically accepted (section 2.4.5).

7.5.1 Pipe-measured data

The results from the EPR analysis for a base case are presented in Table 7.4. The table shows all equations returned by EPR. In total, four levels were reported corresponding to three variables and an intercept. It can be seen that each level shows an improvement in a model fit reported as the CoD value (see 2.4.5.1).

Table 7.4: EPR output for the pipe-measured data from a base case.

Level	Equation reported by EPR	CoD value
1	$Rate(1) = 3.14x10^{-6}$	0.016
2	$Rate(2) = 1.72x10^{-9} StPressure^2 + 0$	0.28
3	$Rate(3) = 8.48x10^{-19} Age^2 StPressure^2 + 0$	0.46
4	$Rate(4) = 2.36x10^{-15} Diam^{1.8} Age^{1.6} StPressure^2 + 0$	0.52

The biggest improvement in the CoD value is observed between level 1 and level 2 (17.5 times). Improvement from level 2 to level 3 and from level 3 to 4 is 1.64 and 1.13 times respectively. Static pressure appears in all equations, age in the last two and diameter only in the last one. According to the equations generated by EPR, the variable importance can be ranked in descending order as static pressure, age and diameter. All variables are in positive relationship with pipe repair rates.

7.5.2 Zone-level data

Table 7.5 presents all equations returned by EPR for the zone-level data. In total, four equations were produced, as shown in the table.

Table 7.5: EPR output for the zone-level data from a base case.

Level	Equation reported by EPR	CoD value
1	$Rate(1) = 8.87x10^{-7}$	0.00013
2	$Rate(2) = 2.54x10^{-7}\frac{1}{Diam^{0.6}} + 0$	0.0019
3	$Rate(3) = 3.97x10^{-9}\frac{Age^{0.4}}{Diam^{0.6}} + 0$	0.0023
4	$Rate(4) = 1.84x10^{-9}\frac{Age^{0.4}StPressure^{0.2}}{Diam^{0.6}} + 0$	0.0024

In comparison with the pipe-measured data, static pressure did not appear as a first variable. The most important variable is diameter (observed in all equations) and subsequently age (last two equations). Static pressure appears in the last level equation. The biggest improvement in the CoD value is observed between level 1 and level 2 (14.62 times). Improvement from level 2 to level 3 and from level 3 to level 4 is 1.21 and 1.04 times respectively. According to the equations generated by EPR, variable importance can be ranked in descending order as diameter, age and static pressure. Positive relations to pipe repair rates is seen for age and static pressure whereas diameter shows an inverse relationship.

EPR results for the pipe-measured and the zone-level data returned four levels of equations. This suggests that additional variables are required to better describe a target data.

7.6 SOMs rate of change of pressure

This section presents the evaluation through SOMs of the categorisation of the rate of change of pressure to select the best choice that can be an input to the EPR. Classifications of the rate of change of pressures were developed in section 6.2.1.1 that included four interpercetnile ranges: 98th, 99.8th, 99.98th and 99.998th. These are evaluated and compared through SOMs according to the criteria described in section 7.4. The selected choice is then taken forward to the EPR.

SOMs are techniques that group, 'map' data comprised of different information/features (called also input space, usually higher dimensional) into a different areas in a two-dimensional lattice (output space, also known as output neurons) representing a topographic map. The features from data are mapped in a topologically ordered manner where a location in the topographic map corresponds to a particular feature of data (see also 2.4.4). The topographic map is successively built through self-organisation process. During this process all neurons in the lattice are exposed to a sufficient number of different input features to ensure that self-organisation has a chance to mature properly. This is also known as a learning (or training) process. After the learning process is completed the input data is mapped to a trained neurons and the features from data are represented in the two-dimensional topographic map. This is also known as a 'mapping' process. Each neuron in the topographic map can be also seen as an analogy to a cluster seen in k-means clustering. For more details on SOMs algorithm see Kohonen (1982).

To explain how SOMs work the example of the football stadium full of people is used. Each person in the stadium has some attributes, such as colour of hair, heights etc. The people were asked to ask each other questions and if they have things in common join into groups. The people kept asking each other questions and kept grouping until the stadium became full of clusters. This represents self-organisation learning process of SOMs algorithm. When the self-organisation process is completed the people are asked to put a red colour if they have the same foot size, i.e. something they did (or did not) cluster on. This is now a mapping process which forms the view across all different attributes at a space. To illustrate it a simple example of mapping process that considers two features is shown in Figure 7.5. The figure shows that in this particular stadium people of large foot size were of medium height, see the top left corners of the foot size and height map where the arrow is drawn. The arrow helps to illustrate how different attributes are related to each other in a space. SOMs are therefore techniques which map objects of the same or similar characteristic/attributes into groups. A location in a topographic map corresponds to a particular attribute of data and provides an assessment on how similar these groups are.

The self-organising map consists of components called neurons (or nodes). Figure 7.5 presents

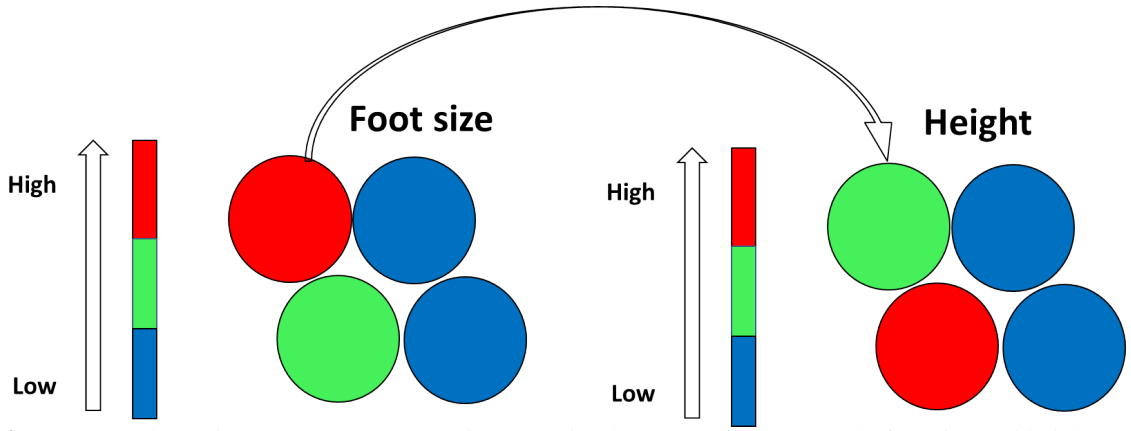


Figure 7.5: Schematic SOMs map representing grouping by two attributes: people foot size and height. It can be seen that people with large foot sizes has also medium heights.

a two-dimensional lattice of neurons. It comprises of four neurons in the 2 x 2 space. The number of neurons (i.e. space size) are usually defined by the user with the condition that the space should be arranged in such a way that empty neurons should not occur. The empty neurons indicate that the map size is too big for available number of samples in the data. Frequently occur features will have larger number of samples in the neuron.

The ordering of the SOMs during the learning process is based on competitive learning during which the output neurons compete among each other. The learning process is controlled by parameters known as learning-rate and neighbouring function (Wehrens and Buydens, 2007). These two parameters pay role in the map convergence during the ordering of SOMs phase. The learning-rate parameter should gradually decrease with time, but remain above 0.01 to prevent configurations with a topological defect from occurring. The second parameter, neighbouring function, represents spatial location (size) of a topological neighbourhood of neurons. This parameter of the SOM algorithm initially includes all neurons in the network and then shrinks with time to finally reduce to one or zero neighbouring neurons. During the learning process the neurons are associated with weight vectors in the sense that neurons adjacent in the lattice will have similar weighs. Feature from a data space is placed in the map by finding the node with the closest weight vector (smallest distance) to the data space. These weights correspond to Euclidean distances which can be also seen as an optimal dissimilarity measure. The neuron weights and neighbour distances can be plotted to observe and assess the SOMs learning process.

For the rate of change of pressure the map size was determined to be 3 x 4 (12 neurons) by performing tests on different space sizes and care was taken to avoid the occurrence of empty neurons. This was achieved using the following SOMs parameters values: SOM function has been used with a number of iterations (rlen) (the number of times the data set will be presented to the map) at the default value 100. The learning rates were from 0.05 to 0.01, and these values enabled the map convergence. Figure 7.6 and Figure 7.7 present the outcome of this process: SOMs space

3 x 4 with 12 neurons. Figure 7.6 (a) Figure 7.7 (a) show node (neurons) counts. The node counts comprise of neurons of different colours. Each colour represents a number of samples assigned (i.e. mapped) to a particular neuron. To explore how strong the clusters are a node quality/distance plot can be assessed, as seen in Figure 7.6 and Figure 7.7 (b) respectively. This plot shows the quality of the samples assigned, represented by a mean distances between units, assigned to a particular cluster. A good assignment should show small distances in each place on a SOMs map and should also identify distinct nodes. It can be seen that there is one distinct node (corresponding to high pipe repair rates) with no empty nodes. Figure 7.6 (c) and Figure 7.7 (c) show the SOMs neighbour distance (also known at the 'U-Matrix') between each node and the neighbour nodes. Light grey colours indicate similar nodes. Dark grey shows nodes that are dissimilar. Figure 7.6 (d) and Figure 7.7 (d) show weight vectors across the map (all dimensions in one diagram). Each weight vector represents the samples mapped to that node. In these diagrams the distribution of samples and variables can be seen. Similar analyses were performed for the 99.8th, 99.98th and 99.998th interpercentile ranges and their SOMs maps are presented in Appendix B.

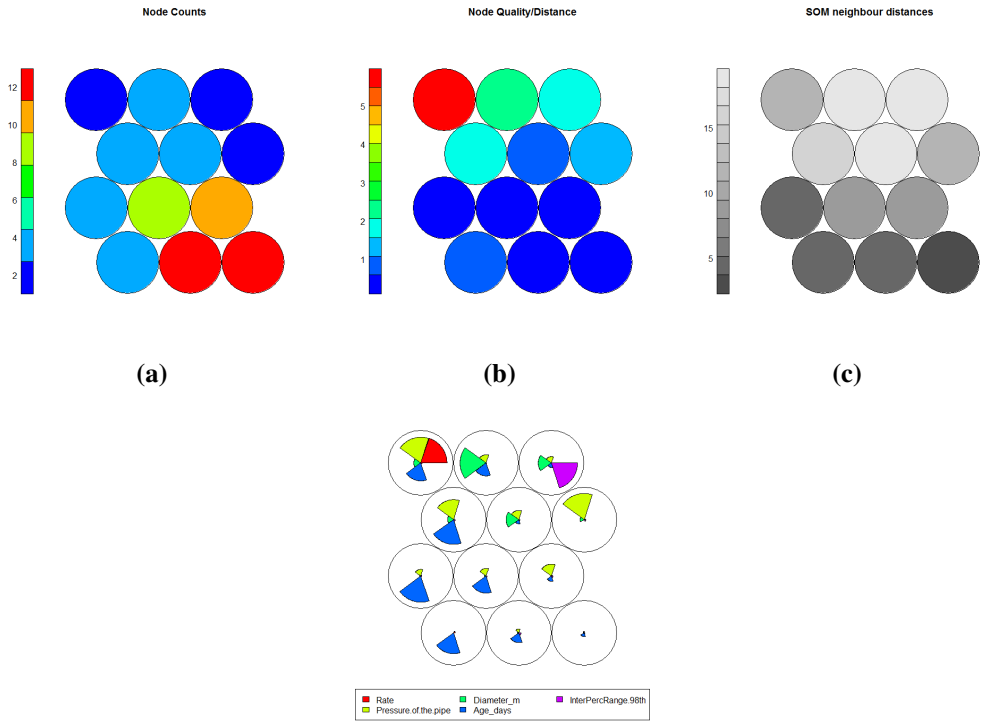


Figure 7.6: Preliminary pipe-measured SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

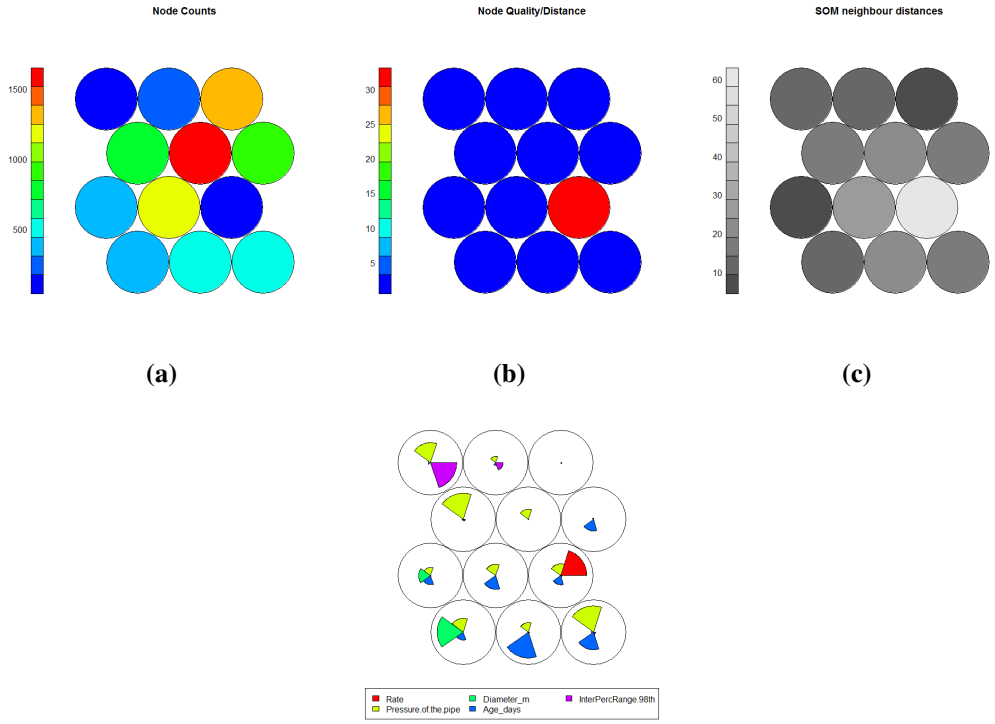


Figure 7.7: Preliminary zone-level SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

Once the map size was determined the analyses were carried out to evaluate each categorisation of the rate of change of pressure to select the best choice to be an input to the EPR by assessing their topographic maps, presented as a heat maps. Multiple heat maps are produced allowing the comparison between variables and identify interesting areas on the topological map. This allows the visualisation of the distribution of a single variable across the map to be assessed and in this respect it is more useful than the visualisation in Figure 7.6 (d) and Figure 7.7 (d) showing all variables on one diagram. Figure 7.8 and Figure 7.9 present the heat maps of the 98th interpercentile range for the pipe-measured and the zone-level data respectively.

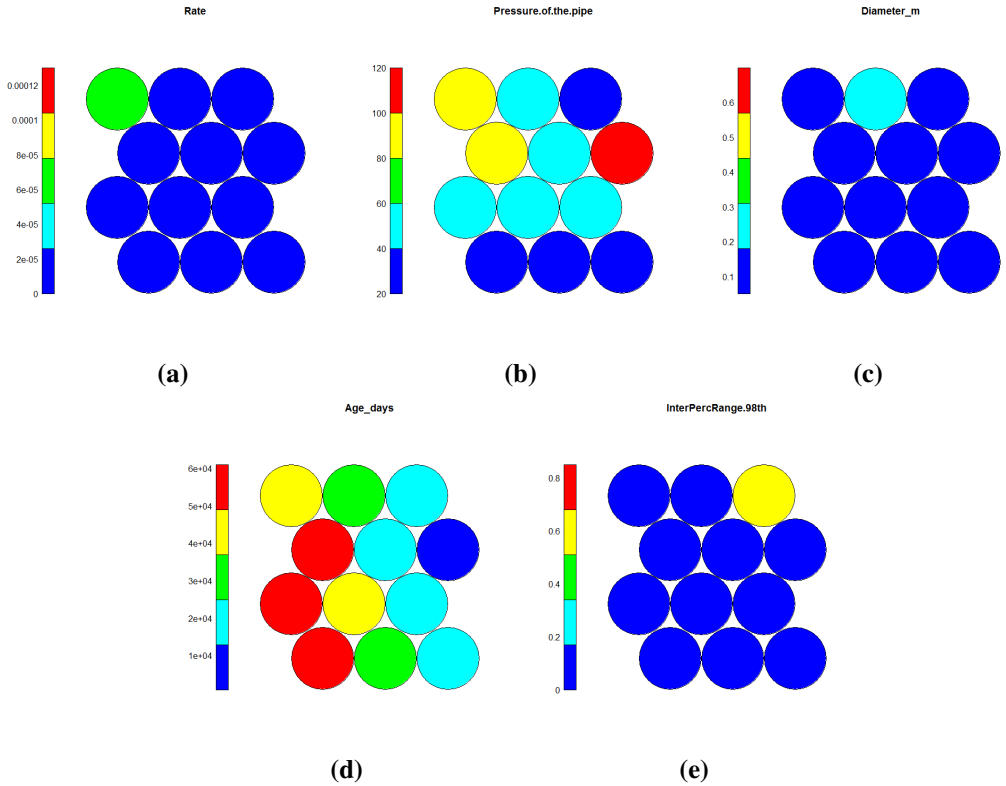


Figure 7.8: Pipe-measured SOMs output (a) pipe repair rate, (b) hydraulic static pressure, (c) diameter, (d) age, (e) the 98th interpercentile range of the rate of change of pressure.

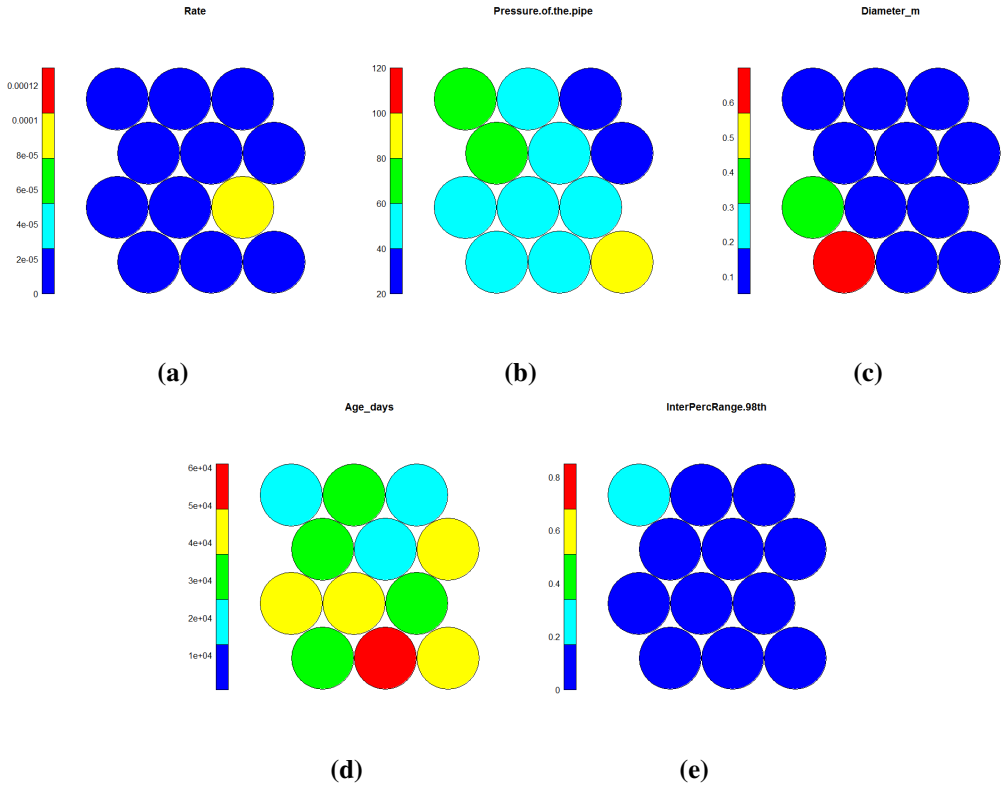


Figure 7.9: Zone-level SOMs output (a) pipe repair rate, (b) hydraulic static pressure, (c) diameter, (d) age, (e) the 98th interpercentile range of the rate of change of pressure.

Similar maps were obtained for the the 99.8th and the 99.98th interpercentile ranges, for the pipe-measured and the zone-level data respectively. All these maps are presented in Appendix B. These were then assessed according to the SOMs assessment criteria as described in section 7.4:

1. retain variables showing correlation with pipe failure rates,
2. identify and retain variables with unique SOMs patterns,
3. discard variables with the same patterns,
4. retain the lowest value of variable for which the variability in SOMs is observed.

According to the SOMs assessment criteria the smallest value for which variability in the SOMs map is observed should be retained. The 98th interpercentile range for the pipe and zone-level data show no variability and correlation of this variable with pipe repair rates, see Figure 7.8 (a) and (e) for the pipe-measured data and Figure 7.9 (a) and (b) for the zone-level data. The same was observed for the 98th interpercentile range (the details of the maps can be find in Appendix B). The variability was observed in the 99.98th interpercentile range and the apparent correlation with pipe repair rates (for details see Appendix B). The 99.998th interpercentile range showed even further an increase in variability and correlation with pipe repair rates was also observed. According to the SOMs assessment criteria the smallest value for which variability in the SOMs map is observed should be retained. The variability was already observed in the previous variable, the 99.98th interpercentile range, therefore the 99.998th interpercentile range is not selected and the 99.98th interpercentile range is taken forward to EPR.

7.7 EPR rate of change of pressure

This section presents the evaluation through EPR of the 99.98th interpercentile range selected during the SOMs analysis as the best variable choice. The evaluation through EPR produces series of equations. Equations that were not previously reported in the base case are highlighted in blue in all tables produced. The change in the CoD, relative to the base case, is reported in the last column and shows the contribution of a variable to the observed pipe repair rates.

7.7.1 Pipe-measured data

Table 7.6 presents the EPR output for the 99.98th interpercentile range for the pipe-measured data. In total, five equations were reported by EPR. This corresponds to five variables being an input to the model. Equations in levels from 1 to 4 are identical to the base case (see Table 7.4). The 99.98th interpercentile range is reported in the last level as a positive variable that provides no improvement in the model fit (with respect to the base case at the same level). The order of importance of the variables observed in the base case (e.g., static pressure, age and diameter) is retained. This is consistent with the SOMs analysis which showed that stronger correlations with repair rates was observed for the BC variables.

Table 7.6: EPR output from the pipe-measured data with the 99.8th interpercentile range. New equations, compared to a base case, highlighted in blue; in black equations reported in the base case EPR.

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 3.14x10^{-6}$	0.016	1
2	$Rate(2) = 1.72x10^{-9} StPressure^2 + 0$	0.28	1
3	$Rate(3) = 8.48x10^{-19} Age^2 StPressure^2 + 0$	0.46	1
4	$Rate(4) = 2.36x10^{-15} Diam^{1.8} Age^{1.6} StPressure^2 + 0$	0.52	1
5	$Rate(5) = 1.38x10^{-15} Diam^{1.6} Age^{1.6} StPressure^2$ $99.98 InterPercRange^{1.1} + 0$	0.52	–

7.7.2 Zone-level data

Table 7.7 shows results from the application of the 99.98th interpercentile range to EPR for the zone-level data. The 99.98th interpercentile range improves the model fit (relative to the base case at the same level) by 1.09 in level 3 and by 1.17 in level 4 respectively. This variable is

also in inverse relationship with pipe repair rates. In comparison to the base case the 99.98th interpercentile range appears to contribute to the overall pipe repair rates replacing age in level 3 (seen in BC) and static pressure in level 4 (seen in BC) respectively. This is consistent with the SOMs analysis that showed stronger correlation of this variable with pipe repair rates than pipe age and static pressure. In comparison to the base case the order of importance of variables changed from diameter, age, static pressure to diameter, the 99.98th interpercentile range and age.

Table 7.7: EPR output from the zone-level data for a base case model and the 99.8th interpercentile range. New equations, compared to a base case, highlighted in blue; in black equations reported in the base case EPR.

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 8.87x10^{-7}$	0.00013	1
2	$Rate(2) = 2.54x10^{-7}\frac{1}{Diam^{0.6}} + 0$	0.0019	1
3	$Rate(3) = 1.25x10^{-7}\frac{1}{Diam^{0.7}99.98InterPercRange^{0.3}} + 0$	0.0025	1.09
4	$Rate(4) = 1.95x10^{-9}\frac{Age^{0.4}}{Diam^{0.7}99.98InterPercRange^{0.3}} + 0$	0.0028	1.17
5	$Rate(5) = 2.39x10^{-9}\frac{Age^{0.4}StPressure^{0.3}}{Diam^{0.7}99.98InterPercRange^{0.3}} + 0$	0.0031	–

7.8 SOMs transient fingerprint

This section presents the results of the SOMs analysis undertaken to decide, which transient fingerprint categorisation into regions (see section 6.2.2.5) should be taken forward to EPR. Four categorisations were identified, i.e. 85%, 90%, 95% and 99%. From these choices SOMs are used to select categorisation that starts to identify separations between regions (e.g., differences between each region).

The similar analyses were carried out as per methodology followed in section 7.6. The SOMs maps can be found in Appendix B. The results have shown that neither the 99% transient categorisation or the 95% transient categorisation produced the separation between regions of the data nor provided any additional information. Therefore both categorisations are rejected.

The 90% transient categorisation shows separations/ differences between each region. For this dataset with SOMs identifying separations between regions variabilities are indicated, e.g, information started to reveal at the level of 90%. This set is, therefore, selected and subjected to the EPR analysis. The 85% transient categorisation provided further considerable variability and

according to the developed assessment criteria, the 85% transient categorisation is rejected and the 90% transient categorisation is taken forward to EPR.

7.9 EPR transient fingerprint

This section presents the EPR analysis for a 90% transient categorisation selected during a SOMs analysis. The results for the pipe-measured and the zone-level data are presented and compared to the base case. Equations which were not previously reported in the base case are highlighted in blue.

7.9.1 Pipe-measured data

In total, eleven equations were generated by EPR. This is one of the examples for which all equations returned by EPR are not presented. The first six equations are in Table 7.8. Beyond this level the equations provided extra complexity without improving the model fit and for these reasons these are not presented. Static pressure is reported in all equations, indicating importance of this variable. Levels 1 and 2 are identical as the base case (see Table 7.4). Region 3 (i.e. pressure transient categorised in region 3) appears from level 3 onwards replacing age (seen in the base case level 3) (see Table 7.4). Region 3 is reported in the denominator in almost all equations from level 3, apart from level 4 where the variable is not reported. $\frac{1}{Region3}$ provides an improvement in model fit by 1.13 with respect to the base case at the same level. Region 1 and region 2 appear together in a form $\frac{Region1}{Region2}$ in level 4. These variables provided an improvement in model fit by 1.1 with respect to the base case. Region 4 is not reported in any EPR equation. Age did not appear within the first five levels.

Pressure transients categorised in region 1 (i.e. fast and large) appeared to contribute to the overall failure rates despite the fact that this group usually represents the least number of counts. The fact that pressure transient variables provided an improvement in model fit suggests that these variables may be important factors contributing to the pipe repair rates.

7.9.2 Zone-level data

Seven equations were reported by EPR. All are presented in Table 7.9.

Table 7.8: EPR output for the pipe-measured data from a base case. New equations, compared to base case, highlighted in blue, in black equations already reported in the base case EPR.

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 3.14x10^{-6}$	0.016	1
2	$Rate(2) = 1.72x10^{-9} StPressure^2 + 0$	0.28	1
3	$Rate(3) = 2.74x10^{-7} \frac{StPressure^2}{Region3^{0.7}} + 1.16x10^{-7}$	0.52	1.13
4	$Rate(4) = 3.15x10^{-7} \frac{StPressure^2 Region1^{1.2}}{Region2^2} + 1.49x10^{-6}$	0.57	1.1
5	$Rate(5) = 8.96x10^{-5} \frac{StPressure^2 Region1^2}{Region2^2 Region3^{1.7}} + 1.85x10^{-6}$	0.59	–
6	$Rate(6) = 5.94x10^{-9} \frac{Age^{1.1} StPressure^2 Region1^{1.9}}{Region2^2 Region3^2} + 1.97x10^{-6}$	0.59	–

Table 7.9: EPR output from the zone-level data for the 90% transient categorisation. New equations, compared to a base case, highlighted in blue, in black equations already reported in the base case EPR.

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 8.87x10^{-7}$	0.00013	1
2	$Rate(2) = 2.54x10^{-7} \frac{1}{Diam^{0.6}} + 0$	0.0019	1
3	$Rate(3) = 3.97x10^{-9} \frac{Age^{0.4}}{Diam^{0.6}} + 0$	0.0023	1
4	$Rate(4) = 4.47x10^{-8} \frac{Age^{0.4}}{Diam^{0.6} Region4^{0.2}} + 0$	0.0024	1
5	$Rate(5) = 4.92x10^{-9} \frac{Age^{0.3} StPressure^{0.4}}{Diam^{0.6} Region1^{0.1}} + 0$	0.0026	–
6	$Rate(6) = 1.09x10^{-7} \frac{Age^{0.4} Region2^{0.1}}{Diam^{0.7} Region1^{0.1} Region4^{0.3}} + 0$	0.0028	–
7	$Rate(7) = 4.36x10^{-8} \frac{Age^{0.4} StPressure^{0.3} Region2^{0.1}}{Diam^{0.6} Region1^{0.1} Region4^{0.3}} + 0$	0.003	–

Levels 1, 2 and 3 are identical as a base case, see Table 7.5. Pressure transient variable appears in level 4 in a form $\frac{1}{Region4}$. It provides no improvement in the model fit with respect to a base case at the same level. In this level static pressure was observed in the BC (see Table 7.5), here static pressure is replaced by region 4. Static pressure appears in levels 5 and 7. Region 4 is replaced by region 1 in level 5. Region 1 is in the denominator from level 5 onwards. Region 3 did not appear in any EPR equation.

7.10 SOMs soil data

Series of soil variables were available for the analysis (i.e., soil class, type, corrosivity, fracturing and workability, see 6.4.6). Based on the visual/qualitative assessment of SOMs and background knowledge the common subset of soil variables is selected and subjected to the EPR analysis.

7.10.1 Pipe-measured data

Figure 7.10 presents evaluation of soil variables through SOMs for the pipe-measured data. From this evaluation the decision is made on which soil variables should be selected and subjected to further analysis, by looking for similar and unique patterns in soil variables and their correlations with pipe repair rates. The initial SOMs variables assessment for the pipe-measured data is presented in Appendix B.

Figure 7.10 shows that there are similarities between soil class and soil type. These variables appear to be similar and they may convey the same information. In contrary, soil corrosivity is a unique variable and therefore it is selected. The next two variables, soil fracturing and workability, appear to be similar to each other. Therefore, which one should be selected is based on the background knowledge.

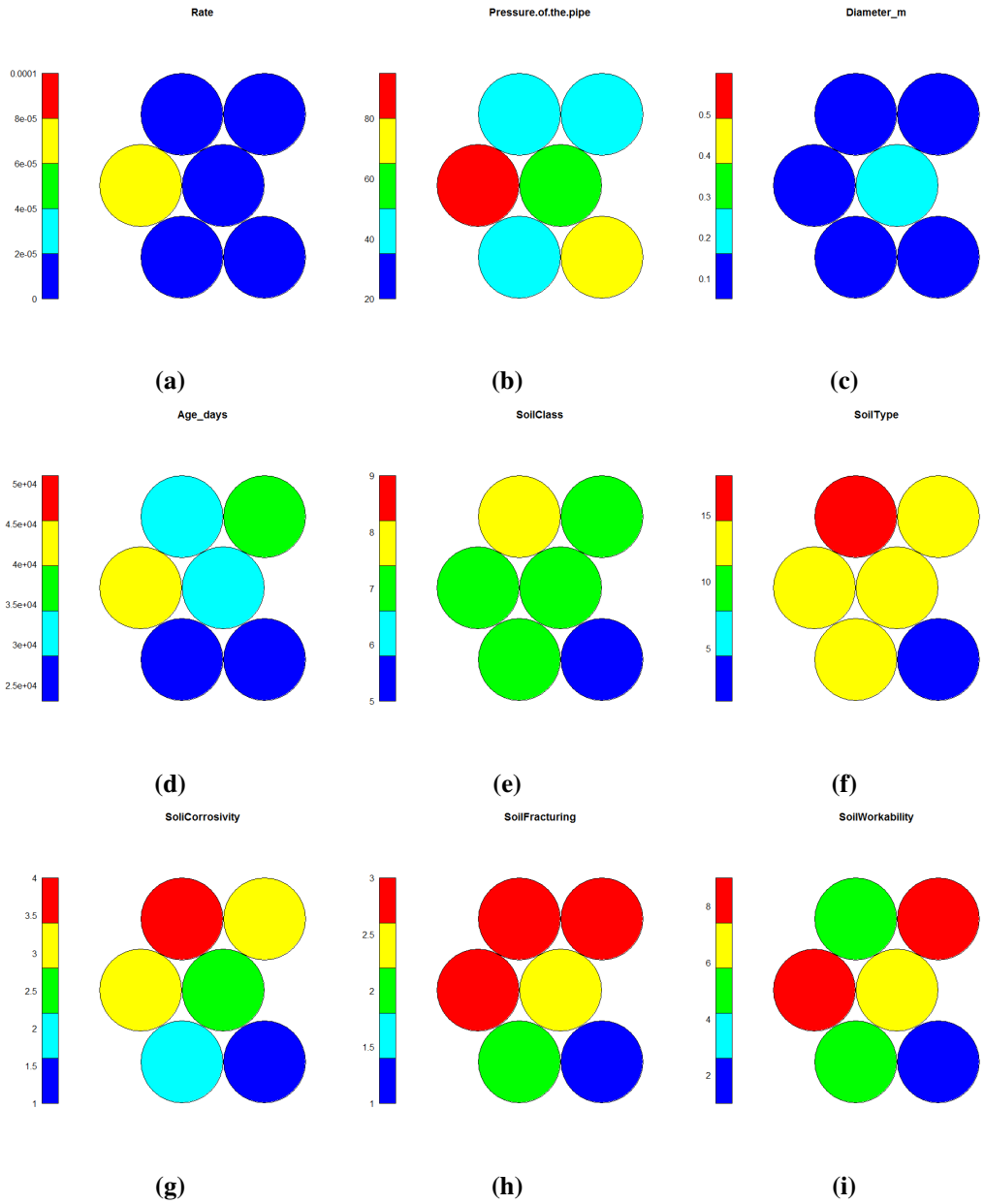


Figure 7.10: Pipe measured SOMs output (a) pipe repair rate, (b) hydraulic static pressure, (c) diameter, (d) age, (e) soil class, (f) soil type, (g) soil corrosivity, (h) soil fracturing, (i) soil workability.

7.10.2 Zone-level data

Figure 7.11 presents soil variables evaluated through SOM for the relationship with a base case variables.

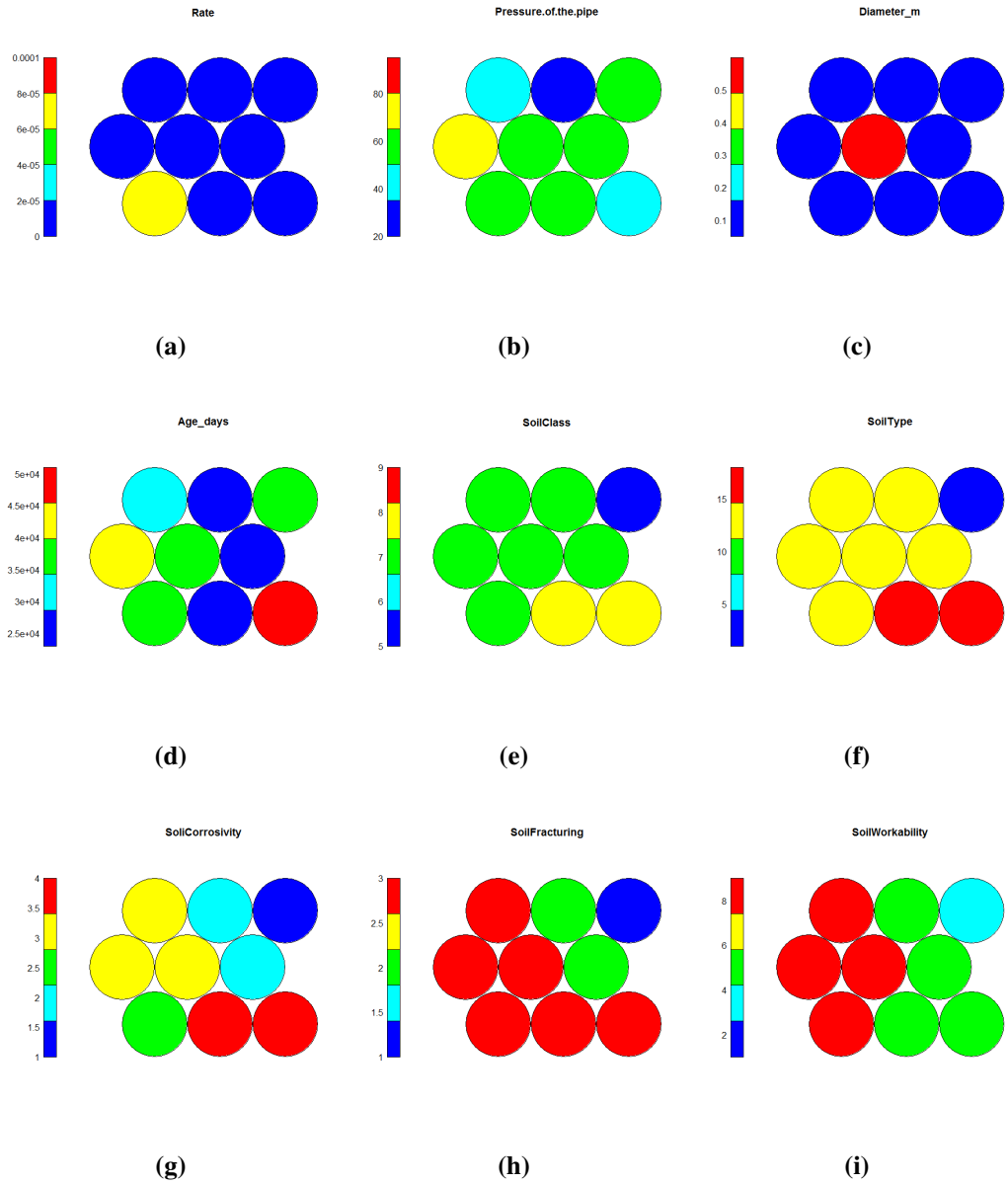


Figure 7.11: Zone level SOMs output (a) pipe repair rate, (b) hydraulic static pressure, (c) diameter, (d) age, (e) soil class, (f) soil type, (g) soil corrosivity, (h) soil fracturing, (i) soil workability.

Figure 7.11 shows similarities between soil class and soil type. Some similarities are also observed between soil fracturing and workability. Soil corrosivity shows an unique relationship.

Comparison of both the pipe-measured (Figure 7.10) and the zone-level (Figure 7.11) datasets was undertaken to reveal any similarities seen in both cases. This was because the same set of variables should be taken forward to the next level of analysis. SOMs obtained from the pipe-measured and the zone-level show that soil class, type are very similar. In addition, soil

class and type are categorical variables (nominal variables) not conveying information. Therefore, these were decided not to be proceed further. Soil corrosivity shows to be a unique variable. Soil corrosivity is a ordinal variable which is ordered according to its severity. In this research cast iron pipes are investigated, which may favour corrosion as a potential informative variable. Furthermore, historically corrosion has been shown to be an important factor in ferrous pipe failure analysis (see 2.3.2). Based on these findings soil corrosivity is selected and subjected to the next level analysis.

Soil fracturing and workability are also ordinal variables organised from the smallest to the largest according to their impact. According to their interpretation (see section 2.3.2) soil fracturing is a natural soil movement due to moisture content and seasonal changes. Soil workability relates to how much strength is needed to move soil. According to the literature review more pipe repairs were observed during cold seasons (winter) caused by the soil fracturing due to frost heave (see section 2.3.2). Using a combination of SOM outcome and background literature knowledge the decision was made that from these two variables fracturing is taken forward to the EPR model.

Other relationships observed in SOMs analyses were with static pressure. High static pressures correlated with high failure rates for the pipe-measured data (see Figure 7.10 (a) and (b)). For the zone-level data this relationship was not observed. High age was correlated with pipe failure rates for the pipe-measured data (see Figure 7.10 (a) and (d)). In the zone-level dataset this relationship was not clear. Correlation with small size pipe diameter with pipe failure rates was seen in both datasets (see Figure 7.10 (a) and (c) and Figure 7.11 (a) and (c)).

7.11 EPR soil data

This section presents the EPR analyses for a base case and soil variables (BC+S) that were selected during the SOMs analysis. The results for the pipe-measured and the zone-level data are presented and compared to the base case respectively. Equations which were not previously reported in the base case are highlighted in blue.

7.11.1 Pipe-measured data

In total, six equations were returned by EPR. Table 7.10 shows all equations reported by EPR for BC+S, i.e. base case and two additional soil variables (soil corrosivity and fracturing) identified during the SOMs analysis. First four levels present equations previously obtained during the evaluation of the base case (Table 7.4). Table 7.10 shows that soil fracturing appears as a positive variable in level 5. Soil corrosivity appears in level 6 in the denominator. Both variables did not provide any improvement in model fit with respect to the base case.

Table 7.10: EPR output for the pipe-measured data from a base case with soil (BC+S). New equations, compared to a base case, highlighted in blue, in black equations already reported in the base case EPR.

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 3.14x10^{-6}$	0.016	1
2	$Rate(2) = 1.72x10^{-9} StPressure^2 + 0$	0.28	1
3	$Rate(3) = 8.48x10^{-19} Age^2 StPressure^2 + 0$	0.46	1
4	$Rate(4) = 2.36x10^{-15} Diam^{1.8} Age^{1.6} StPressure^2 + 0$	0.52	1
5	$Rate(5) = 8.87x10^{-14} Diam^{1.9} Age^{1.1} StPressure^2 SoilFrac^{1.9} + 0$	0.54	–
6	$Rate(6) = 1.7x10^{-13} \frac{Diam^{1.6} Age^{1.1} StPressure^2 SoilFrac^2}{SoilCorrosivity^{1.2}} + 0$	0.54	–

Variables traditionally associated with causing pipe repairs appeared in the first four levels (Table 7.10). The addition of the soil variables did not show to be more important (no improvement in model fit) than static pressure, age and diameter.

7.11.2 Zone-level data

Table 7.11 shows all equations reported by EPR during the evaluation of a base case and soil variables (BC+S) for the zone-level data.

Table 7.11: EPR output for the zone-level data from BC+S. New equations, compared to a base case, highlighted in blue; in black equations already reported in the base case EPR.

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 8.87x10^{-7}$	0.00013	1
2	$Rate(2) = 2.54x10^{-7} \frac{1}{Diam^{0.6}} + 0$	0.0019	1
3	$Rate(3) = 1.05x10^{-7} \frac{SoilFracturing^{0.7}}{Diam^{0.7}} + 0$	0.0025	1.09
4	$Rate(4) = 1.14x10^{-7} \frac{SoilFracturing^{2}}{Diam^{0.7} SoilCorrosivity^{1.3}} + 0$	0.0034	1.42
5	$Rate(5) = 6.9x10^{-10} \frac{Age^{0.5} SoilFracturing^{2}}{Diam^{0.7} SoilCorrosivity^{1.4}} + 0$	0.0039	–
6	$Rate(6) = 6.21x10^{-10} \frac{Age^{0.4} StPressure^{0.3} SoilFracturing^{2}}{Diam^{0.7} SoilCorrosivity^{1.4}} + 0$	0.0041	–

Soil fracturing appears as a positive variable in level 3 onwards. This shows the improvement in model fit by 1.09 in comparison to the base case at the same level. Soil fracturing replaced age at level 3 (seen in BC at the same level). Soil corrosivity appears in three last levels. This variable improves the model fit with respect to the base case at the same level by 1.42. Age is reported from level 5 in two last equations. Static pressure appears only in the last level.

Base case variables showed that diameter is the important variable. However, the addition of soil data shows that soil variables are more important than age and static pressure, as these showed an improvement in model fit with respect to the base case at the same level. The fact that soil variables replaced age and appeared in high levels may suggest that for this data soil variables are important factors. These results are in contrary to the pipe-measured data, which showed that soil varialbes provided no improvement in model fit.

7.12 EPR rate of change of pressure and soil

Previous analyses investigated a base case variables and the transient fingerprint (BC+T) and a base case variables and soil parameters (BC+S). This section compare soil variables and the rate of change of pressure simultaneously in one case type (BC+S+T) evaluated through EPR. This is investigated to show whether the rate of change of pressure may interact with soil.

7.12.1 Pipe-measured data

In total, seven equations were reported by EPR. All equations are reported in Table 7.12. Levels from 1 to 4 were previously reported in a base case (BC), see Table 7.4. In addition, Levels 5 and 6 were previously reported in a base case and soil (BC+S), see Table 7.10.

Table 7.12: EPR output from pipe-measured data for soil and the 99.98th interpercentile range. New equations, compared to a base case, highlighted in blue, in black equations already reported in the base case EPR.

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 3.14x10^{-6}$	0.016	1
2	$Rate(2) = 1.72x10^{-9} StPressure^2 + 0$	0.28	1
3	$Rate(3) = 8.48x10^{-19} Age^2 StPressure^2 + 0$	0.46	1
4	$Rate(4) = 2.36x10^{-15} Diam^{1.8} Age^{1.6} StPressure^2 + 0$	0.52	1
5	$Rate(5) = 8.87x10^{-14} Diam^{1.9} Age^{1.1} StPressure^2 SoilFrac^{1.9} + 0$	0.54	–
6	$Rate(6) = 1.7x10^{-13} \frac{Diam^{1.6} Age^{1.1} StPressure^2 SoilFrac^2}{SoilCorr^{1.2}} + 0$	0.54	–
7	$Rate(7) = 6.52x10^{-14} \frac{Diam^{1.2} Age^{1.1} StPressure^2 SoilFrac^2}{99.89 InterPrecRange^{0.2} SoilCorr^{1.3}} + 0$	0.55	–

As can be seen in Table 7.12 static pressure is reported from level 2 and appears in all equations. This is followed by age (from level 3 onwards) and diameter (from level 4 onwards). The 99.98th interpercentile range appears in the last level in denominator. Soil fracturing is reported in three last levels: 5, 6 and 7. Inverse relationship with soil corrosivity is shown in two last levels. No improvement in fit with respect to the base case at the same levels is seen.

7.12.2 Zone-level data

Seven equations were returned by EPR, all are presented in Table 7.13. Level 1 and 2 equations were previously seen in Table 7.5. Levels 3 and 4 in Table 7.11 respectively. From Table 7.13 it can be seen that soil variables appear in level 3 and 4. These variables provide an improvement in model fit (relative to a base case) by 1.09 and 1.42 respectively. The 99.98th interpercentile range appears in denominator from level 5 onwards. Only soil variables showed the improvement in model fit with respect to the base case at the same level.

Table 7.13: EPR output from zone-level data for soil and the 99.98th interpercentile range. New equations, compared to a base case, highlighted in blue; in black equations reported in the base case EPR.

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 8.87x10^{-7}$	0.00013	1
2	$Rate(2) = 2.54x10^{-7} \frac{1}{Diam^{0.6}} + 0$	0.0019	1
3	$Rate(3) = 1.05x10^{-7} \frac{SoilFracturing^{0.7}}{Diam^{0.7}} + 0$	0.0025	1.09
4	$Rate(4) = 1.14x10^{-7} \frac{SoilFracturing^{2}}{Diam^{0.7} SoilCorrosivity^{1.3}} + 0$	0.0034	1.42
5	$Rate(5) = 8.66x10^{-8} \frac{SoilFracturing^{2}}{Diam^{0.7} 99.98InterPercRange^{0.3} SoilCorrosivity^{1.5}} + 0$	0.004	–
6	$Rate(6) = 5.24x10^{-10} \frac{Age^{0.5} SoilFracturing^{2}}{Diam^{0.7} 99.98InterPercRange^{0.3} SoilCorrosivity^{1.6}} + 0$	0.0045	–
7	$Rate(7) = 2.72x10^{-10} \frac{Age^{0.4} StPressure^{0.4} SoilFracturing^{2}}{Diam^{0.7} 99.98InterPercRange^{0.4} SoilCorrosivity^{1.6}} + 0$	0.005	–

7.13 EPR transient fingerprint and soil

This section compares soil variables and the transient fingerprint in one case type (BC+S+T). This is examined to show whether and how this transient fingerprint categorisation may interact with soil.

7.13.1 Pipe-measured data

In total 18 equations were reported by EPR. From level 7 the complexity of the reported equations increased but did not provide improvement in model fit. Therefore, first eight levels are reported in Table 7.14 (one more than the last significant level). Levels 1 and 2 were previously reported in a base case (BC), see Table 7.4. Levels 3, 5 and 6 were reported in BC+T, see Table 7.8.

Table 7.14 shows that static pressure is reported from level 2 and appears in all equations. Region 3 appears in the denominator in level 3 and provides the improvement fit by 1.13 (relative to a base case at the same level). Region 1 is reported as a positive variable in level 4 and provides an improvement in model fit by 1.08 with respect to the base case at the same level. From level 5 a relationship $\frac{Region1}{Region2Region3}$ is reported in every equation. This, however, provides no improvement in model fit. Soil fracturing is seen in two last levels: 7 and 8 with no improvement in model fit. Only pressure transient variables showed the improvement in model fit with respect to the base case at the same level.

Table 7.14: EPR output from the pipe-measured data for soil and transient classification 90%. 18 equations were reported in total. New equations, compared to base case, highlighted in blue, in black equations already reported in base case EPR.

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 3.14x10^{-6}$	0.016	1
2	$Rate(2) = 1.72x10^{-9} StPressure^2 + 0$	0.28	1
3	$Rate(3) = 2.74x10^{-7} \frac{StPressure^2}{Region3^{0.7}} + 1.16x10^{-7}$	0.52	1.13
4	$Rate(4) = 6.73x10^{-6} \frac{StPressure^2 Region1^{0.8}}{Region3^2} + 1.37x10^{-6}$	0.56	1.08
5	$Rate(5) = 8.96x10^{-5} \frac{StPressure^2 Region1^2}{Region2^2 Region3^{1.7}} + 1.85x10^{-6}$	0.59	–
6	$Rate(6) = 5.94x10^{-9} \frac{Age^{1.1} StPressure^2 Region1^{1.9}}{Region2^2 Region3^2} + 1.97^{-6}$	0.59	–
7	$Rate(7) = 8.52x10^{-9} Region3^{0.4} SoilFracturing^2 + 9.11x10^{-5} \frac{StPressure^2 Region1^2}{Region2^2 Region3^{1.7}} + 0$	0.6	–
8	$Rate(8) = 2.18x10^{-8} Region3^{0.3} SoilFracturing^2 + 6.01x10^{-9} \frac{Age^{1.1} StPressure^2 Region1^{1.9}}{Region2^2 Region3^2} + 0$	0.6	–

7.13.2 Zone-level data

Nine equations were returned by EPR, that are presented in Table 7.15. Level 1 and 2 were previously reported in Table 7.5 and levels 3, 4 and 5 in Table 7.11 respectively. Soil variables appear in level 3 and level 4 and provide an improvement in model fit (relative to a base case) by 1.09 and 1.42 respectively. Pressure transient data appears from level 6 onwards in the following forms $\frac{1}{Region4}$, $\frac{1}{Region1}$, $\frac{1}{Region1 Region4}$ and $\frac{Region3}{Region1 Region4}$, however no improvement in model fit is seen.

Table 7.15: EPR output from the zone-level pipe data for soil and transient classification 90%. New equations, compared to a base case, highlighted in blue, in black equations reported in the base case EPR.

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 8.87x10^{-7}$	0.00013	1
2	$Rate(2) = 2.54x10^{-7} \frac{1}{Diam^{0.6}} + 0$	0.0019	1
3	$Rate(3) = 1.05x10^{-7} \frac{SoilFracturing^{0.7}}{Diam^{0.7}} + 0$	0.0025	1.09
4	$Rate(4) = 1.14x10^{-7} \frac{SoilFracturing^{2}}{Diam^{0.7} SoilCorrosivity^{1.3}} + 0$	0.0034	1.42
5	$Rate(5) = 6.9x10^{-10} \frac{Age^{0.5} SoilFracturing^{2}}{Diam^{0.7} SoilCorrosivity^{1.4}} + 0$	0.0039	–
6	$Rate(6) = 2.62x10^{-8} \frac{Age^{0.5} SoilFracturing^{2}}{Diam^{0.7} Region4^{0.3} SoilCorrosivity^{1.4}} + 0$	0.0042	–
7	$Rate(7) = 6.5x10^{-10} \frac{Age^{0.4} StPressure^{0.5} SoilFracturing^{2}}{Diam^{0.7} Region1^{0.1} SoilCorrosivity^{1.5}} + 0$	0.0046	–
8	$Rate(8) = 2.47x10^{-8} \frac{Age^{0.4} StPressure^{0.5} SoilFracturing^{2}}{Diam^{0.7} Region1^{0.1} Region4^{0.3} SoilCorrosivity^{1.5}} + 0$	0.0048	–
9	$Rate(9) = 1.71x10^{-8} \frac{Age^{0.5} StPressure^{0.4} Region3^{0.1} SoilFracturing^{2}}{Diam^{0.7} Region1^{0.1} Region4^{0.4} SoilCorrosivity^{1.4}} + 0$	0.0049	–

7.14 EPR transient fingerprint for asbestos cement pipes

The additional EPR analyses are undertaken on small size, pipe-measured, asbestos cement dataset to further evaluate the impact of transient fingerprints on repair rates. This data was the second largest dataset (see Table 6.11) available after cast iron. This data had 25 pipes with 5 pipe failure records.

Transient fingerprint showed to be an important variable for the cast iron pipe-measured dataset. This was seen by comparison BC and BC+T. To evaluate whether this can also be observed in asbestos cement pipes BC and BC+T comparison is undertaken. Figure 7.16 and Figure 7.17 show EPR results for a base case and base case with pressure transient fingerprint for the asbestos cement pipe-measured data.

Table 7.16: EPR output for the pipe-measured data from a base case model (asbestos cement pipes).

Level	Equation reported by EPR	CoD value
1	$Rate(1) = 7.1x10^{-7}$	0.048
2	$Rate(2) = 1.48x10^{-9} StPressure^{1.6} + 0$	0.11
3	$Rate(3) = 6.54x10^{-11} \frac{StPressure^2}{Diam^{0.7}} + 0$	0.12
4	$Rate(4) = 3.31x10^{-12} \frac{Age^{0.3} StPressure^2}{Diam^{0.7}} + 0$	0.12

Table 7.16 shows the order of variable importance as static pressure, diameter and age. Similar ordering of variables was seen in cast iron base case with the exception of diameter in the denominator for asbestos cement data.

Table 7.17: EPR output for the pipe-measured data from a base case model with pressure transient classification (90%). New equations, compared to a base case, highlighted in blue, in black equations reported in the base case EPR (asbestos cement pipes).

Level	Equation reported by EPR	CoD value	Improvement in fit with respect to the BC at the same level
1	$Rate(1) = 7.1x10^{-7}$	0.048	–
2	$Rate(2) = 1.4x10^{-11} Region1^{1.3} + 0$	0.35	3.18
3	$Rate(3) = 9.9x10^{-14} StPressure^2 Region1 + 0$	0.52	4.33
4	$Rate(4) = 2.67x10^{-24} StPressure^2 Region1 Region4 + 0$	0.7	5.8
5	$Rate(5) = 3.41x10^{-31} StPressure^2 Region1^2 Region2^{0.7} Region4^2 + 0$	0.78	–
6	$Rate(6) = 2.08x10^{-31} Diam^2 StPressure^2 Region1^2 Region2^{1.2} Region4^2 + 0$	0.8	–
7	$Rate(7) = 1.65x10^{-31} \frac{Diam^2 StPressure^2 Region1^2 Region2^2 Region4^2}{Region3^{0.8}} + 0$	0.8	–

Table 7.17 shows that region 1 replaced static pressure in level 2 (base case, Figure 7.16). The observed order of importance of variables is: region 1, static pressure and region 4. All these variables are in a positive relationship with pipe repair rates. Region 1 provides an improvement in fit by 3.18 with respect to the base case at the same level. Static pressure improves the fit by 4.33 and region 4 by 5.8 respectively.

7.15 Summary

This chapter presented the results from the fieldwork transient pressure data collection. This data collection was undertaken to investigate the contribution of pressure transients to the observed repair rates. In the presented SOMs analysis and EPR modelling the strength of the relationships among variables was examined by type of cases. The pipe-measured equations, i.e. subset of pipes, which were directly measured, showed a better fit than the equations reported for the zone-level data.

For the pipe-measured data variables traditionally associated to cause pipe repairs, represented by the base case with diameter, age and static pressure, achieved model fit (CoD) equal to 0.52 (level 4). Static pressure was the most important variable. The 99.98th interpercentile range did not provide any improvement in fit, but was seen in positive relationship with pipe failure rates. The transient fingerprint categorisation showed the improvement in the model fit by 1.13 and by 1.08 in comparison with a base case at the same level. The addition of soil data did not show any improvement in model fit.

For the zone-level data EPR reported a weak fit. Significant variables were found in the equations returned, but static pressure was seen as the least important. For the zone-level data the rate of change of pressure provided some improvement in fit, but an inverse relationship to the pipe failure rates. The transient fingerprint show some marginal impact on pipe repair rates. The addition of soil data showed to provide some contribution to the observed repair rates.

Chapter 8

Discussion

8.1 Introduction

The aim of this chapter is to discuss the results and to explore assumptions made in this research, before drawing conclusions. The key objective of this research was to investigate and assess the contribution of pressure transients to pipe repairs, and determine whether pressure transients can help in quantifying and understanding historical pipe repair rates in the context with other contributing variables. To answer this research question, a large and systematic data collection was designed and implemented across a range of water networks. The findings from the fieldwork are discussed first.

A series of assumptions for model variables are then presented and discussed followed by discussion of the assumptions made for the EPR model, this includes discussion on the performance of EPR on the data containing only failed pipes.

Results from the modelling of different types of cases are then assessed. The base case results with standard variables (without pressure transients) for the pipe-measured and the zone-level cast iron pipes are examined. The discussion of the results showing the impact of soil variables (base case + soil) on pipe repair rates is presented next. Case types with pressure transients (base case + transient) included are then discussed and compared with the base case. The impact of soil is also compared to the impact of pressure transients in one case type (base case + soil + transient) to show whether pressure transients and soil interact. Additional discussion follows on asbestos cement results for the pipe-measured data.

8.2 Field data

The transient fingerprint method (as described in section 5.5) allowed the quantification of pressure transients from the continuously recorded time-series data. None of the previous studies could assess how often pressure transients occur in a real network, as they lacked both pressure transient data and a suitable pressure transient identification and assessment method. This research addresses this gap by introducing a transient fingerprint method, which allows characterisation and quantification of discrete pressure transient events by their magnitudes, durations and number of occurrences.

8.2.1 Occurrence of pressure transients

The extensive fieldwork data collection described in Chapter 4 provided an overview of the existence of pressure transients in water distribution networks based on the study of 67 DMAs divided into 79 zones (see section 6.3). The unbiased data collection process was designed to provide data from sites with high and low pipe repair frequencies and from sites traditionally associated and not associated with pressure transients. The examples of zone specific transient fingerprints captured the complexity, variety and widespread nature of pressure transients within studied networks (section 7.2). These data revealed greater transient activity in WDS than had been initially estimated (section 7.2, Table 7.1) thus providing a greater understanding of 'real' pressure transient occurrences. The results have shown (see Table 7.1) that pressure transients of magnitudes between 5 - 10 m were dominant (recorded in 78 zones), followed by 10-20 m (recorded in 56 zones) and 20-50 m (26 zones). This result suggests that pressure transients exist widely throughout water networks.

Following advances in technology, recent studies were able to measure pressure transients across wider networks (in up to 19 sites (Besner et al., 2010)). However, only one study (Karim et al., 2003) confirmed pressure transients in all monitored sites. The uncertainty around the existence of pressure transients can now be addressed with the evidence provided by this research.

The results chapter (see section 7.2) presented the examples of pressure traces and corresponding transient fingerprints, showing that varieties of pressure transients can be observed in water networks. Prior research generally targeted specific pressure transients (Kirmeyer et al., 2001; Gullick et al., 2004; Friedman et al., 2005; Besner et al., 2010); therefore, it provided a hard evidence of only extreme (mainly low or negative) pressure events. These types of pressure transients were targeted because they were thought to have an impact on water quality failures. This research expanded on previous studies by showing that both low and high magnitude pressure transient of diverse magnitudes are observed in water networks, see Table 7.1. It also provided the hard evidence which shows that overall from 79 zones, five experienced pressure transients of magnitudes between

50-100 m, and in only one zone were magnitudes greater than 100 m recorded (see Table 7.1). Therefore, it can be concluded that extreme pressure transients exist, albeit rarely.

8.2.2 Damping of pressure transients

This research has documented pressure transients induced not only by pumps but also by local industry and these events were seen propagating throughout the system. Early research attempted to model pressure transient propagation in simple networks (Karney and McInnis, 1990; McInnis and Karney, 1995) and compared the results with engineered pump 'trips'; these showed the quick damping of pressure transients. Computational simulations were also undertaken by some later studies (Friedman et al., 2005; Besner et al., 2010), suggesting a much quicker damping of real pressure transients in comparison to the simulated results. Therefore, the general existence of pressure transients were expected to be relatively quickly damped, but this remained unverified by real data. Current advances in the modelling of pressure transients also consider the effect of viscoelastic (Ramos et al., 2004a) and mechanical damping effects (Stephens et al., 2011) on pressure transient propagation. These studies have not performed analysis on more complex real networks due to its complexity and uncertainties. The findings from this research showed the persistence of pressure transients throughout various complex networks. This finding indicates that damping mechanisms can possibly affect pressure transients to a lesser extent than is currently suggested. This research shows that under normal operating conditions pressure transients are persistent across wide networks.

8.2.3 Sources of pressure transients

Pressure transients regularities observed during weekday working hours confirmed their association with an industrial demand as a source, see pressure traces and corresponding transient fingerprints in Figure 7.1 (c), (d), (e), (f) and Figure 7.2 (a), (b). A regular pressure transient occurrences unrelated to weekdays and weekends caused by pumping operations were also detected, see Figure 7.2 (c), (d). This data confirms findings from previous studies highlighting pump trips or some significant industrial activities as pressure transient sources. The occurrence of pressure transients was only anecdotally known beyond special cases close to pumping stations. This research expanded on previous knowledge by providing unique results from sites traditionally not associated with pressure transients, as described in Table 4.2. The results from the fieldwork showed that pressure transients existed in sites which were initially classified as 'traditionally not associated with pressure transients'. Some zones showed high numbers of small pressure variations with a high intensity transients occurring once or twice a week (see Figure 7.1 (a) and

(b)) potentially associated with malfunction of equipment installed in the network. For instance, a malfunction of PRVs can create pressure transients with possible adverse effect on the network. The propagation of pressure transients through PRVs requires further studies as it could not be verified whether these devices are able to 'damp'/stop pressure transients. PRVs may be too slow to react on high-speed pressure transients.

This research identified that regular pressure transients can be associated with pumping stations or industrial demand as pressure transient sources. Random events were also seen and these can possibly be associated with some maintenance operations or malfunction of water machinery. In contrast with previous studies, this research did not target any specific transients, such as extreme events generated by pumps, but determined occurrence of pressure transient through wide spectrum monitoring under normal operating conditions. This monitoring confirmed the existence of random and regular pressure transients, but it also showed that regular pressure transients are prominent in complex networks.

8.2.4 Locations

Time series data usually show that pressure transients measured within the same zone are similar; however, the transients fingerprint characterisation method has shown differences between them (see section 7.3). It was found that pressure transient magnitudes depend on location within a zone likewise pressure transient durations. This was demonstrated with time series data from the same zone (see pressure traces and corresponding transient fingerprints in Figures 7.3 and 7.4). Counts of pressure transient magnitudes (see Table 7.3) further clarified that these pressure traces are not the same. From this, it can be concluded that pressure transients change with respect to network characteristics (physical and hydraulic) and network layout: each site or location has a unique dynamic characteristic and pressure transients should be considered as location sensitive parameters. Previous attempts to model pressure transients in simple networks had only limited success, as discussed in section 8.2.2. The findings from this research show variabilities of pressure transients measured at different locations. Therefore, modelling pressure transients on a large scale can be a challenge and, to verify the results, pressure transient data should be collected at fine spatial resolution.

8.2.5 Number of measuring points

To measure and to identify pressure transient occurrence one high-speed measurement within a zone may be sufficient (despite previous assumptions of quick pressure transients damping occurring in real networks, see section 8.2.2). The unique pressure transient data was obtained

through a high-speed (100 Hz) pressure measurements (see section 4.5). According to the review of previous studies it was unclear how pressure transients should be measured across different networks to capture their existence and reveal their possible contribution to pipe repair rates. However, the results from the zone-level data showed that high spatial resolution describing pressure transient propagation (see also section 8.2.4) is required to accurately assess pressure transient impact on any large scale networks. The results from the pipe-measured data were not transferable/ comparable to a larger scale (zone-level) where accurate pressure transient data was not available. The methodology of data collection showed that one measuring point was sufficient to capture pressure transients, but not enough to fully determine their contribution to the observed pipe repair rates for which more fine spatial data resolution is required. Clearly, a spatial resolution can be a limiting factor for variables showing high sensitivity to location, like pressure transients.

8.3 Data assumptions and limitations

A series of modelling assumptions were made during the analysis. These included static pressure values collected over two week period, that were used in the analysis, pipe materials, diameter the cut offs values on variables (diameter, static pressure, length), the enhanced peak-trough detection algorithm and digital pipe repairs data available. The impact of cut offs on the results is presented first, followed by the discussion on the algorithm, the available span of pipe repair records, static pressure values collected over two weeks recording periods and finally the limited variety of pipe diameters and materials. Further limitations of the data were also identified that included low CoD values for the zone-level results which is discussed in a separate section.

8.3.1 Cut off choices

The impact of cut offs applied to variables as presented in Table 6.10, Table 6.13, Table 6.16 were assessed. These cut offs included: 1 m cut off on pipe length, 50 mm cut off on diameter and 10 m on static pressure respectively. The analyses were repeated for the base case with all data (before the cut offs were applied) and had no impact on final results. The analyses were also repeated for each cut off individually (also on the base case) and these did not provide any additional information or change to the final results. Therefore, it can be concluded that applied cut offs had no effect on the analysis.

8.3.2 Counting of pressure transient events

The enhanced peak-trough detection algorithm, see section 5.5, allowed the differences between pressure transients from diverse zones (and locations within a zone) to be quantified. The peak-trough

detection algorithm may appear to be counting of pressure transient events 'twice', see section 5.5, Figure 5.8. Counting the events 'twice' has the potential to explore what is useful and what is not and its importance in association with impact explored. Therefore, the algorithm was designed to count 'all' pressure transient events and acknowledges them as equally valid parts of the signal. Subsequently, 'all' pressure transient events, comprising a transient fingerprint, were categorised into four regions (as a reminder see Figure 8.1, and for more details section 6.2.2).

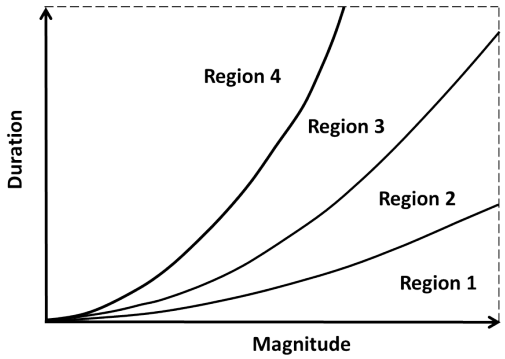


Figure 8.1: Schematic plot showing locations of transient fingerprint regions.

The events of large magnitude and short duration are located in region 1, medium size magnitude and duration in region 2, small magnitude and long duration in region 3. All these events were separated from region 4 which was designed to hold events typically observed in each location, therefore, resembling common underlying transient behaviour, see Figure 8.1. Events counted 'twice' are most likely to be accumulated in region 4 and less likely in regions 3 and 2 respectively. The EPR equations for the pipe-measured data returned the following combinations of regions : $\frac{Region1}{Region2}$ and $\frac{Region1}{Region2Region3}$. This may indicate that the effect of counting of pressure transient events utilised by the enhanced algorithm may have a limited impact on the network performance due to the fact that regions 2 and 3 are in the denominator.

Based on the results from the pipe-measured data it can also be concluded that the choice to divide transient fingerprint into four regions was a reasonable decision as each region showed some or none contribution to the observed pipe repair rates. Regions 1 (pipe-measured data) and region 2 (zone-level data) showed to have a strong positive relationship with pipe repair rates.

8.3.3 Pipe failure records

Digitalization of the pipe failure records enabled such data to be used for diverse analysis. The span of available data, however, is usually quite short. This fact has been also identified by other researchers (Savic and Banyard, 2011). Short spans may not be informative and may not reveal potential trends in pipe repairs. Therefore, records of a longer time scale are required. Pipe repair records available for this research were 11 years long (see section 6.4.1). The pipes with repair

records consisted of about 8.2% (zone-level) of the overall cast iron data. In this research the available span was sufficient to find relationships between variables and pipe repair records which gives a confidence in the results.

Due to the disparate time series of the data collection (11 years for pipe repairs and 11 months for the pressure transient data at discrete locations) and the fact that there were only a few examples (none confirmed) of pipe repairs during a typical recording period in each zone, it was felt that the pressure transient data should initially be collected as a structural (time invariant) property of each zone. Therefore, the collected pressure traces were assumed to be representative of the pressure experienced in each zone over 11 year period during which pipe repair records were collected. Note that some of the pipes were 160 years old. The dynamic pressure history for each pipe lifetime was not available. Available was a two weeks snapshot of 'now' and, despite this monitoring period, it has been demonstrated that pressure transient data was sufficient to draw conclusions from equations describing pipe repair records, which may be considered as a good result. A two weeks monitoring period appear to be an effective strategy to identify transient risks in water networks.

8.3.4 Static pressure

It has been decided that for the purpose of this research the static pressures recorded from the field, as part of the field monitoring programme, will be used in the analysis, see section 6.4.5. The results (see Table 7.4, 7.6, 7.8, 7.10, 7.12, 7.14) showed that static pressure was the most important variable for the pipe-measured data. It appeared in level 2 showing the importance of this variable in explaining pipe repair rates. The importance of static pressure for the pipe-measured data is with the agreement with previous studies (Brandon, 1984; Skipworth et al., 2002), however it should be noted that the static pressure values were only collected over two weeks recording during the fieldwork. This could have impacted the results mainly because historical static pressure values were not available. This result should also be interpreted in the context of pressure management zones which were historically introduced to reduce and control leakage across the UK in the early 1980s (section 2.1.1) and are relatively modern strategies to reduce pipe repairs (Ulanicki et al., 2008; Gomes et al., 2011). In this context, static pressure values used in this research are mainly a descriptor of recently introduced changes. As the pipes have been present for longer, it is unlikely that static pressure was the same throughout a pipe life span and static pressures may have caused damage to the pipe prior to these network changes. This indicates that more precise static pressure data, preferably considering historical values, is needed to improve pipe-measured analysis. Despite these limitations, the static pressure values collected over a typical two weeks

recording period were sufficient to reveal relationship with pipe repair rates for the pipe-measured data.

8.3.5 Pipe material

The analysis focused on cast iron pipes due to the highest repair records and data size, see Table 6.11. These analyses were carried out at the pipe-measured and the zone-level. In addition, asbestos cement pipe-measured data were also investigated to further explore the findings observed in the cast iron pipe-measured data. However, these data were limited to number of records available. The use of a low variety of pipe materials is a limitation; however this could not be avoided for this research. Other material types could therefore be investigated as a future work. Greater variability in pipe materials would provide broader overview of the impact of pressure transients on water networks.

8.3.6 Pipe diameter

The cast iron pipe-measured data had 78% of pipes with a diameter of 101.6 mm.This shows that the majority of field measurements were done on limited variety of diameters. This could not be avoided during the fieldwork as at that point the occurrence of pressure transients in water networks was unknown. This means that this variable was not effectively studied and that the results are applicable to a specific diameter range. The results for the pipe-measured data showed that diameter was in a positive relationship with pipe repair rates. The results obtained from the zone-level data had greater variations in diameter values and showed a reduced and negative exponent of diameter.

8.3.7 CoD value

The coefficient of determination for pipe repair rates is data specific (relative to data, see 2.4.5.1). Previous studies which used a CoD as a measure of fit, usually obtained high values for cast iron pipes, around 0.82 (Kettler and Goulter, 1985; Kakoudakis et al., 2016), and lower, i.e. 0.75, for asbestos cement pipes (Kettler and Goulter, 1985). In this research similar values were observed for cast iron pipe-measured data (around 0.6), but values less than 0.005 were obtained for the zone-level data. High values of the CoD are usually desirable but it could be that the CoD value is high but the observed relationship is not useful as it may not be appropriate, sensical or does not follow expectations from the literature. Therefore, the obtained form of the relationships should also be assessed, not only its CoD value. In majority of cases when the CoD is low the relationship does not exist. However, there are cases that show expected forms of relationships with a low CoD

values. There are also cases that prefer the equations with lower CoD values over the equation with higher CoD because the former provides better description of the underlying phenomenon (Savic et al., 2009). This is observed in the zone-level results that firstly formed a relationship and secondly the obtained relationship follows trends previously reported in the literature. Therefore, the results from the zone-level data were considered valid. To enhance the understanding of the impact of pressure transients on pipe repair rates, this research focused on *changes* in CoD values for a sensical relationships between different type of cases.

8.4 EPR assumptions

Series of EPR model settings were selected before obtaining final model equations. These included an intercept, choice of the exponent values and the assumption that EPR can perform on data with failed and non-failed pipes. These assumptions are discussed in the following sections.

8.4.1 Intercept

In this research the intercept was included in all equations returned by EPR. The intercept equal to zero indicates that all data were explained by the input variables in the equation. In contrary, the intercept different than zero suggests that some repairs are not described by the variables. Base case with transient fingerprint for the pipe-measured data had the intercept not equal to zero and is discussed.

The equations for the pipe-measured data from level 3 onwards had the intercept not equal to zero, see Table 7.8. In level 3 the intercept was equal to $1.16x10^{-7}$ $(NoF/m/day)$. However, the average failure rate value was $3.14x10^{-6}$ $(NoF/m/day)$ (level 1, CoD was equal to 0.016, i.e 1.6%). In level 3 the intercept is 27 times less than overall average. The intercept, therefore, shows that some part of the data was explained by the constant (intercept) and the rest by the variables. The important variables for this equation were $\frac{StaticPressure}{Region3}$, level 5. Level 4, had the intercept equal to $1.49x10^{-6}$ $(NoF/m/day)$, 2.1 times less than overall average, which also shows that part of the data is explained by this constant. The rest of the data was explained by the variables from which the important ones were $\frac{StaticPressure \cdot Region1}{Region2}$.

For the rest of the EPR returned equations the intercept was equal to zero and therefore, did not have an impact on the results. Only for two cases, the pipe-measured data, the intercepts were not equal to zero (discussed). Therefore, it gives confidence in the EPR equations and the variables chosen.

8.4.2 Exponents choice

For this research the possible exponents values were chosen to be between -2 to 2 with the separation between successive values of 0.1, i.e. [-2,-1.9,-1.8 (...), 0, (...), 1.8, 1.9, 2]. Previous research considered the most well-known relationships, e.g. linear, quadratic, inverse linear, square root, therefore possible power values of [-2,-1.5,-1,-0.5, 0, 0.5, 1, 1.5, 2] were used and the separation between values were 0.5 (Savic et al., 2009; Laucelli et al., 2014; Mounce et al., 2016). In this research the separation between successive values was increased to allow EPR a greater 'freedom' to produce a better fit. The CoD with these exponent choices were then assessed against CoD suggested by the literature for all cases for the pipe-measured and the zone-level data. The results showed to only slightly reduce CoD values and had no impact on equation forms. Therefore, the choice of increased separation between successive exponent values was retained.

The choice of exponents to the power of two did not have an impact on only one type of case, i.e. base case with rate of change of pressure, for the zone-level data. For this case all exponents had values lower than 2. The exponents to the power of 2 were observed for the rest of cases. The possible increase of the exponent limit beyond the power of two was investigated in this research, however, there was no obvious limit of the exponents to restrain the data. Therefore, the decision to bound the exponents within limits from -2 to 2 was made based on engineering judgement and previous studies (section 2.4.5).

8.4.3 Performance on non-failed pipe data

This section investigates the assumption whether EPR can perform on data with excessive numbers of non-failed pipes. This was especially important for the zone-level data which had 91.8% pipes with zero repair rates. The assumption was checked for the base case of reduced data (non-failed pipes removed), which was then compared to the base case at full size (with non-failed pipes). The obtained equations were than compared. This was performed on the pipe-measured and the zone-level data respectively.

If this assumption was satisfied EPR could be used on the remaining type of cases. Other modelling methods were investigated, see section 6.6 but were found not to be suitable for data with excessive zero repair rates, e.g. standard GLMs described in section 6.6.2.2

8.4.3.1 Pipe-measured results

The pipe-measured data contained 22 failed pipes. The EPR results for only failed pipes are presented in Table 8.1.

Table 8.1: EPR output from failed pipe-measured level data for a base case model input.

Level	Equation reported by EPR	CoD value
1	$Rate(1) = 9.15x10^{-6}$	0.045
2	$Rate(2) = 3.69x10^{-9}Pressure^2 + 0$	0.63
3	$Rate(3) = 4.23x10^{-10}Age^{0.2}Pressure^2 + 0$	0.63
4	$Rate(4) = 6.13x10^{-3}\frac{1}{Pressure^2} + 9.811x10^{-9}Diam^{0.5}Pressure^2 + 0$	0.63
5	$Rate(5) = 7.4x10^{-3}\frac{1}{Pressure^2} + 3.3x10^{-9}Diam^{0.5}Age^{0.1}Pressure^2 + 0$	0.63

In total 10 equations were produced by EPR. The first five are included in the table, one more than the last significant level. The results for the full size (with non-failed pipes) for the pipe-measured data are presented in Table 7.4. It can be seen that higher CoD values were obtained for failed pipes data. However, from level 2 the fit did not improve beyond CoD value of 0.63. The full size data (with non-failed pipes) showed an improvement in the CoD value at each level, however, the values were lower (in comparison to failed pipes) with a maximum CoD value of 0.52 obtained in the last level. Furthermore, the same forms of the equations were seen for both data, i.e., static pressure appeared in level 2, age in level 3 and diameter in level 4 respectively and a positive relationships with pipe repair rates were observed for all variables. However, the EPR on failed pipe data returned more complex equations from level 4, where an inverse relation with static pressure can be seen. Therefore, it can be concluded that the results on full size data are not only comparable to failed pipe data but are also more conclusive. EPR has shown to perform satisfactory on the full size pipe-measured data where non-failed pipes were included.

8.4.3.2 Zone-level results

The comparison analysis was carried out on the zone-level data with non-failed pipes removed for the base case. The failed pipes data included 655 pipes. EPR modelling results for this data are presented in Table 8.2.

The equations for the full size data (with non-failed pipes) are presented in Table 7.5. It can be observed that higher CoD values were obtained for failed pipes data, i.e. in level 4 CoD was equal to 0.0079 in comparison to 0.0024 for the full size data at the same level. However, simpler equations were obtained for the full size data. The equations for failed pipe data were more complex from level 5 onwards. In addition, level 5 and 6 showed an inverse relationship with age.

Table 8.2: EPR output from failed zone-level data for a base case model input.

Level	Equation reported by EPR	CoD value
1	$Rate(1) = 1.08x10^{-5}$	0.0015
2	$Rate(2) = 2.33x10^{-7}Age^{0.4} + 2.92x10^{-6}$	0.0043
3	$Rate(3) = 2.33x10^{-8}\frac{1}{Diam^2} + 1.34x10^{-7}Age^{0.4} + 0$	0.0074
4	$Rate(4) = 2.36x10^{-8}\frac{1}{Diam^2} + 9.06x10^{-8}Age^{0.4}Pressure^{0.1} + 0$	0.0079
5	$Rate(5) = 1.76x10^{-7}\frac{1}{Diam^2Age^{0.2}} + 9.22x10^{-8}Age^{0.4}Pressure^{0.1} + 0$	0.008
6	$Rate(6) = 1.76x10^{-7}\frac{1}{Diam^2Age^{0.2}} + 1.91x10^{-9}Pressure^{1.1} + 9.07x10^{-8}.Age^{0.4}Pressure^{0.1} + 0$	0.008

Overall, the investigation of the assumption whether EPR is able to return model equations on data with excessive number of non-failed pipes showed that this assumption is valid. This is also conclusive with previous studies (Kakoudakis et al., 2016), which used data with failed and non-failed pipes for the EPR analysis. The results with and without failed pipes for the pipe-measured and the zone-level are comparable. Furthermore, for the zone-level it was observed that full data returned simpler and more concise equations despite lower CoD values. However, the overall CoD value (as discussed in section 8.3.7) was not considered important. The pipe-measured data results showed to be slightly better, however this may be affected by data size for removed non-failed pipes in the pipe-measured data (22 pipes). However, previous studies (Kakoudakis et al., 2016) reported that for the EPR analysis number of samples should be less or equal to number of parameters to be estimated.

The analysis on failed pipe data supports the decision that full data was a good choice for the pipe-measured and the zone-level data respectively. EPR showed to be performing well on data with failed and non-failed pipes.

8.5 Base case modelling results

A research aim was to present a method to enhance the understanding of the contribution of pressure transients to the observed pipe repair rates. Cast iron pipes were investigated due to the highest repair records and data size, see Table 6.11, but at the pipe-measured and the zone-level. In addition, asbestos cement pipe-measured data was also investigated to further explore the findings observed in the cast iron pipe-measured data. This data was chosen because it was the second largest dataset (pipe-measured) available. Common variables describing pipe failure rates formed a base case (BC) input to EPR. The base case comprised three core, traditional variables related to

pipe repairs, i.e. static pressure, age and diameter (section 6.5). The introduction of sophisticated data analysis techniques, i.e. EPR (section 6.6.6) facilitated quantifying a relationship between these variables and pipe repair rates.

The pipe-measured data consisted of 64 pipes of which 22 had a recorded repair (s), i.e. 34%. In contrast, the zone-level had 7978 pipe records, from which 655 were pipes which failed, i.e. 8.2%. The average pipe repair rate for the pipe-measured data was $3.14x10^{-6}$ $(NoF/m/day)$, equivalent to 1.15 $(NoF/km/year)$ as described by a constant in level 1 Table 7.4. The zone-level data has a slightly lower average repair rate of $8.87x10^{-7}$ $(NoF/m/day)$, equivalent to 0.32 $(NoF/km/year)$ as described by a constant in level 1 Table 7.5. Comparison of repair rates shows that these two data are not representative of each other and this was also seen through the analysis.

The results from this work identified static pressure, age and diameter as important variables. Based on the EPR ranking of returned model equations by both the coefficient of determination and structural complexity (Giustolisi and Savic, 2009) the variable ranking was determined for the base case input as shown in Table 8.3. The table presents variable importance ranking in descending order for the pipe-measured and the zone-level data, based on Tables 7.4 and 7.5. As can be seen the pipe-measured and the zone-level data demonstrated different variable ranking. In addition, the way each variable interacts with each other and its exponent value, were also considered.

Table 8.3: Comparison of variable ranking for the base case pipe-measured and zone-level data.

Variable ranking	Pipe-measured	Zone-level
1	Static hydraulic pressure^2	$\frac{1}{Diameter^{0.6}}$
2	Age^2	$Age^{0.4}$
3	$Diameter^{1.8}$	Static hydraulic pressure$^{0.2}$

Historical studies have identified a number of key variables associated with pipe failures. Table 8.4 lists key variables: age (section 6.4.3), static pressure (section 6.4.5) diameter (section 6.4.4) and how they may be expected to correlate with pipe failures based on published work.

Table 8.4: Available variables and expected relationship with pipe failure rates. Ordering based on published work, section 2.3.2.

Variable	Expected relationship
Age	positive
Static pressure	positive
Diameter	inverse

8.5.1 Static pressure

This research has shown a positive relationship between static pressure and pipe repair rates. This was observed in both (pipe-measured and zone-level) data results. Static pressure is expected to be a positive variable, as higher pressure would cause more pipe repairs and was shown in previous studies (Skipworth et al., 2002; Lambert, 2002; Ulanicki et al., 2008; Gomes et al., 2011; Ghorbanian et al., 2016). Therefore, the results agree with previously published work.

Static pressure was the most important variable for the pipe-measured data (see Table 8.3 and for more details Table 7.4). This is with an agreement with previous studies, which highlighted the importance of static pressure in relation to pipe repair rates (Ghorbanian et al., 2016; Skipworth et al., 2002; Brandon, 1984). For the pipe-measured data the values of static pressure were directly measured during the fieldwork (accurate/real values). Static pressure appeared in level 2 and improved the model fit by 17.5 times with respect to the previous level showing the importance of this variable in explaining pipe repair rates.

For the zone-level data the static pressure was raised to the power of *0.2*. This suggests that variations of static pressures are less important than for the pipe-measured data to describe pipe repair rates. For the zone-level data, static pressure appeared in the last level. This may be related to the fact that for this data static pressure values were not directly measured but extrapolated for each pipe (see section 6.4.5) based on network topography; adjusted by elevation and the division of the network into zones according to the presence of PRVs. Previous studies showed that static pressure is highly related to network topography and elevation (Lambert, 1998). However, for this research the extrapolated static pressure values (considering elevation) showed to have the least impact on pipe repairs for the zone-level data. The extrapolation of static pressures may, therefore, explain why it is a poor descriptor of pipe repair rates. Furthermore, some discrepancies between measured and calculated values of static pressures were observed after the extrapolation that can further explain this results. These can be attributed to a number of uncertain parameters in operation in real networks. Such issues were also identified by previous research (Ghorbanian et al., 2016). Numbers of uncertain factors (head loss, burial depth) could therefore impacted the extrapolation of static pressure values. For the zone-level data the redefining of zones improved the static pressure values by providing more accurate static pressure values, however, this was not sufficient to improve the overall zone-level results. Static pressure values for the zone-level were found to be not entirely informative due to extrapolation and numbers of uncertain factors. This indicates that more precise static pressure data is needed for the zone-level data to accurately assess their contribution on pipe repair rates.

8.5.2 Age

It should be expected that high pipe age would correlate with high pipe repair rates. The base case analysis supports this assumption for the pipe-measured and the zone-level data (see Table 8.3, and for more details Tables 7.4 and 7.5). In both groups, age was observed as a positive variable suggesting that older pipes experience more repairs.

An exponential relation with age was reported by Shamir and Howard (1979) and by Walski and Pelliccia (1982), and time-powered by Kleiner and Rajani (2001). Kettler and Goulter (1985) found that for cast iron pipes overall repair rates increase with time. A strong positive correlation with age, understood as the time a pipe is exposed to various external and internal factors, has been well documented (see section 2.3.2) and a number of environmental conditions, as well as different external and internal load histories, were previously associated with age (Skipworth et al., 2002); for these reasons age is considered as a complex factor. Boxall et al. (2007) concluded that annual burst rates in cast iron pipes showed a complex, i.e. unclear association with age, with age being seen as a surrogate for other factors (e.g. quality and strength of the material). This may explain the ordering of variables for the pipe-measured data showing age as the second important variable after static pressure.

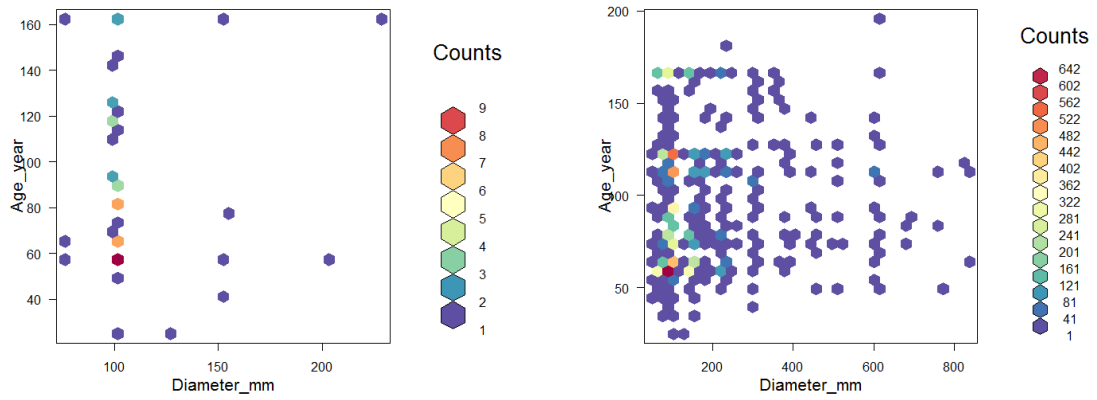
The value of the exponent for age in the pipe-measured data was high, i.e. *2* and *1.6* while for the zone-level data it was *0.4* in all model equations returned by EPR. A low value of the exponent suggests that age will have less effect on pipe repair rates over longer periods. Age of pipes follows the 'bathtub curve' previously reported in the literature review (see section 2.3.2). An exponent of *0.4* could therefore be associated with describing the flat bottom of the 'bathtub curve'. However, the average pipe age in the zone-level data is 100 years, whereas in the pipe-measured, 90 years. Higher exponent values for the pipe-measured data suggest that a different part of the 'bathtub curve' may be described in this data. Higher pipe repair rates for the pipe-measured data may support this assumption.

8.5.3 Diameter

The pipe-measured data results (see Table 8.3 and for more details Table 7.4) show a positive relationship with diameter. This is contrary to expectations from the literature, as shown in Table 8.4. Smaller diameter pipes have traditionally been associated with higher repair rates due to their inability to resist bending when the same force is applied (Skipworth et al., 2002). The inverse relationship between pipe repairs and diameter were widely reported (Shamir and Howard, 1979; Kettler and Goulter, 1985; Andreou, 1986; Walski et al., 1986). EPR model equations for the pipe-measured data returned diameter in the last level, providing only a slight improvement in

the measure of fit, small decrease in the exponent of age with no effect on the exponent of static pressure.

The average diameter for the pipe-measured and the zone-level data were 110 mm and 155 mm respectively. However, for the pipe-measured data 78% of the pipes had a diameter of 101.6 mm. Therefore a lack of variation in diameter limits its value as a variable. This seems to be confirmed by the results obtained from the zone-level data, which had greater variations in diameter values and showed a reduced and negative exponent of diameter. The relationship between age and diameter is presented in Figure 8.2 for the pipe-measured and the zone-level data respectively. For the pipe-measured data, there appears to be some relationship between adding diameter and reducing the exponent of age, suggesting that these variables may not be independent. However, for the zone-level data the addition of age does not affect diameter, suggesting that these variables are more likely to be independent.



(a) **(b)**

Figure 8.2: 2D histogram of age and diameter (a) pipe-measured data (b) zone-level data.

The zone-level data analysis showed that the most important variable is diameter (Table 7.5). Its exponent value was not altered by subsequent addition of age and static pressure. In this data diameter appears in the denominator and therefore fits expectations from published work (Shamir and Howard, 1979; Walski and Pelliccia, 1982; Kettler and Goulter, 1985; Boxall et al., 2007).

8.5.4 Other factors

Number of factors were identified in the literature to explain pipe repair rates (see section 2.3.2). The most important were considered in this research including various soil variables.

8.5.4.1 Temperature differential

Temperature differential was historically identified as one of possible factors affecting pipe repairs. Many studies have shown an increase in pipe repairs during winter season or during periods of frost and low temperatures (Habibian, 1994; Rajani and Tesfamariam, 2005; Tesfamariam et al., 2006). This can be related to the soil movement due to volume expansion of water due to low temperatures. Temperature is therefore an important factor but not predominant. In this research variations in temperature were indirectly captured by soil movement (soil fracturing variable).

Some of the other causes of pipe repairs previously indented by researchers (Morris, 1967) are:

- Inadequate design
- Improper installation
- Manufacturing defects
- Traffic impact

Inadequate design, improper installation and manufacturing defects are often considered with age. This research considered age as one of core variables (base case) which indirectly account for these causes. Traffic impact data was not available for this research. Other studies (Brevis et al., 2014) highlighted that interactions of cyclic loading of traffic with internal circumferential and longitudinal loads can have adverse effect on pipes. In this research the external pressure resulting from traffic was indirectly accounted for by including soil variables in a base case + soil (BC+S) type of input to EPR, see section 6.5.

8.5.4.2 Impact of soil

SOMs analysis identified soil fracturing and soil corrosivity as important, unique and independent parameters describing pipe repair rates. This was in agreement with previous studies (see section 2.3.2).

The addition of soil corrosivity and soil fracturing to the base case data formed a new case type, i.e. base case + soil (BC+S) (see section 6.5) which was then evaluated through EPR. Soil variables did not provide much improvement for the pipe-measured data as both parameters appeared in the last two levels, see Table 7.10 and Table 8.5 showing the variable importance ranking. None of the additional soil variables described the pipe repair rates better than the base case variables.

For the zone-level data, soil fracturing was identified as an important variable after diameter. Soil corrosivity appeared as the next important variable before age and static pressure, as summarised

in Table 8.5. The important fact is that all core variables have been retained and their relation to pipe repair rates was as previously observed in the base case.

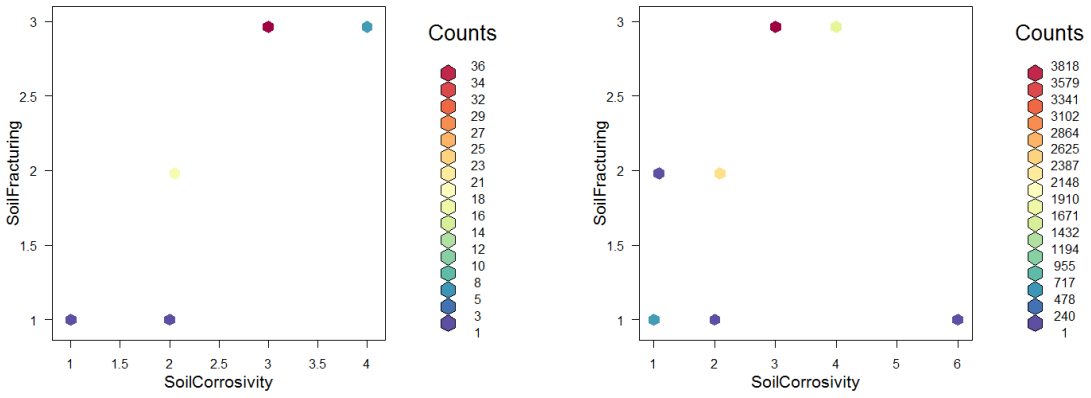
Table 8.5: Comparison of variable ranking for the base case+soil (BC+S) pipe-measured and zone-level data.

Variable ranking	Pipe-measured base case + soil	Zone-level base case + soil
1	Static pressure^2	$\frac{1}{Diameter^{0.6}}$
2	Age^2	Soil fracturing$^{0.7}$
3	Diameter$^{1.8}$	$\frac{1}{Soilcorrosivity^{1.3}}$
4	Soil fracturing$^{1.9}$	Age$^{0.5}$
5	$\frac{1}{SoilCorrosivity^{1.2}}$	Static hydraulic pressure$^{0.3}$

Soil fracturing appeared as a positive variable for the pipe-measured and the zone-level data. It was an important variable for the zone-level data where a positive relationship (as age) was formed in level 3 and level 4. This relationship is expected as older ferrous pipe would be more affected by soil fracturing due to them being subject to corrosion weakening a pipe walls over time. Previous research identified soil fracturing (e.g. ground movement) as an important factor affecting pipes (Skipworth et al., 2002). As it has been already mentioned (section 8.5.3) small diameter pipes have low ability to accommodate bending. Previous research (Skipworth et al., 2002) concluded that ground movement affects cast iron pipes more than other pipe materials, including plastic and more rigid pipes (steel and ductile iron). Ground movement was thought to cause fatigue effect on pipes (Skipworth et al., 2002) and clay soils were considered as especially prone to movement due to shrinking and swelling related to temperature changes and moisture content. A positive relation found in this research between soil fracturing and pipe failure rates gives a confidence in analysis results.

Soil corrosivity: The inverse relationship with soil corrosivity was observed in level 6 for the pipe-measured data and in level 4 for the zone-level data. This is contrary to literature expectations (Skipworth et al., 2002) suggesting that external corrosion is a main factor of pipe failures due to structural deterioration. Internal corrosion caused by chemical interaction with supplied water was considered less structurally damaging. As such, the internal corrosion is not considered as a pipe failure driver in the literature. Usually the opinion is that more corrosive soils would cause more pipe failures. However, some researchers concluded that soil corrosivity cannot be concluded from chemical properties. Higher amount of clay in soil represents greater moisture retention capacity which leads to higher soil corrosivity (Sadiq et al., 2004). External galvanic corrosion occurs in wet environment where current between soil and a pipe is created (O'Day, 1989). The inverse relationship with soil corrosivity found by this research can be explained by pipe laying techniques

requiring gravel or other external protections for cast iron pipes. Pipe laying techniques could protect a pipe against the adverse effect of corrosive soil, but would not prevent it from a ground movement (associated with soil fracturing). The pipe-measured data reports that soil does not have an effect on pipe failure rates. However, for the zone-level data soil variables were important parameters (the same importance as age). The relation between both variables (soil fracturing and soil corrosivity) was therefore further investigated as shown in Figure 8.3.



(a) **(b)**

Figure 8.3: 2D histogram of soil fracturing and soil corrosivity (a) the pipe-measured data (b) the zone-level data.

The figure shows that there may be a positive correlation between two variables. The calculated value of the correlation was equivalent to 0.8. This suggests that both variables may convey the same information. This was not identified during SOMs analysis (section 7.10). SOMs analysis, being mainly a visual interpretation of patterns in the data, showed to be prone to errors arising from a human interpretation. The use of this method, therefore, may be a limitation. Furthermore, the resolution of soil data was 100 m (section 6.4.6), which indicates that some information about soil is extrapolated. For instance, for rural areas such resolution may be sufficient, however, it may not be high enough for industrial sites where changes in soil are likely to be induced by environmental pollution form various industry activities. The resolution of soil data being 100 m may explain the fact why soil variables did not have an impact on pipe failure rates in the pipe-measured data. The values of soil are possibly less accurate than other variables which for this data were directly measured for each pipe (i.e., static pressure).

The zone-level data showed some impact of soil parameters on pipe failure rates. The extrapolation of static pressure for the base case also showed to be affecting the final results. This may suggests that static pressure extrapolation is less accurate than soil. Soil data indirectly convey information about load (traffic load), therefore this variable was found to have some importance in explaining

pipe repairs for the zone-level data.

8.6 Contribution of pressure transients to pipe repairs

The contribution of pressure transients to pipe repair rates was assessed by adding variables categorising pressure transients to a base case. This formed a base case + transient (BC+T) type of case input to EPR. The final type of case comprised soil and pressure transient variables, i.e. base case + soil + transient (BC+S+T). The results from different cases are discussed in separate sections. A methodology for assessing the contribution of pressure transients on the observed repair data was designed as described in Figure 6.7. Analysis of pipe-measured asbestos cement pipes were compared to cast iron pipe-measured results.

8.6.1 Rate of change of pressure

The interpercentile ranges category derived from the rate of change of pressure were identified as an effective, simple and fast method to initially describe pressure transients (see section 6.2.1). Pressure transients expressed by this method and their possible contribution to the pipe failure rates in cast iron pipes was undertaken by the comparison to the base case as described in Figure 6.7.

8.6.1.1 Pipe-measured results

During the SOMs analysis (see section 7.6) the 99.98th interpercentile range was selected as an input to EPR. Table 8.6 summarises the ranking of variables for the pipe-measured and the zone-level data respectively. The EPR results for the pipe-measured data have shown that the 99.98th interpercentile range appeared in the last model equation. However, no improvement in the CoD value with respect to the base case was seen. The 99.98th interpercentile range was not a main variable but showed to be in positive relationship with pipe repair rates. The 99.98th interpercentile range reduced the exponent of diameter, but did not affect age and static pressure. The simple quantification of changes in pressure utilised by the 99.98th interpercentile range may not be enough to accurately characterise pressure transients in terms of their magnitudes and durations and subsequently their contribution to pipe repair rates. The results from the high quality pipe-measured data suggest that better description of pressure transients is required.

Table 8.6: Comparison of variable ranking for the base case pipe-measured and zone-level data.

Variable ranking	Pipe-measured	Zone-level
1	Static pressure^{2}	$\frac{1}{Diameter^{0.6}}$
2	Age^{2}	$\frac{1}{\text{99.98th interpercentile range}^{2}}$
3	Diameter$^{1.8}$	Age$^{0.4}$
4	99.98th interpercentile range$^{1.1}$	Static pressure$^{0.3}$

8.6.1.2 Zone-level results

The 99.98th interpercentile range in the zone-level data created a better description of pipe repairs than the base case for the zone-level. The 99.98th interpercentile range provided improvement in model fit by 1.09 with respect to the base case at the same level and by 1.17 in level 4, as presented in Table 7.7. The 99.98th interpercentile range appeared in the denominator suggesting that high changes in pressure cause less pipe repairs. This is contrary to the pipe-measured analysis which showed a positive relationship of the 99.98th interpercentile range with pipe repair rates. The inverse relationship observed in the zone-level data may be caused by extrapolated values derived from the assumption that one measuring point is representative of the rest of the system.

Overall, the results from the pipe-measured and the zone-level data suggest that pressure transients represented by the simple categorisation, the 99.98th interpercentile range, have none or marginal effect (respectively) on pipe repair rates according to criteria utilised by this research. The conflicting results show the positive (pipe-measured) and the inverse (zone-level) relationships with pipe repair rates. The inverse relationship observed in the zone-level data may be caused by extrapolation of pressure transients across the zones. The extrapolation of static pressure was already found to have an adverse impact on analysis, as previously discussed in section 8.5.1.

The interpercentile ranges provide an quick option for the companies to use while processing large amounts of high-speed pressure data. Quick data analysis allow faster reaction time. However, the results showed to be limited in terms of precisely describing pressure transient characteristics (magnitude, duration and counts), for which accurate, method was developed, i.e., transient fingerprint.

8.6.1.3 Rate of change of pressure and soil

Analysis of a base case + soil + transient (BC+S+T) for the pipe-measured and the zone-level data were carried out respectively. The results for the pipe-measured and the zone-level are summarised in Table 8.7, for more details see Table 7.12 and Table 7.13.

Table 8.7: Comparison of variable ranking for the base case with soil and the 99.98th interpercentile range (BC+S+T) for the pipe-measured and the zone-level data.

Variable ranking	Pipe-measured base case+soil+transient	Zone-level base case+soil+transient
1	Static pressure^{2}	$\frac{1}{Diameter^{0.6}}$
2	Age^{2}	Soil fracturing$^{0.7}$
3	Diameter$^{1.8}$	$\frac{1}{SoilCorrosivity^{1.3}}$
4	Soil fracturing$^{1.9}$	$\frac{1}{\text{99.98th interpercentile range}^{0.3}}$
5	$\frac{1}{SoilCorrosivity^{1.2}}$	Age$^{0.5}$
6	$\frac{1}{\text{99.98th interpercentile range}^{0.2}}$	Static pressure$^{0.4}$

The EPR analysis of the rate of change of pressure for the pipe-measured data showed that the 99.98th interpercentile range appeared at the last level after all base case and soil variables. It also showed no improvement in the CoD value with respect to the base case, i.e., no contribution to the observed pipe repair rates. The decreasing contribution of this variable was also seen in its inverse relation with pipe repair rates, also seen in the zone-level results. The 99.98th interpercentile range showed no improvement in the CoD value with respect to the base case and it also appeared in the denominator.

The 99.98th interpercentile range for the zone-level data appeared after soil variables, but before age and static pressure. This result is consistent with previous findings (see section 8.6.1.2), where soil variables showed some contribution to the pipe repair rates. The extrapolation of the 99.98th interpercentile range is likely to explain the zone-level results as extrapolation of static pressure with elevation data only poorly correlated with the observed pipe repair rates (discussed in section 8.5.1).

8.6.2 Transient fingerprint

A methodology to select and characterise (by magnitudes and durations) the discrete pressure transient events from time series data was developed using the enhanced peak-trough detection algorithm (see section 5.5). The discrete pressure transient events were then categorised into regions (section 6.2.2), as a reminder see Figure 8.1, and analysed through SOMs and through EPR to facilitate identification of regions that contribute to pipe repairs.

8.6.2.1 Pipe-measured results

The EPR results for the pipe-measured data are summarised in Table 8.8 by variable importance ranking, for more detail see Table 7.8. For the pipe-measured data it is observed that pressure transients contributed to the observed pipe repair rates and improved the model fit by 1.13 and

1.1 in level 3 and 4 receptively (see Table 7.8). The pressure transient categorisation into regions (section 6.2.2) created a better description of pipe repair rates than the base case. Furthermore, it is the first time the improvement in the CoD value with respect to the base case is seen for the pipe-measured data. It therefore means that transient fingerprints have more contribution to the observed repair data than the base case variables.

Specific transient fingerprint categories were identified by EPR as contributing to the pipe repairs. The pressure transient events categorised in region 3 and expressed as $\frac{1}{Region3}$ was returned by EPR in level 3. This category was combined with static pressure (level 3) and provided an improvement in model fit with respect to the base case at the same level by 1.13, see Table 7.8. It also 'replaced' age (seen in level 3 in the base case), i.e. showed to be more important. Level 4 showed region 1 and region 2 combined into $\frac{Region1}{Region2}$. The equation with this ratio improved the model fit by 1.1 in level 4. Therefore, these results support the idea that pressure transients are contributing factors towards pipe repairs and provides an answer to the research question. The results show that pressure transient events from region 1 (high magnitudes and short durations) cause risk for water networks. This ratio was also multiplied by static pressure showing that high static pressure, high number of pressure transients in region 1 and low number of pressure transient events in region 2 are the combination of variables which increase pipe repairs. Furthermore, it shows that an increase of the pressure transient occurrences in region 1 (fast and large) increase a risk of repairs in the water network.

Table 8.8: Comparison of variable ranking for the base case pipe-measured and zone-level data.

Variable ranking	Pipe-measured	Zone-level
1	Static pressure ^{2}	$\frac{1}{Diameter^{0.6}}$
2	$\frac{1}{Region3^{0.7}}$	$\text{Age}^{0.4}$
3	$\frac{Region1^{1.2}}{Region2^2}$	$\frac{1}{Region4^{0.2}}$
4	$\text{Age}^{1.1}$	$\text{Static pressure}^{0.4}$, $\frac{1}{Region1^{0.1}}$
5		$\frac{Region2^{0.1}}{Region1^{0.1}Region4^{0.3}}$

The EPR results for the pipe-measured data showed that only static pressure and age (appeared in the last level) were retained from the set of core variables. As previously seen, age appeared as a positive variable which supports the idea that older pipes fail more often. Diameter did not appear in any EPR outputs in Table 7.14 and also Table 8.8. Therefore, it may not be an important variable as confirmed by previous analysis (base case, section 8.5.1) which showed that most of the data had 101.6 mm diameter and there was very little data of other diameter sizes available.

Large magnitude (but not extreme) transients regularly occurring are less common in water

networks than the events categorised in regions 2, 3 and 4 respectively. However, events captured in region 1 over a typical two week recording period were sufficient to evidence their adverse effect on a water infrastructure. Medium size (in terms of magnitude and duration) pressure transients, represented in region 2, are easier to capture, see Table 7.1. However, for the pipe-measured data these events were in the denominator, i.e. $\frac{Region1}{Region2}$. This shows that high magnitude and regularly occurring events contribute to observed repair rates more than any other region. It also shows that a typical two weeks recording period was sufficient to draw conclusions and that this duration can be applied for monitoring pressure transients in water networks.

8.6.2.2 Zone-level results

The zone-level data results are summarised in Table 8.8, for more details see Table 7.9. Pressure transients provided only a marginal improvement in a CoD value with respect to the base case. The main factors were diameter and age (appeared in the first two levels). Diameter appeared in the denominator, which agrees with the base case. Age appeared as a positive variable supporting previous findings and the idea that older pipes should correlate with pipe repairs. It should be noted that all base case variables have been retained. The observed order of appearance of pressure transient categories were: $\frac{1}{Region4}$ (impact on CoD), $\frac{1}{Region1}$ (no improvement in the CoD) and $\frac{Region2}{Region1Region4}$ (no improvement in the CoD). This is contrary to the findings for the pipe-measured data, i.e. $\frac{1}{Region3}$ (improvement in the CoD), $\frac{Region1}{Region2}$ (improvement in the CoD). The contradicting results from the zone-level data may be explained by the extrapolation of the transient fingerprints due to limited monitoring points per zone. This is consistent with the findings discussed in section 8.2.4 showing that pressure transients are sensitive to location and should, therefore, be measured (or accurately modelled) for each pipe. Therefore, the results for the zone-level data require improvement by providing more accurate pressure transient data, possibly by further field data collection because current models are not sufficient to accurately model pressure transients even in simple networks, as discussed in section 8.2.2. By improving the accuracy of pressure transient data the overall model fit may then improve.

The inconclusive results due to extrapolation were also previously observed in the base case where static pressure values were extrapolated across zones. Therefore, for the zone-level data, the extrapolation of much complex variable as pressure transients are, which showed to be location sensitive, is a limitation. The extrapolation did not provide a reliable results for any variables extrapolated, which was required for the zone-level data. The zone-level data showed that to draw any meaningful conclusions there is a need to obtain better pressure transient data for each pipe.

8.6.2.3 Transient fingerprint and soil

This section explores interaction or pressure transients and soil variables. The pipe-measured results for a base case + soil + transient (BC+S+T) showed that transient fingerprints are contributing to the observed pipe repair records more than soil, see Table 8.9 and for more details Figure 7.14 and Table 7.15. The analysis of transient fingerprint showed that for the pipe-measured data pressure transients appear to be less important than static pressure, but more important than age and soil data, see Table 8.9. This was not observed in the zone-level data, where pressure transients were less important than soil. The limited contribution of pressure transients is likely to be expected based on elevation extrapolated static pressures, which poorly correlated with the observed pipe repair records (see section 8.5.1). Therefore, extrapolated pressure transients for the zone-level data are even less likely to contribute to the observed pipe repairs. Soil data showed a greater impact on pipe repair rates.

Table 8.9: Comparison of variable ranking for the base case pipe-measured and zone-level data.

Variable ranking	Pipe-measured	Zone-level
1	Static pressure^2	$\frac{1}{Diameter^{0.6}}$
2	$\frac{1}{Region3^{0.7}}$	Soil fracturing$^{0.7}$
3	Region $1^{0.8}$	$\frac{1}{SoilCorrosivity^{1.3}}$
4	$\frac{1}{Region2^2}$	Age$^{0.5}$
5	Age$^{1.1}$	$\frac{1}{Region4^{0.3}}$
6	Soil fracturing^2	Static pressure$^{0.5}$, $\frac{1}{Region1^{0.1}}$
7		Region $3^{0.1}$

The accurate measurements and categorisation of pressure transients in the pipe-measured data showed that pressure transients have an adverse effect on pipe failure rates (see section 8.6.2). This evidence provides the answer to the aim of this research. The impact of pressure transients characterised by categorisation of transient fingerprint for the pipe-measured data was more significant than soil (base case + soil) and a simple method characterising pressure transient, i.e., the rate of change of pressure (base case + transient). This may be because the transient fingerprint provide a possible link to fatigue as it tries to account for repeating occurrences of pressure transient events. Fatigue analysis is designed to deal with cycling loading imposed on a pipe (Little and Jebe, 1975; Lemaitre, 1984). Fatigue failures occur when pipe/material is exposed to prolonged and intense (magnitude) cyclic forces (number of occurrences). Therefore, if a pipe is exposed to prolonged pressure transients, the fatigue failure may be a concern. Findings from this research showed the evidence that repeated occurrences of large magnitude pressure transients have an effect on

pipe repairs. The pressure transients counted in region 1 were less frequent than in other regions. Suggesting that even limited number of recurving events of this type captured over a typical two weeks recording period may be sufficient to cause fatigue damage.

The zone-level data showed some improvement in the CoD value, and therefore contribution to the observed pipe repair rates yet this was also affected by extrapolation. Extrapolated transient fingerprints showed to be lesser descriptor than soil variables.

8.6.3 Asbestos cement pipes

Asbestos cement pipe data (section 7.14 Tables 7.16 and 7.17) was investigated to further explore the findings observed in the cast iron pipe-measured data. The order of variable importance for the base case and the base case with the transient fingerprint categorisation is summarised in Table 8.10. The zone-level data was not utilised, as previous findings had already shown that the extrapolation of pressure transients is a limitation adversely affecting the results.

Two data were compared for asbestos cement data following the methodology implemented for cast iron pipes. A base case form was compared with a base case + transient (transient fingerprint categorisation).

Table 8.10: Comparison of variable ranking for the base case and the base case + transient.

Variable ranking	Pipe-measured base case	Pipe-measured base case + transient fingerprint
1	Static pressure$^{1.6}$	Region $1^{1.3}$
2	$\frac{1}{Diameter^{0.7}}$	Static pressure^{2}
3	Age$^{0.3}$	Region 4
4		Region $2^{0.7}$
5		Diameter^{2}
6		$\frac{1}{Region3^{0.8}}$

The results have shown that large and short duration events are more important factors in describing pipe repairs than static pressure. Region 1 counts appeared as the most important variable and improved the fit by 3.18 with respect to the base case at level 2 (see Table 7.17). This finding confirms the results for cast iron analysis (pipe-measured data) where large magnitude and short duration pressure transients were seen as an important factor explaining pipe repair rates. The results also highlighted the fact that region 4 was associated with high repairs which is contrary to cast iron pipe-measured data. The size of asbestos cement data, however, was smaller than for cast iron pipes which may give less confidence in the results obtained.

Static pressure and age appeared in nominator, diameter in the denominator, following trends observed in the literature. However, the case with transient fingerprints showed diameter in

nominator. Some previous studies (Kettler and Goulter, 1985) observed a weak inverse relationship with diameter for asbestos cement pipes.

8.7 Practical applications - lesson learned

This research developed methods to analyse and meaningfully categorise large amounts of high-speed pressure data, which can now be used by water companies in various fields such as planning, proactive network management and maintenance or decision making. Figure 8.4 presents the pictogram results from the pipe-measured analysis according to their risk impact. It represents graphically which regions of pressure transient categorisation are likely to generate a risk to water supply infrastructure.

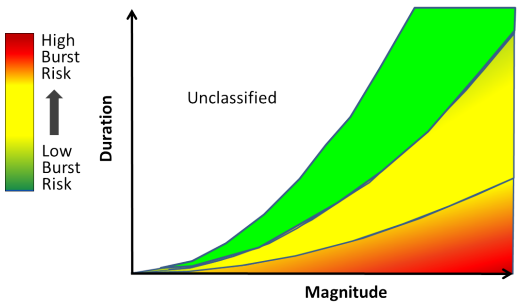


Figure 8.4: Pictogram from the pipe-measured data interpreted according to its risk impact.

Specific pressure transient types which cause the risk to water infrastructure were identified by transient fingerprint categorisation as region 1 (regularly occurring high magnitude and short duration events). The prevention of this specific type of pressure transient will allow cost effective, proactive water infrastructure management against future pipe repairs. Furthermore, a careful control of these pressure transients can be implemented to further reduce costs of repairs.

The developed transient fingerprint categorisation into regions can be used as an effective tool for identification of areas (DMAs, zones) which are at greater risk of failure due to pressure transients. The identification of areas sensitive to pressure transients allows investments to focus on these specific parts of a network providing these with greater attention, monitoring and control. A careful control of areas of high pressure transient risks will reduce costs by implementing more effective (targeted) network improvements or renewal. The understanding provided by this research may be used to also identify pipes with greater risk of repairs or prioritise those to be replaced to help reduce current pipe repairs and aid future replacement and repair decisions. An improved risk identification tool utilising propagation of pressure transients within networks could enhance the prediction of pressure transients related repairs across wider networks. The information about pressure transients at the pipe level, combined with the developed tool, could

further increase accuracy of pressure transient risk alerts and more precise monitoring of not only high pressure prone areas (DMAs, zones) but also specific pipes.

The developed tool could also be applied by cost effective targeting sources of specific type of pressure transients (identified by this research as region 1). This would allow water utilities to directly target pressure transient causes to control and better manage the pressure transient related incidents. Alternatively, pressure transient mitigation devices could be implemented in areas where risk of pressure transients was identified but their sources are challenging to localise.

The tool developed by this research also provide a cost effective alternative for storage of large amounts of 100 Hz pressure transient data. Water companies can now focus on storing only the selected information categorised in region 1 (high risk) to gain an insight on pressure transient risk in their networks at low cost and therefore better manage network structural performance. The pressure transients in water network can be better understood by learning from the additional data. More pressure transient data would also provide more robust and comprehensive EPR model equations for larger network scales.

8.8 Summary

Despite the fact that the pipe-measured data was of limited size and the zone-level data was dominated by non-failed pipes (i.e. 91.8%), the expected forms of the model equations were captured by EPR. Through the discussion of the results it has been observed that static pressure was the main variable explaining pipe repairs for the pipe-measured data. For this data the field measurements of static pressures over the typical two weeks recording period were available for each pipe. The impact of static pressure was not apparent in the zone-level data and static pressure was seen as the least important variable. This is most likely due to extrapolation of static pressure values for this dataset.

The rate of change of pressure for the pipe-measured data showed no improvement in the CoD value with respect to the base case. Therefore, no impact on pipe repair rates can be concluded. However, for the pipe-measured data the 99.98th interpercentile ranges appeared in the last level as a positive variable, which should be noted. As it was already highlighted, the categorisation of pressure transients by the 99.98th interpercentile ranges do not fully capture pressure transient properties, i.e. do not accurately convey information about their magnitudes, durations and number of occurrences (associated with fatigue).

The zone-level results showed some marginal improvement in the CoD value, however, the observed relation with pipe repair rates was inverse. This may be the result of extrapolation of the 99.98th interpercentile ranges for the zone-level data. For the zone-level data soil variables

were more important than the extrapolated 99.98th interpercentile ranges. The base case results showed that extrapolation of static pressure adversely affected the results, despite the fact that the static pressure values were adjusted by elevation and redefined zones. It is therefore concluded that extrapolation of the 99.98th interpercentile ranges also adversely affected the results. Furthermore, more precise characterisation of pressure transients in terms of their exact magnitudes, duration and number of occurrences is necessary to accurately assess their contribution to the observed pipe repair rates as the 99.98th interpercentile ranges did not fully capture pressure transient properties, as also concluded for the pipe-measured data.

The more advanced characterisation of pressure transients was provided by the transient fingerprint. The pipe-measured results showed that pressure transients contribute to the pipe repair rates. Transient fingerprint developed by this research, showed, by the EPR eqations, an improvement in the CoD value with respect to the base case. This also shows that it was possible to answer the aim of this research. Pressure transient events in region 1 showed an adverse impact of pressure transients to observed repair rates in the pipe-measured data. The EPR equations showed that high static pressure, high numbers of events in region 1 and low number of events in region 2 are combination that contributed to the description of the pipe repair rates. The association of these variables with high repairs was proven despite the fact that number of occurrences of large events was lower than other types of pressure transients captured over a typical two weeks recording period. The pressure transient events in region 1 showed an adverse effect on pipes which can possibly be explained by fatiguing material which eventually may trigger pipe failures. The adverse impact of repeatedly occurring high magnitude pressure transients was also seen for asbestos cement pipe material. For the asbestos cement data, pressure transients from regions 1, 2 and 4 showed a contribution to observed repair rates.

The zone-level results for transient fingerprint showed no apparent contribution of pressure transients for the zone-level data which is expected as this require extrapolation. Based on the fact that it was shown that extrapolation was suspect for the static pressure extrapolation with elevation and knowing that transient fingerprints are far more complex and sensitive to location within a network, no effect of transient fingerprint can be expected for the zone-level data. An improvement of pressure transient data for the zone-level data is therefore desirable, as current results are suspect due to extrapolation. The precise pressure transient data that was used in the pipe-measured dataset allowed to answer the aim of this research.

Chapter 9

Conclusions and Future work

This work has developed novel methods to identify, characterise and classify pressure transient events measured in operational drinking water distribution systems. From these methods it becomes possible to assess the contribution of pressure transients to observed pipe repair data. A novel method characterising pressure transients, transient fingerprint, showed that pressure transients contributed to the observed pipe repair data. Applying this understanding allows pipes with pressure transients that contribute to pipe repairs to be identified, such that remedial strategies could be implemented. This could lead to reduction in future pipe repairs by enabling focused investment in improvements and network renewal.

Conclusions from this work are:

1. The use of high speed pressure data collection at 100 Hz allowed the wide spread occurrence of pressure transients across diverse networks to be observed, based on 79 monitored zones, for the first time. High speed monitoring has given new insights into the behaviour of pressure transients. The majority of monitored sites had transients between 5-10 m. Extreme (infrequent) and small (regular) scale transients were recorded and all zones had transient activity.

2. The following assumptions and limitations of the work needs to be taken into account when interpreting the results: two weeks of data collected during the field work were used for static pressure and pressure transient analysis, low values of the CoD were seen for all the results from the zone-level dataset. The research analysed the largest data available, however this data were limited due to small variety of pipe diameters and material types.

3. Evolutionary polynomial regression identified plausible relationships between variables and pipe repair data on 64 cast iron pipes. These trends were previously reported in the literature which supports the results. The analysis for cast iron pipes showed that pipes less prone to

repairs had the following characteristics: large diameter, relatively young with complex relation to soil parameters.

4. An initial analysis using a simple method to characterise pressure transients, the rate of change of pressure, derived from traditional ways of describing variability in the continuously recorded time series signal, showed some impact of pressure transients on water networks. Evolutionary polynomial regression and self organising maps analysis identified a potential impact of pressure transients based on the interpercentile ranges. However, the results were inconclusive.

5. Enhanced peak-trough detection algorithm was developed that facilitated identification of discrete pressure transient events that sought their magnitude, duration and number of occurrences from continuously recorded time series data. The novel method yielded an unique transient fingerprint output for each time signal recorded.

6. Pressure transient classification developed from transient fingerprint allowed, for the first time, identify magnitude and duration of pressure transients that contribute to pipe repairs and their association with fatigue impact. Evolutionary polynomial regression analysis showed that short duration and large magnitude pressure transients have adverse impact on pipes.

7. Measuring points were sufficient to show contribution of pressure transients to pipe repair rates from the pipe-measured data. Zone-level results were limited due to extrapolation of pressure transients. For the zone-level analysis more pressure transient data or accurate modelling of pressure transient propagation is required.

8. Analysis carried out on small subset of pipe-measured asbestos cement pipes confirmed an adverse impact of pressure transients.

Future work

The following areas of future work have been identified as providing the potential to further benefit understanding gained by this research:

- This research identified that pressure transients are parameters sensitive to locations (location specific). To obtain accurate pressure transient propagation data a high spatial resolution should be considered when collecting the data. Alternatively, a pipe level pressure transient model is required precisely representing pressure transient propagation in complex networks. This additional data could be used to verify the results from the zone-level analysis. Greater

understanding of pressure transient propagation would therefore provide more data to target those networks which are at greater risk of failures due to pressure transients.

- This research has found that time (duration), an additional characteristic of pressure transient, could benefit a future pressure transient analysis through laboratory tests. Structural tests would allow assessment of which magnitudes and durations cause a structural pipe failures. Such analysis could ultimately lead to the development of novel 'S-N-T' curves.

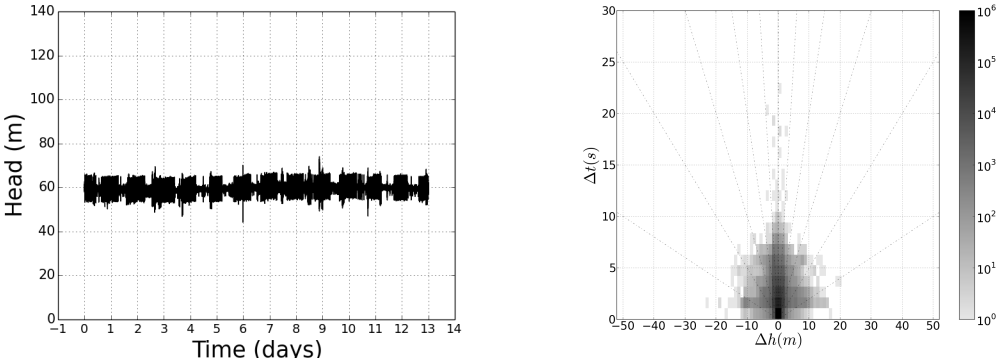
- This research focused on cast iron and a subset of asbestos cement pipes for analysis due to the limited number of pipe failure records in the remaining material datasets. Other material types could be investigated as future work, including the wide range of viscoelastic materials. Ideally these materials should have pressure transients directly measured or accurately modelled. In addition, the majority of cast iron pipes were of 101.6 mm diameter which makes the results applicable to a specific diameter range. Future work should consider greater variability of pipes in terms of pipe characteristic such as materials, diameters. This could possibly improve the low CoD values seen in the zone-level dataset.

- This research focused on investigating the contribution of pressure transients to the observed pipe repair data. There is a possibility, however, that pressure transients could also contribute to water quality failures. Further work would address whether pressure transients are contributing factors towards material mobilisation (discolouration or mobilisation of deposited material) due to dynamic accelerating and decelerating velocity profiles exerting shear stress for material mobilisation from pipe walls.

Appendices

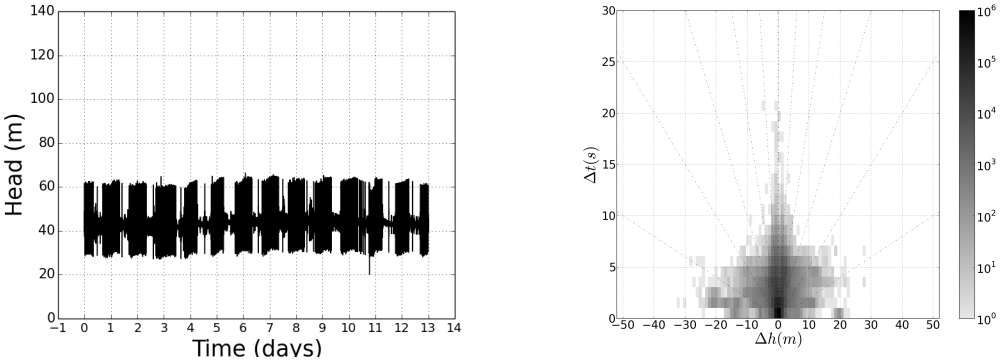
Appendix A

Transient fingerprint



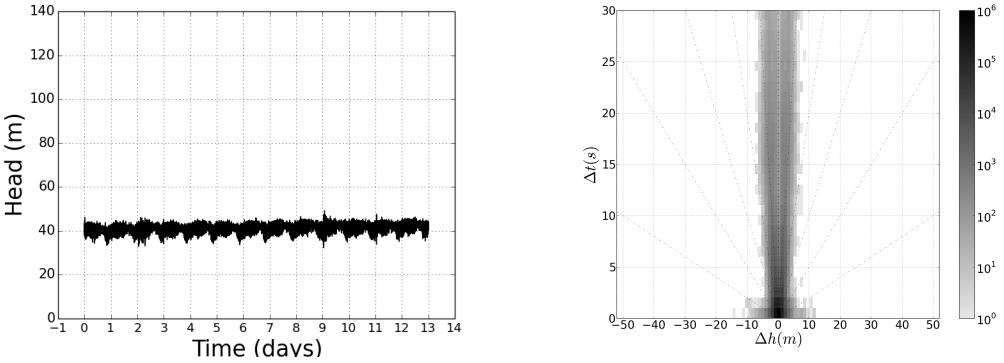
(a) **(b)**

Figure A.1: Zone 1, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



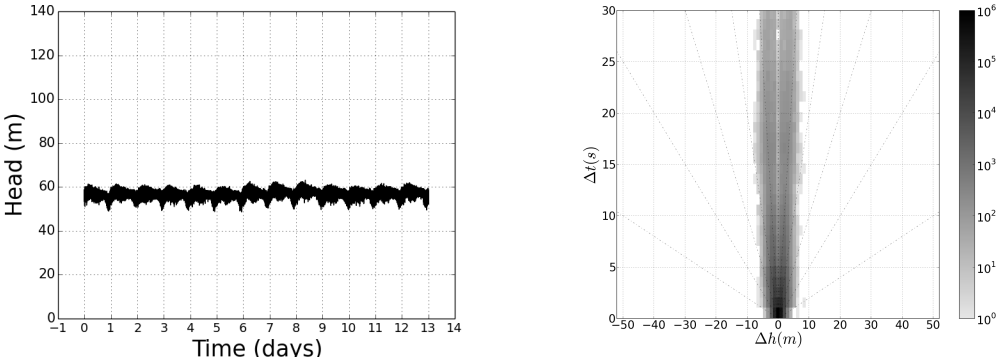
(a) **(b)**

Figure A.2: Zone 1, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



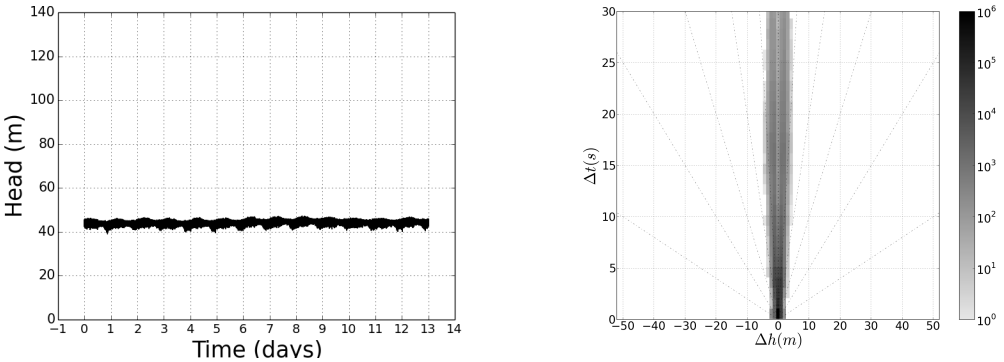
(a) **(b)**

Figure A.3: Zone 2, location 1, (a) pressure transients recorded at 100Hz and (b) its transient fingerprint.



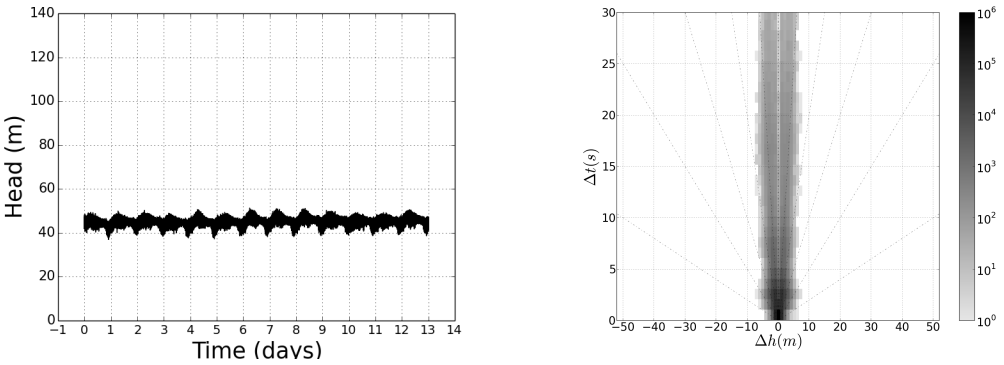
(a) **(b)**

Figure A.4: Zone 2, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



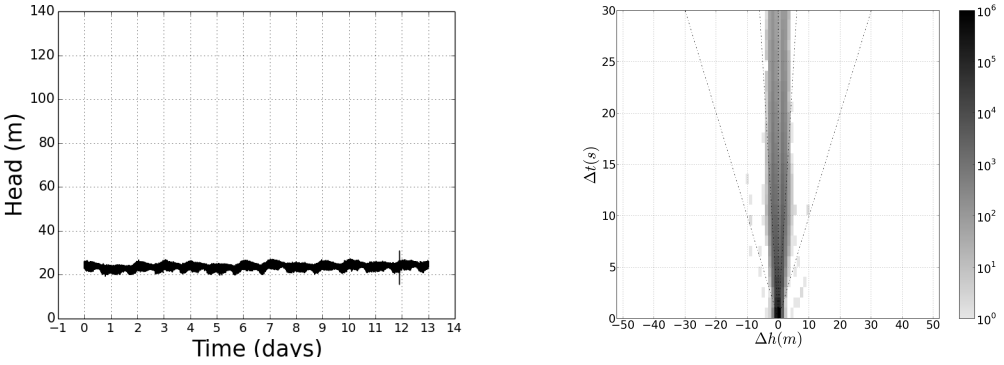
(a) **(b)**

Figure A.5: Zone 2, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



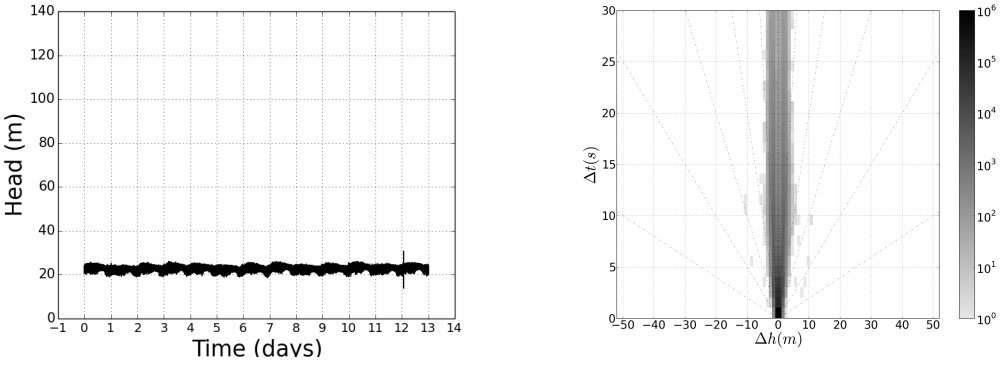
(a) **(b)**

Figure A.6: Zone 2, location 4, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



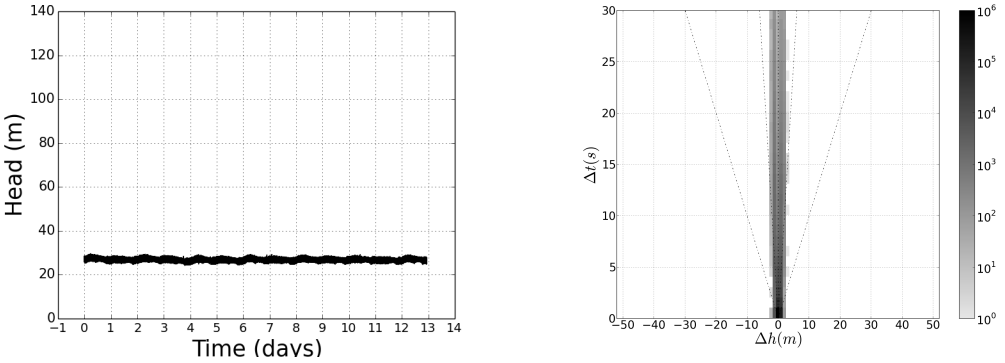
(a) **(b)**

Figure A.7: Zone 3, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



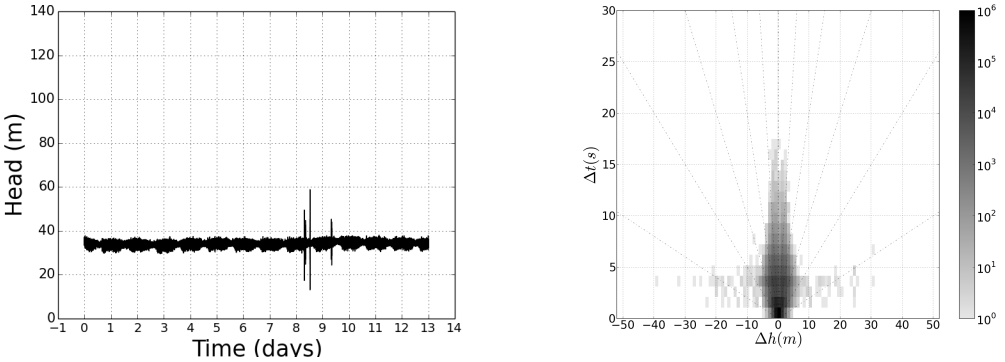
(a) **(b)**

Figure A.8: Zone 3, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



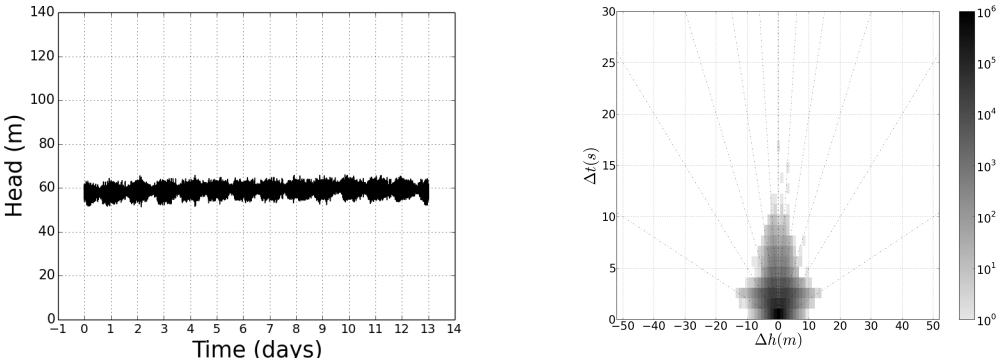
(a) **(b)**

Figure A.9: Zone 3, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



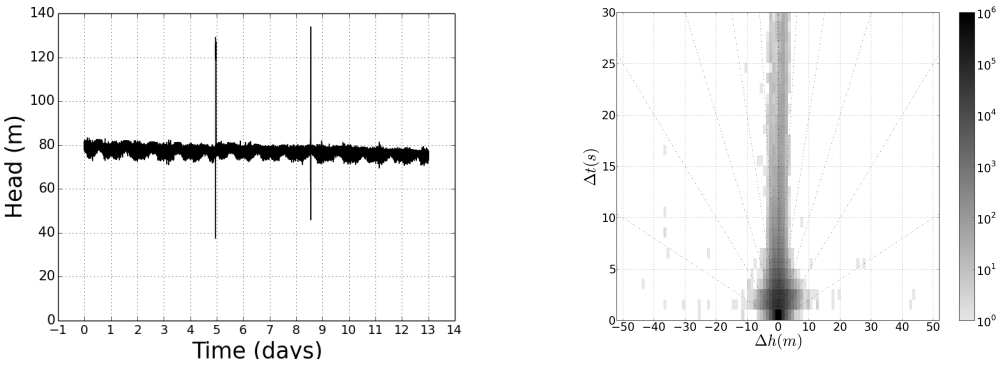
(a) **(b)**

Figure A.10: Zone 4, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



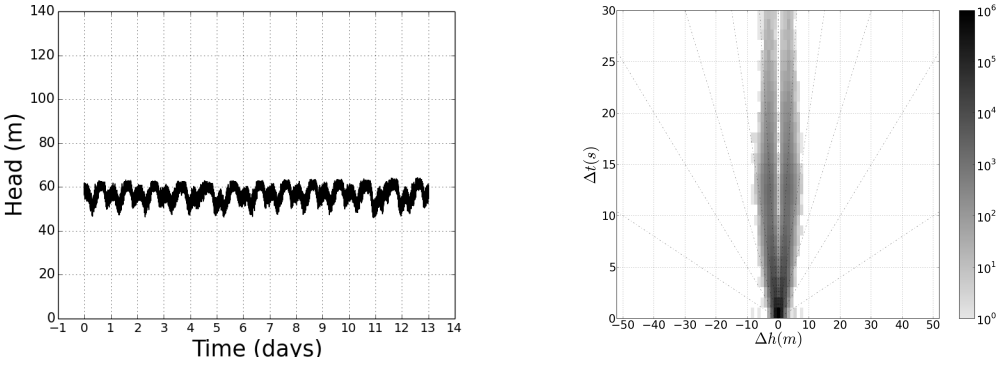
(a) **(b)**

Figure A.11: Zone 4, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



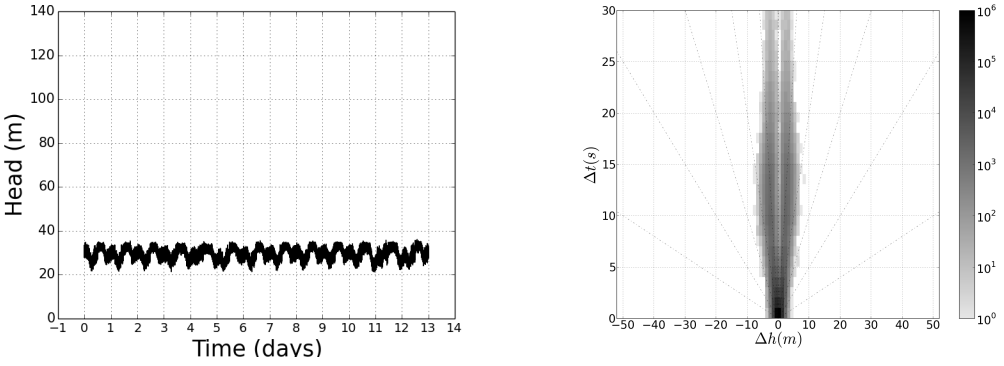
(a) **(b)**

Figure A.12: Zone 5, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



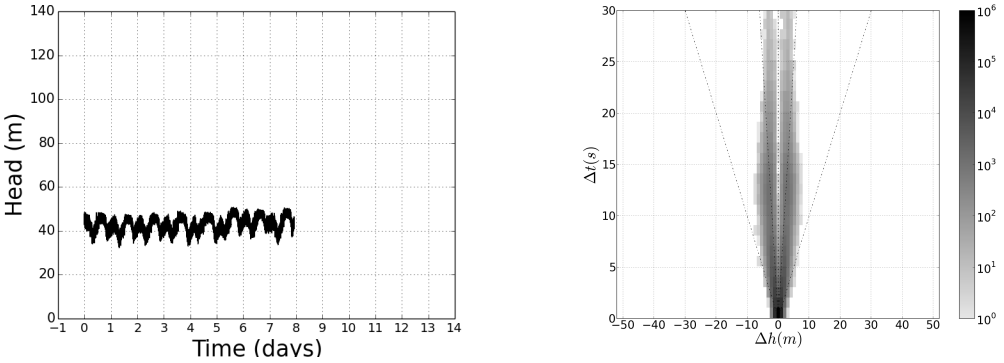
(a) **(b)**

Figure A.13: Zone 6, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



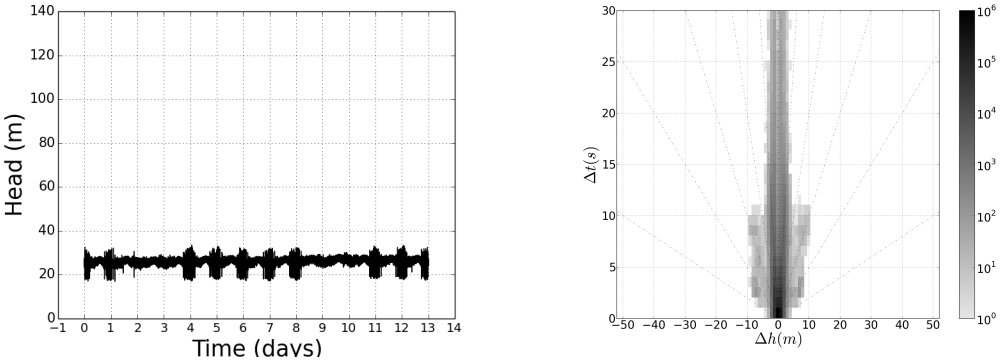
(a) **(b)**

Figure A.14: Zone 6, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



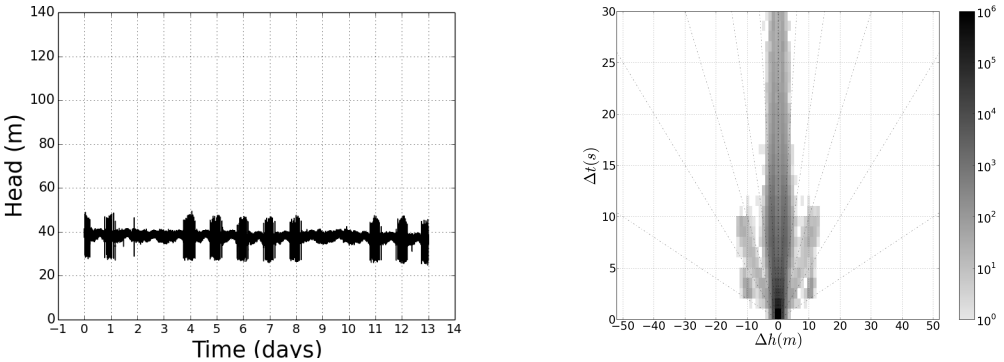
(a) **(b)**

Figure A.15: Zone 6, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



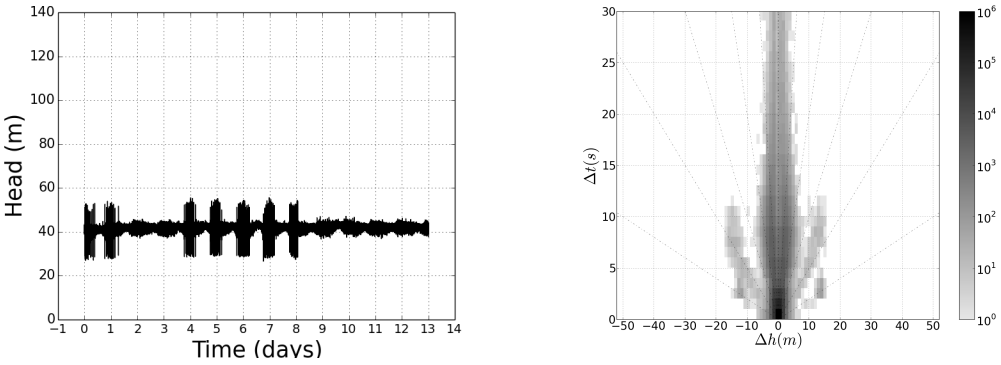
(a) **(b)**

Figure A.16: Zone 7, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



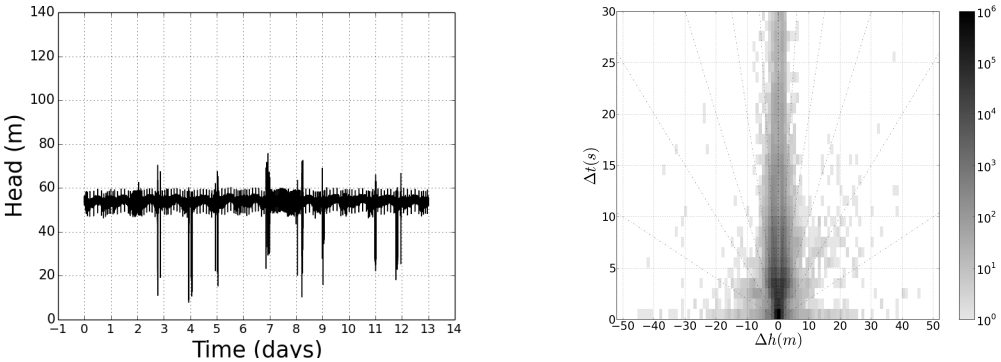
(a) **(b)**

Figure A.17: Zone 7, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



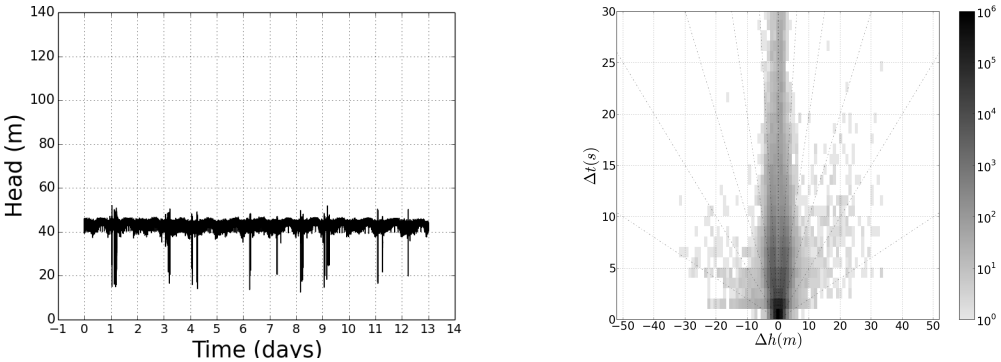
(a) **(b)**

Figure A.18: Zone 7, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



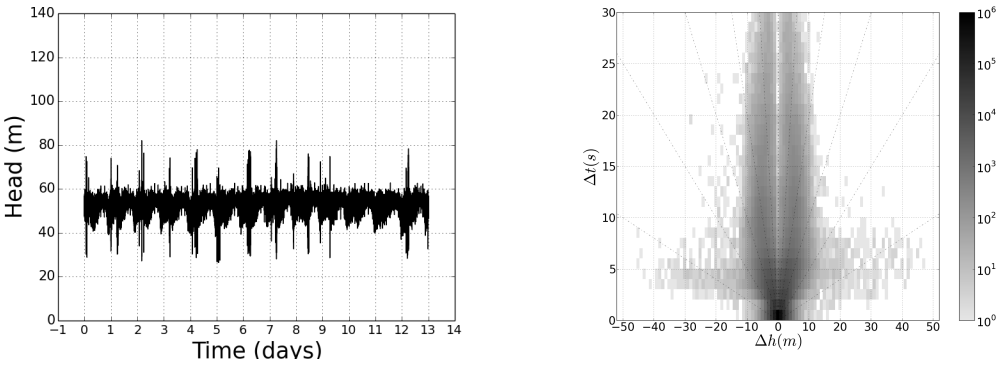
(a) **(b)**

Figure A.19: Zone 8, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



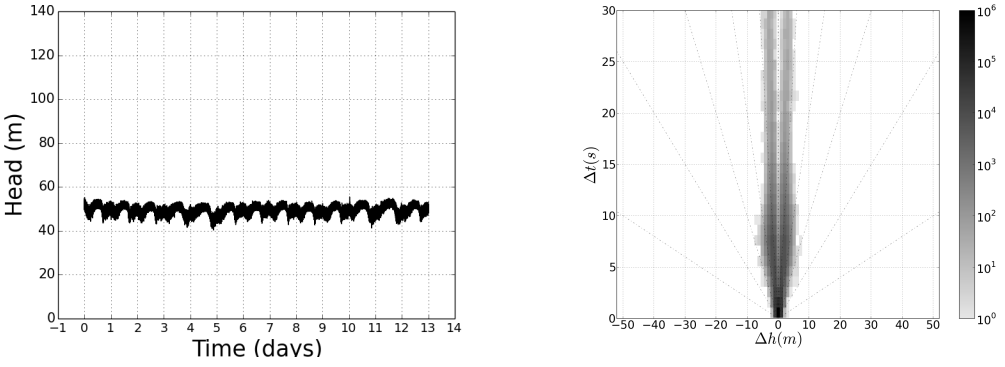
(a) **(b)**

Figure A.20: Zone 8, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



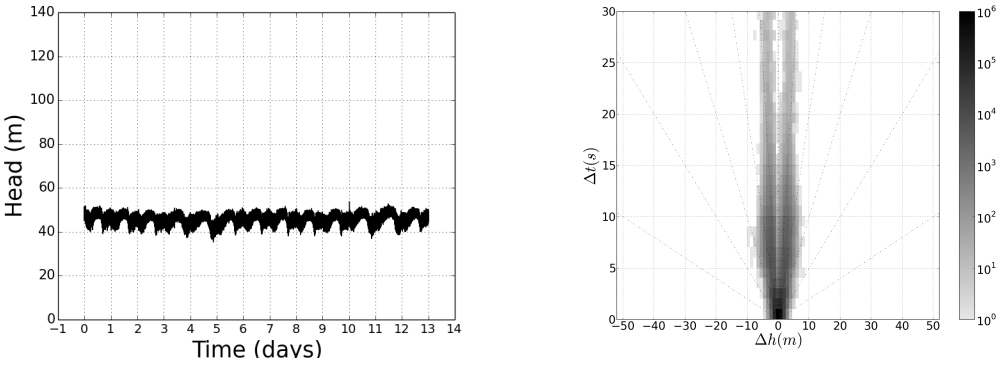
(a) **(b)**

Figure A.21: Zone 9, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



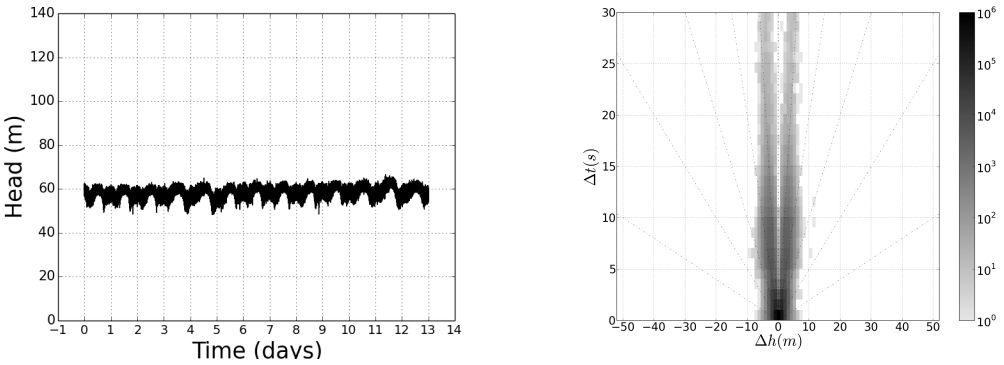
(a) **(b)**

Figure A.22: Zone 10, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



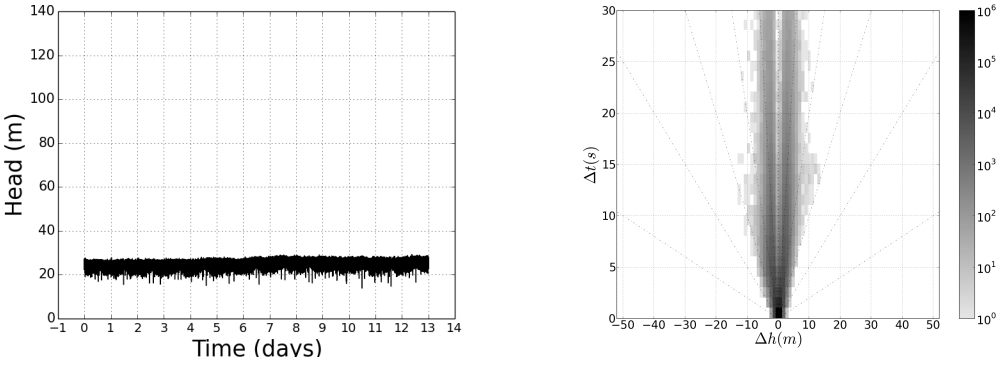
(a) **(b)**

Figure A.23: Zone 10, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



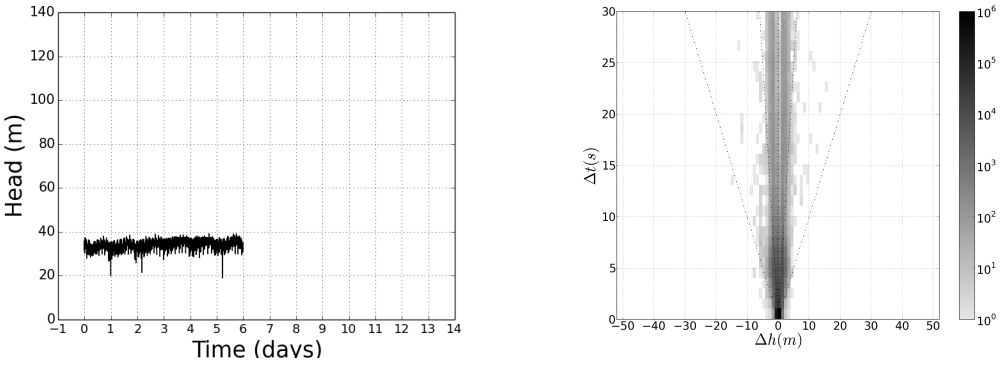
(a) **(b)**

Figure A.24: Zone 10, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



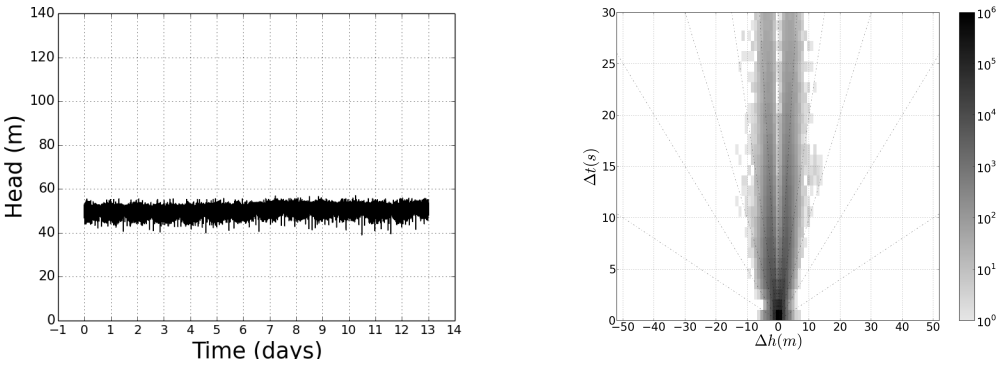
(a) **(b)**

Figure A.25: Zone 11, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



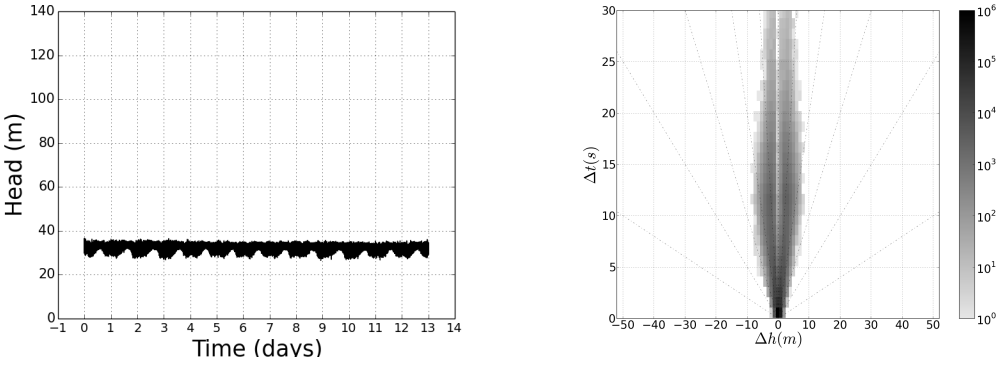
(a) **(b)**

Figure A.26: Zone 11, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



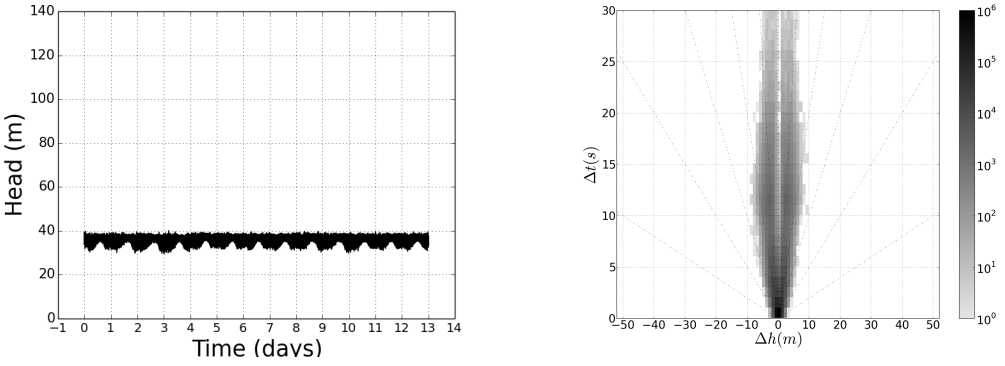
(a) **(b)**

Figure A.27: Zone 12, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



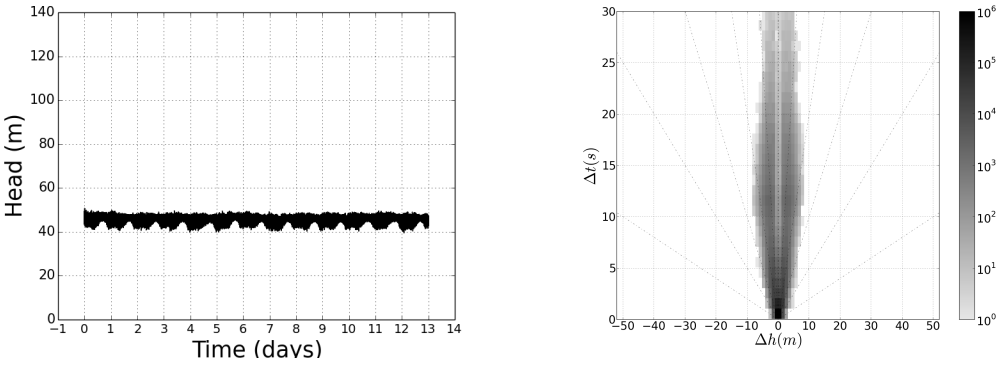
(a) **(b)**

Figure A.28: Zone 13, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



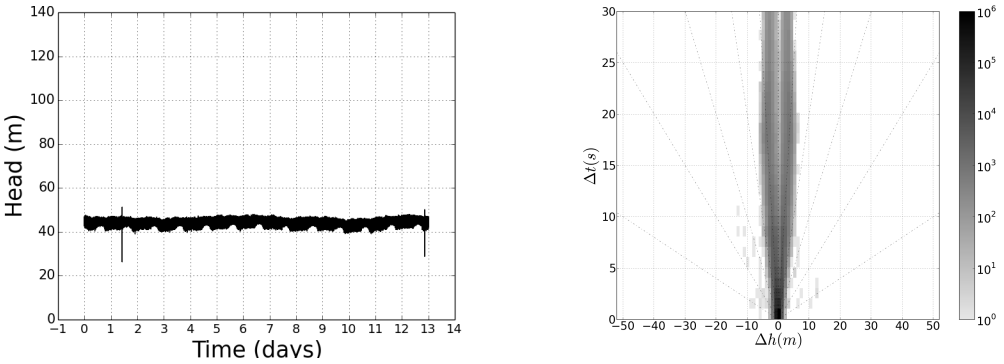
(a) **(b)**

Figure A.29: Zone 13, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



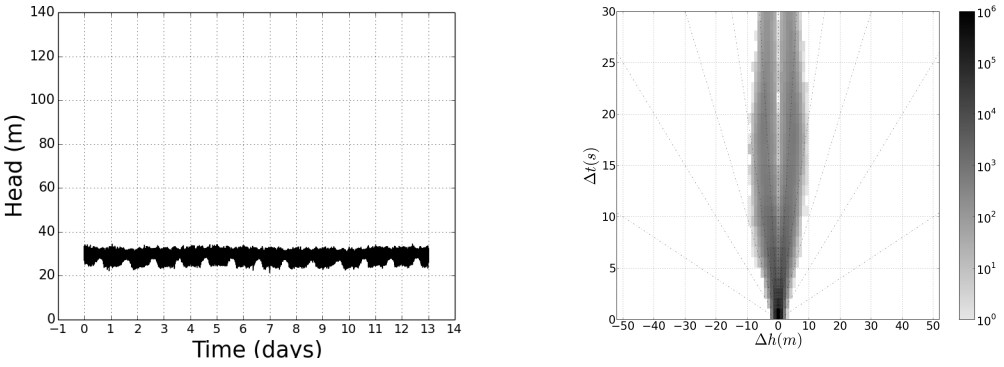
(a) **(b)**

Figure A.30: Zone 13, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



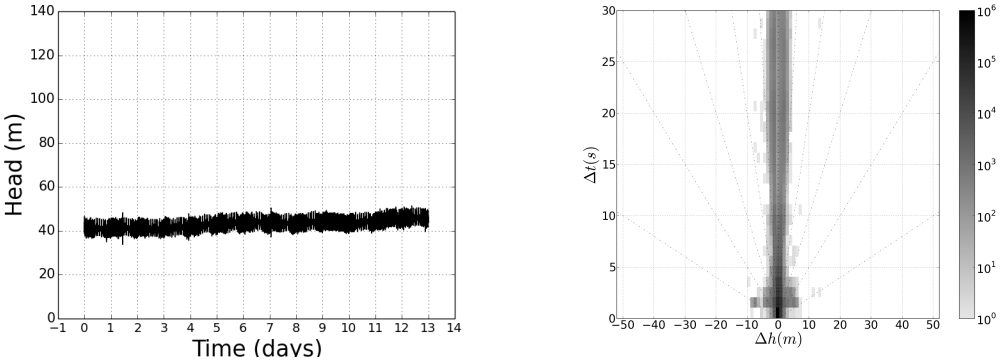
(a) **(b)**

Figure A.31: Zone 14, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



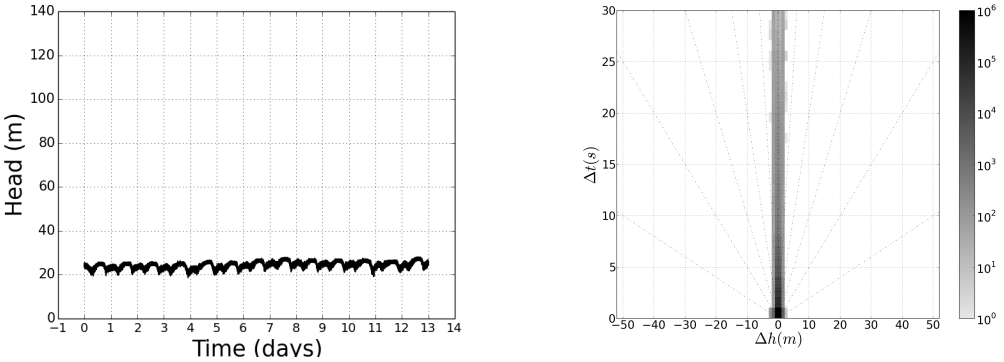
(a) **(b)**

Figure A.32: Zone 14, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



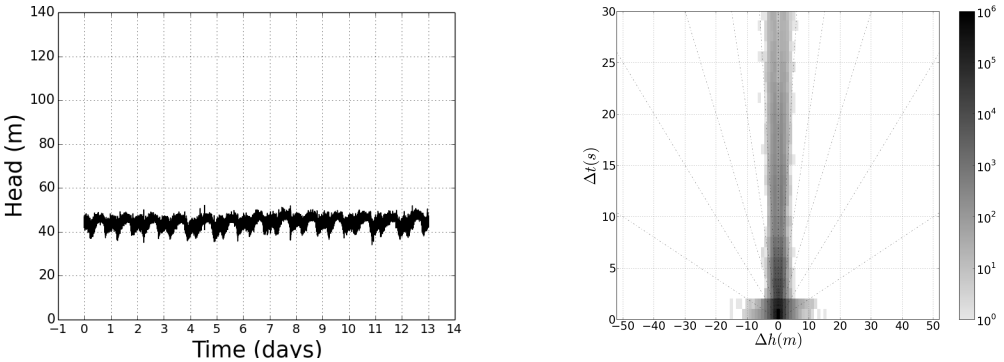
(a) **(b)**

Figure A.33: Zone 15, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



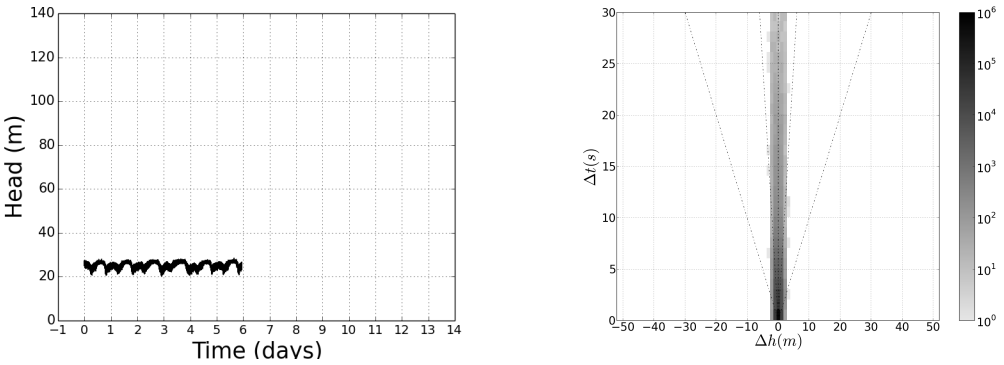
(a) **(b)**

Figure A.34: Zone 16, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



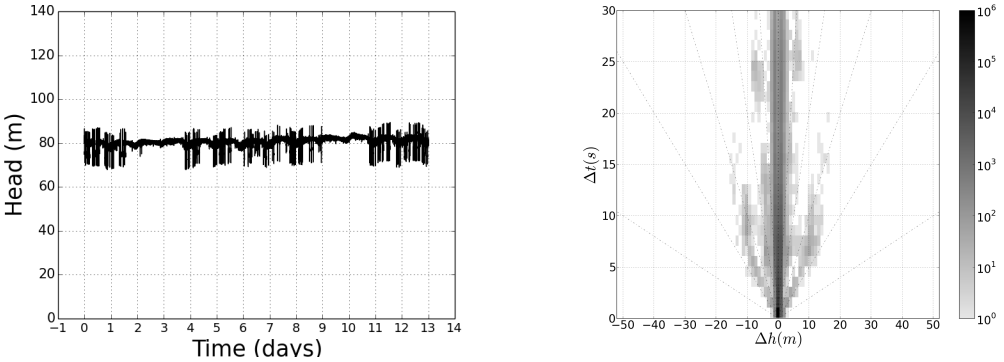
(a) **(b)**

Figure A.35: Zone 16, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



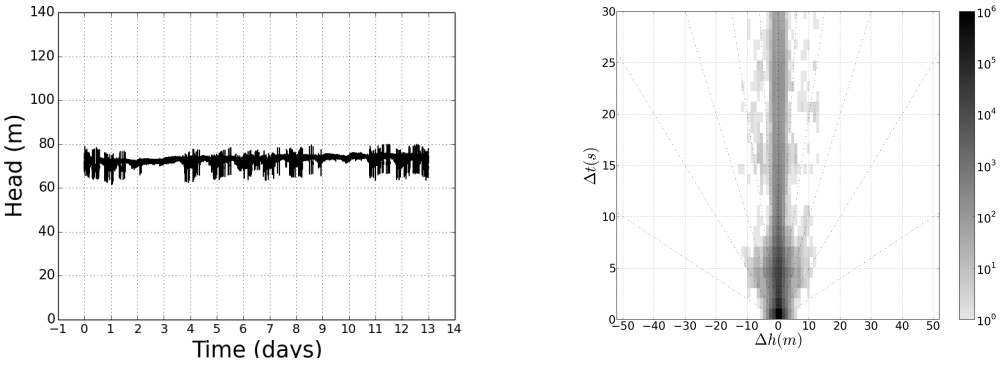
(a) **(b)**

Figure A.36: Zone 16, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



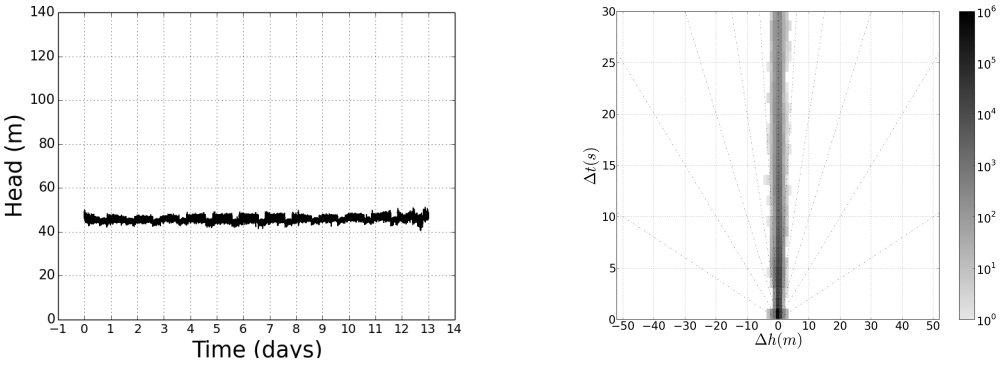
(a) **(b)**

Figure A.37: Zone 17, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



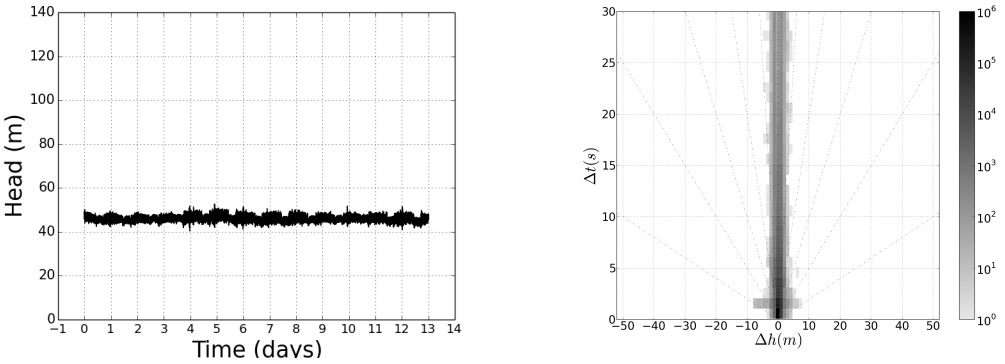
(a) **(b)**

Figure A.38: Zone 17, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



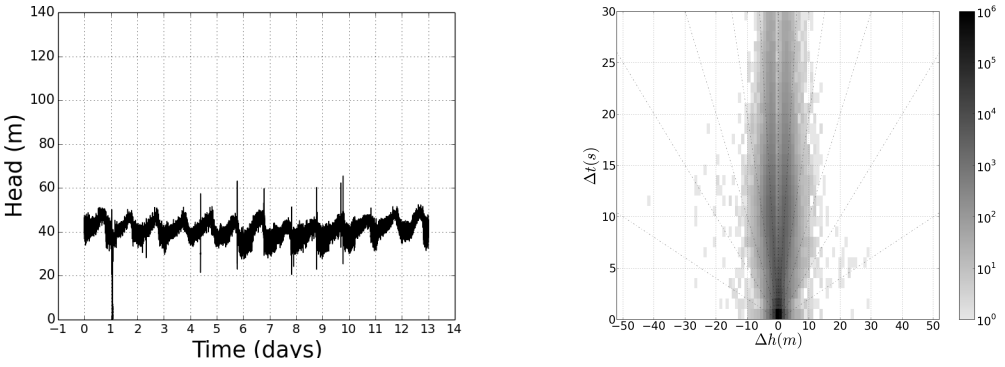
(a) **(b)**

Figure A.39: Zone 18, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



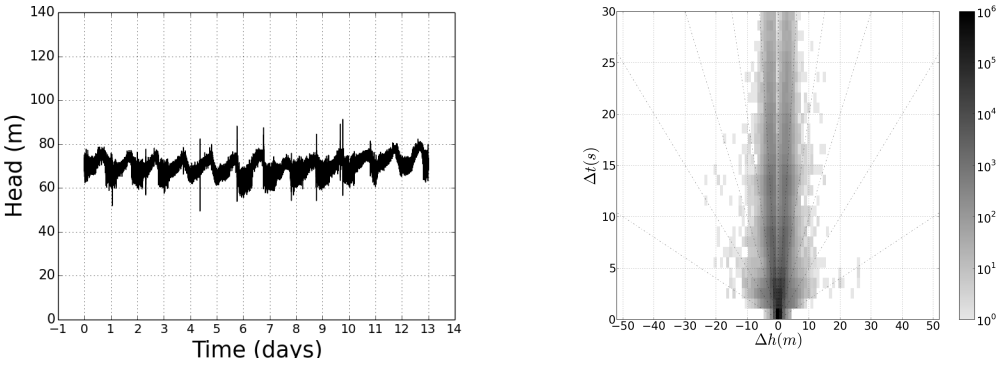
(a) **(b)**

Figure A.40: Zone 18, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



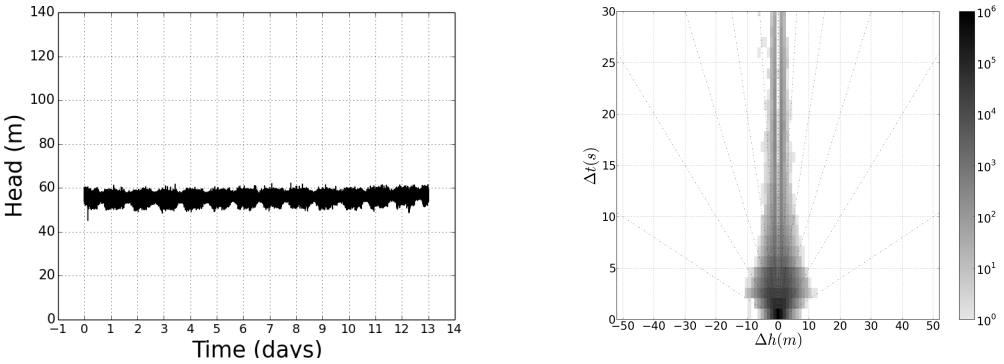
(a) **(b)**

Figure A.41: Zone 19, location 1, (a) pressure transients recorded at 100 Hz and its transient fingerprint.



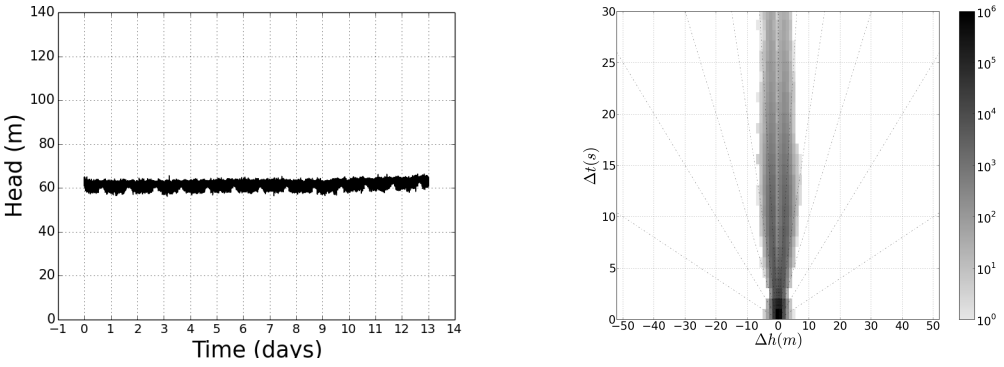
(a) **(b)**

Figure A.42: Zone 19, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



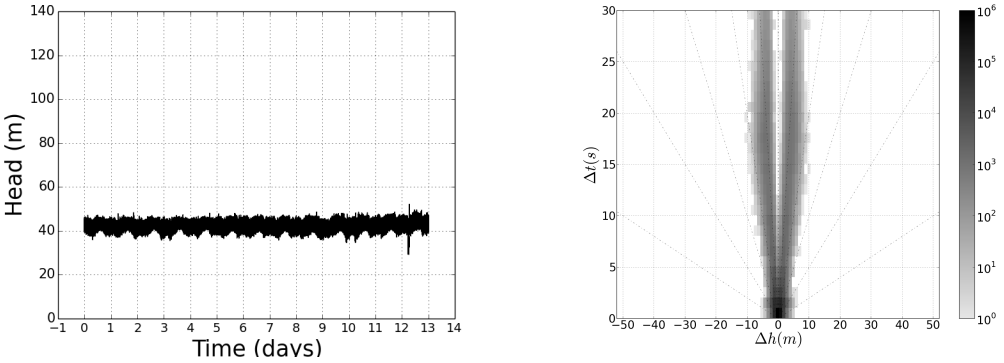
(a) **(b)**

Figure A.43: Zone 20, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



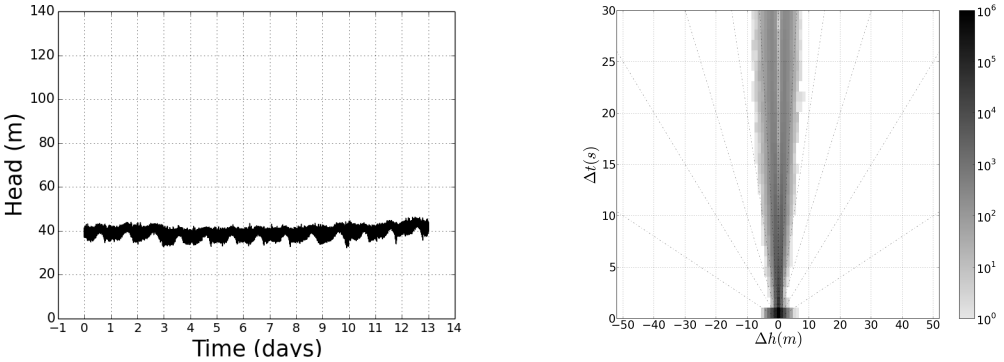
(a) **(b)**

Figure A.44: Zone 21, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



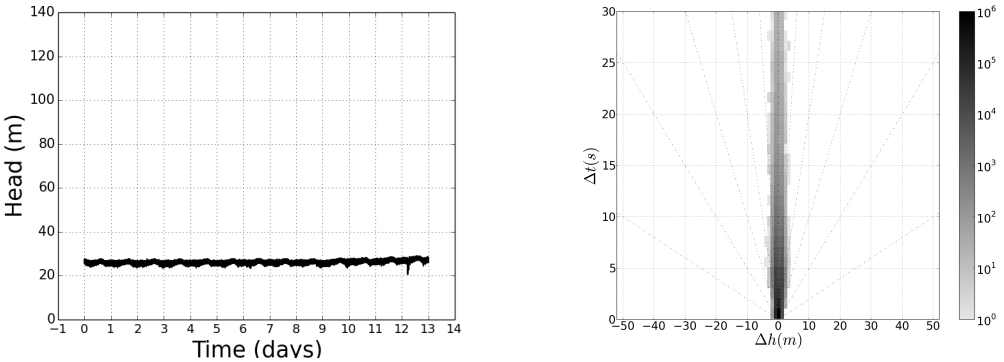
(a) **(b)**

Figure A.45: Zone 22, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



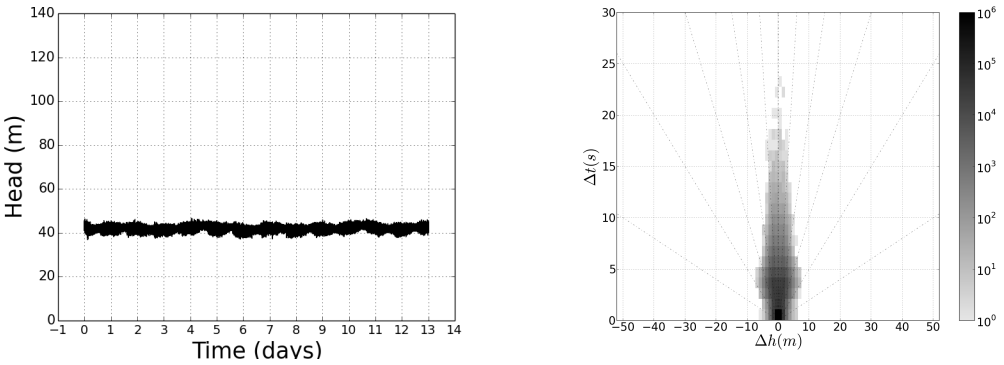
(a) **(b)**

Figure A.46: Zone 23, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



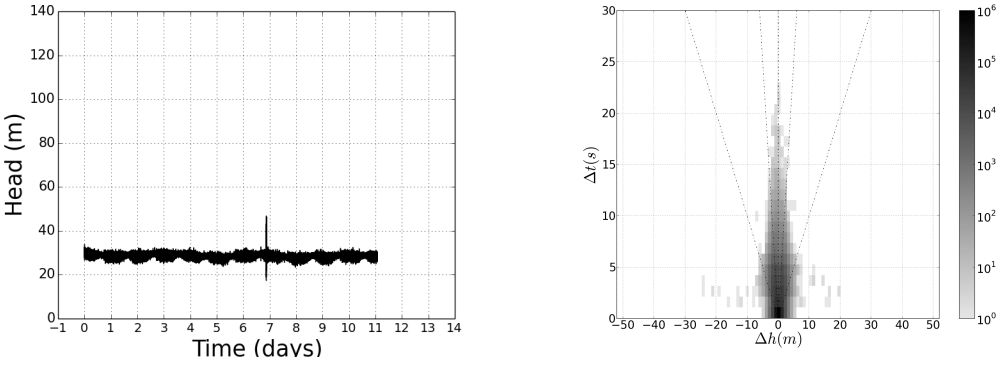
(a) **(b)**

Figure A.47: Zone 24, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



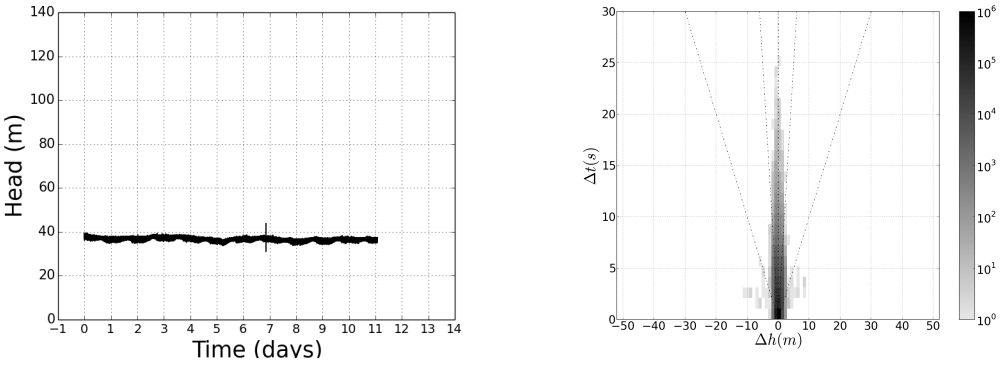
(a) **(b)**

Figure A.48: Zone 25, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



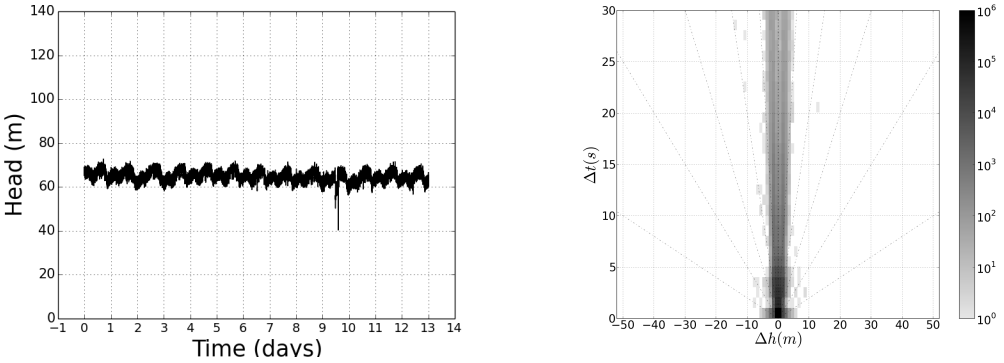
(a) **(b)**

Figure A.49: Zone 25, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



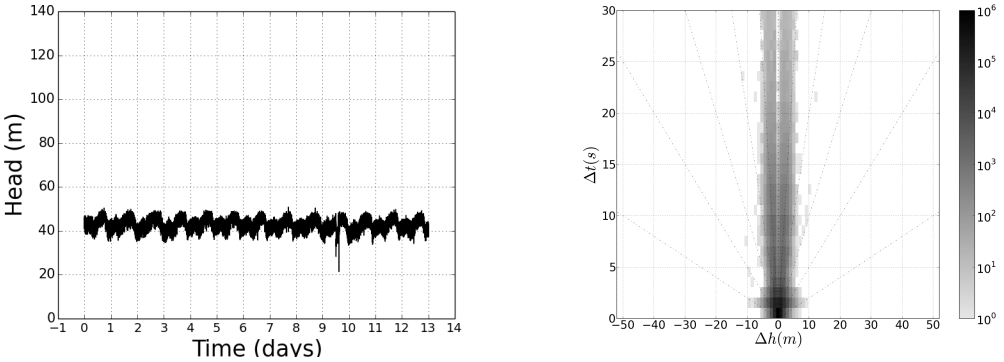
(a) **(b)**

Figure A.50: Zone 25, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



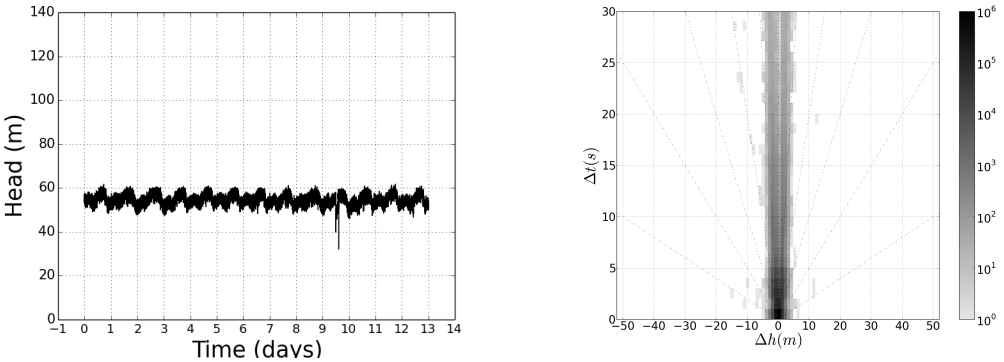
(a) **(b)**

Figure A.51: Zone 26, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



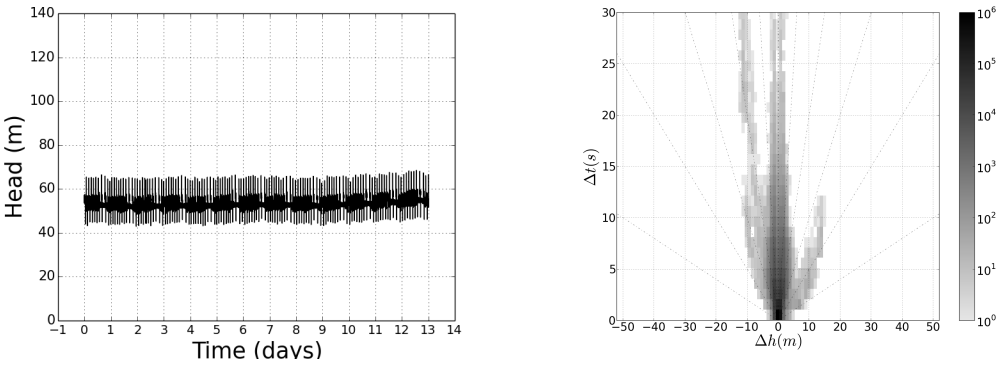
(a) **(b)**

Figure A.52: Zone 26, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



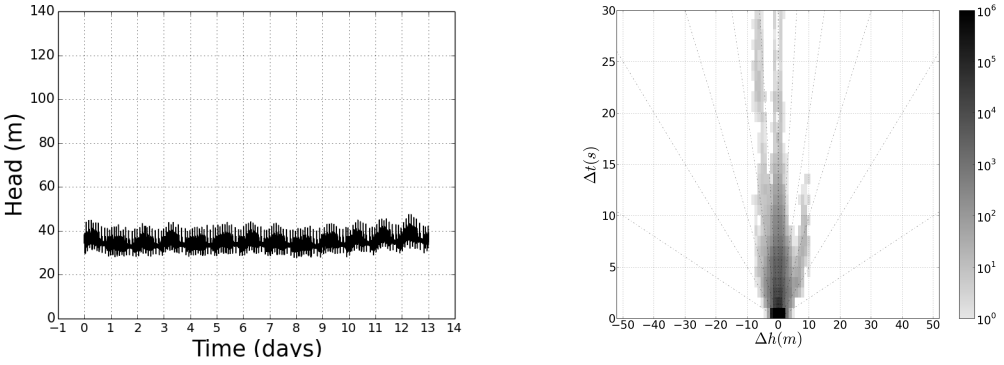
(a) **(b)**

Figure A.53: Zone 26, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



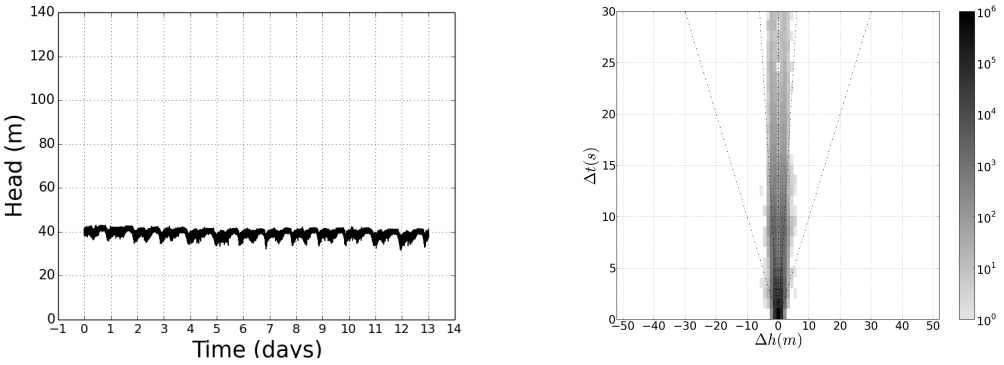
(a) **(b)**

Figure A.54: Zone 27, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



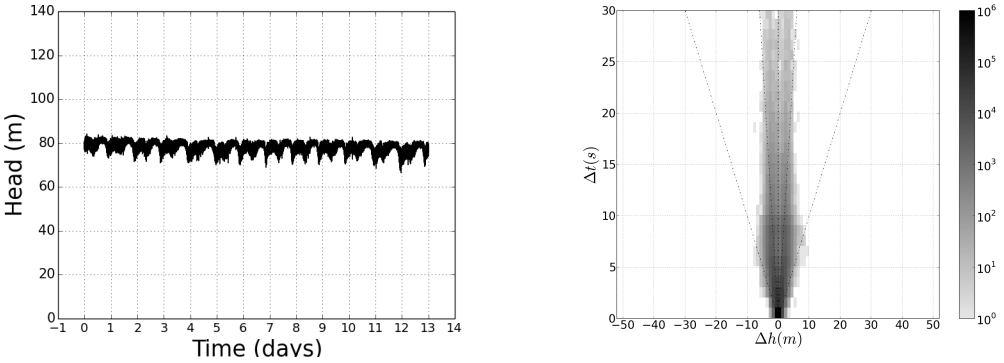
(a) **(b)**

Figure A.55: Zone 27, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



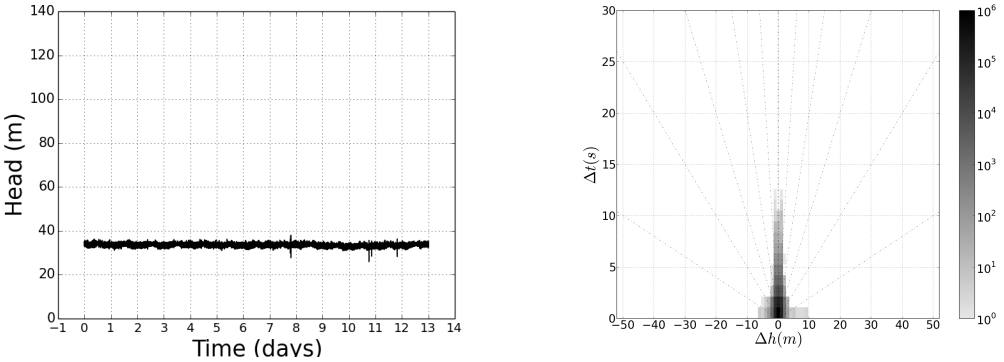
(a) **(b)**

Figure A.56: Zone 28, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



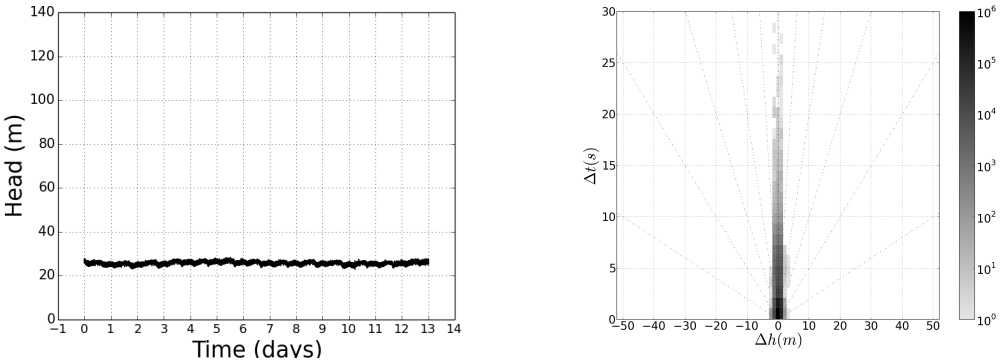
(a) **(b)**

Figure A.57: Zone 28, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



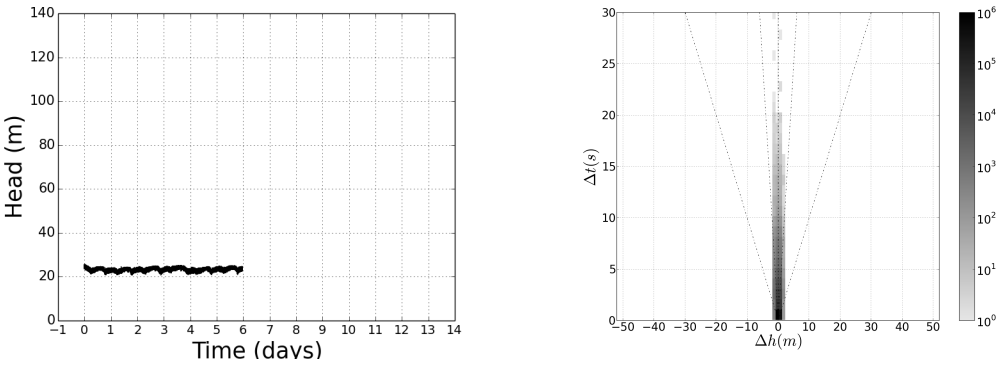
(a) **(b)**

Figure A.58: Zone 29, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



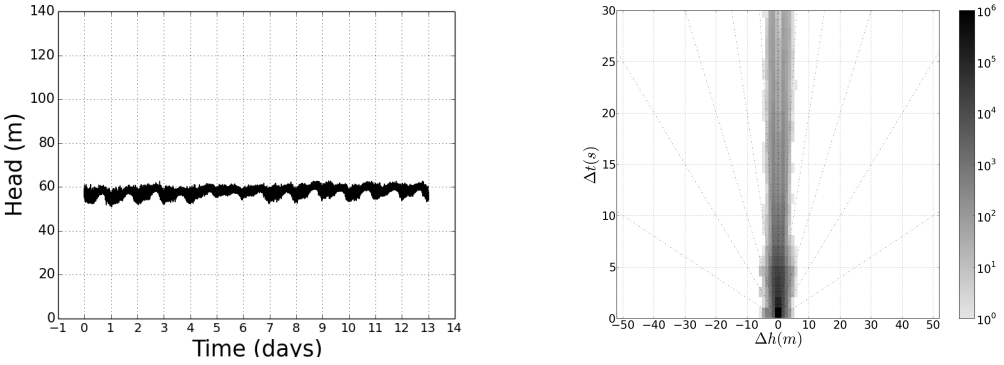
(a) **(b)**

Figure A.59: Zone 29, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



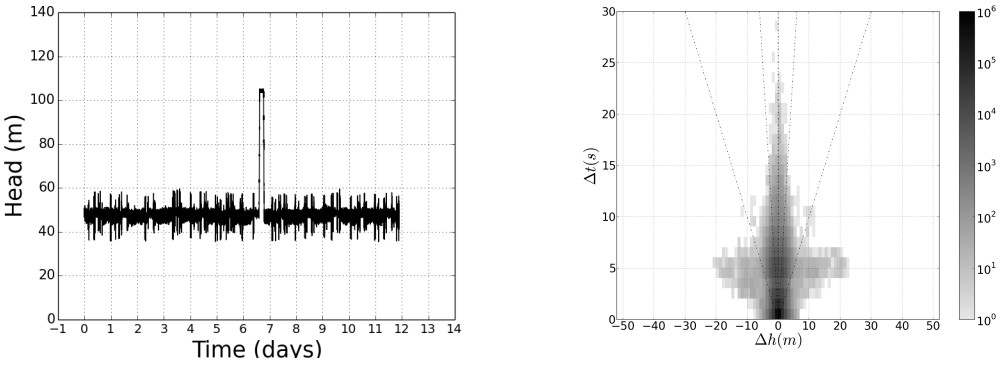
(a) **(b)**

Figure A.60: Zone 29, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



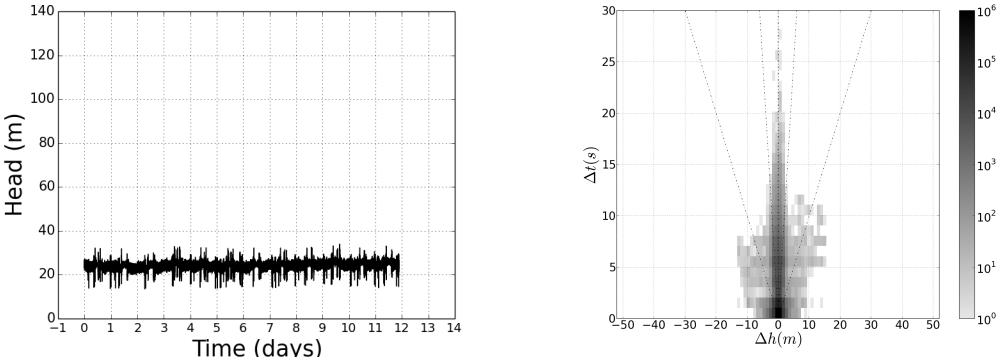
(a) **(b)**

Figure A.61: Zone 30, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



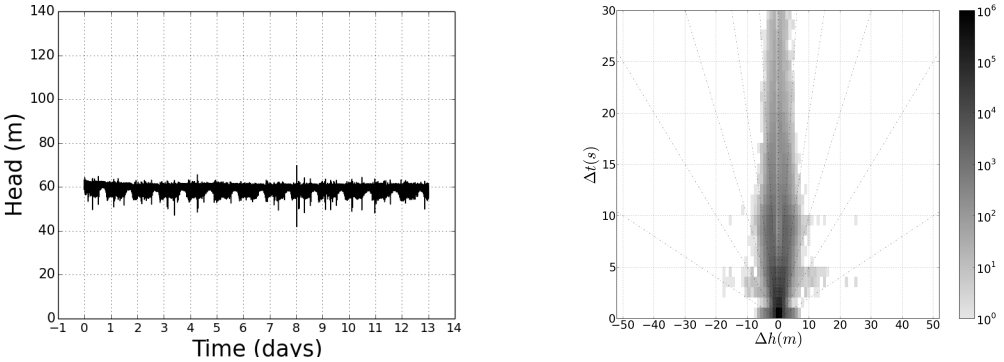
(a) **(b)**

Figure A.62: Zone 31, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



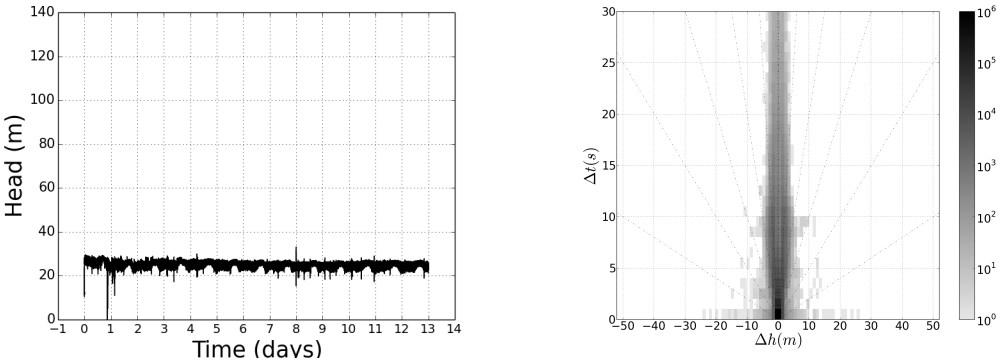
(a) **(b)**

Figure A.63: Zone 31, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



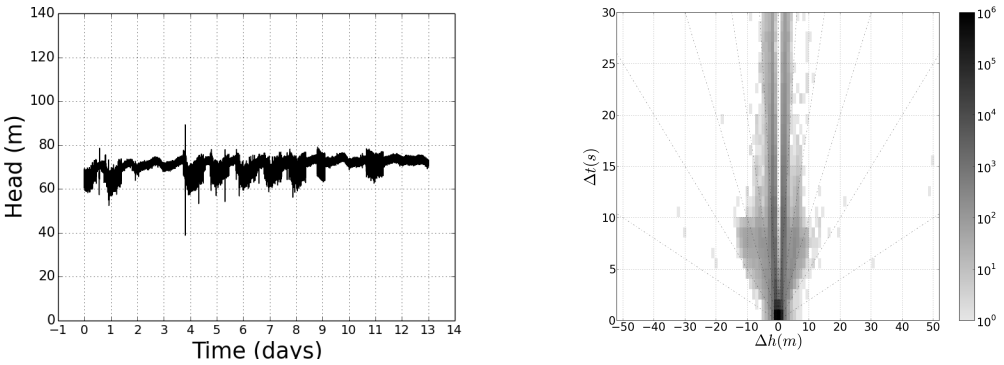
(a) **(b)**

Figure A.64: Zone 32, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



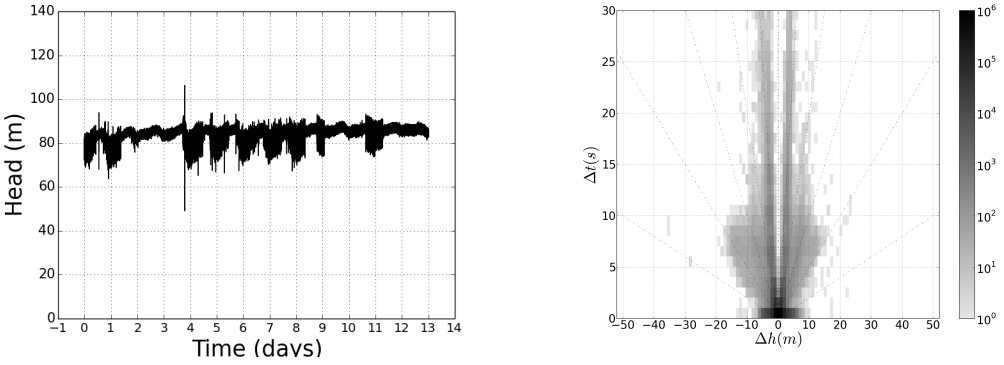
(a) **(b)**

Figure A.65: Zone 32, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



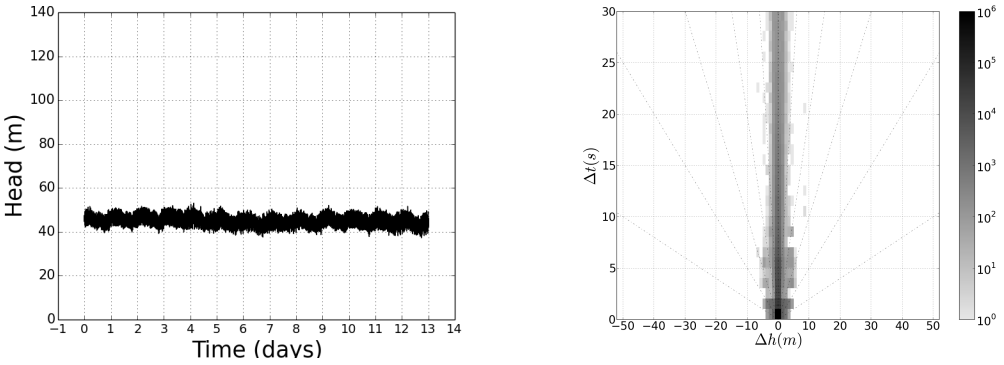
(a) **(b)**

Figure A.66: Zone 33, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



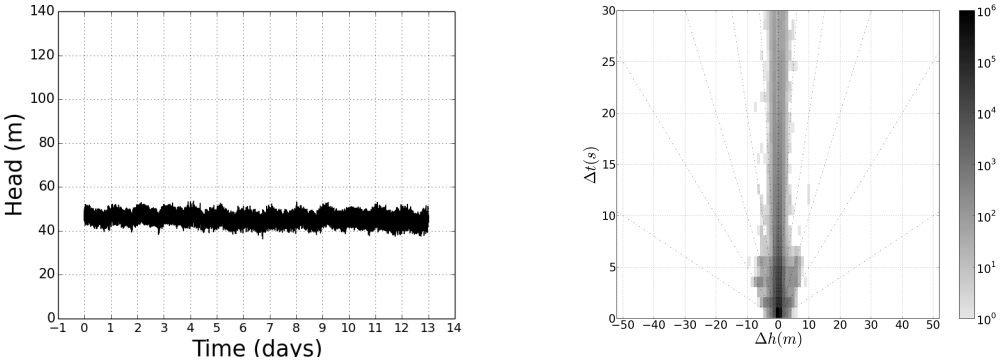
(a) **(b)**

Figure A.67: Zone 33, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



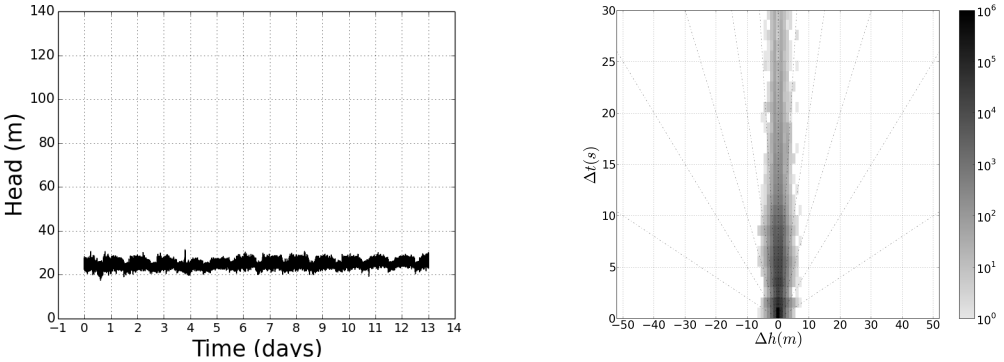
(a) **(b)**

Figure A.68: Zone 34, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



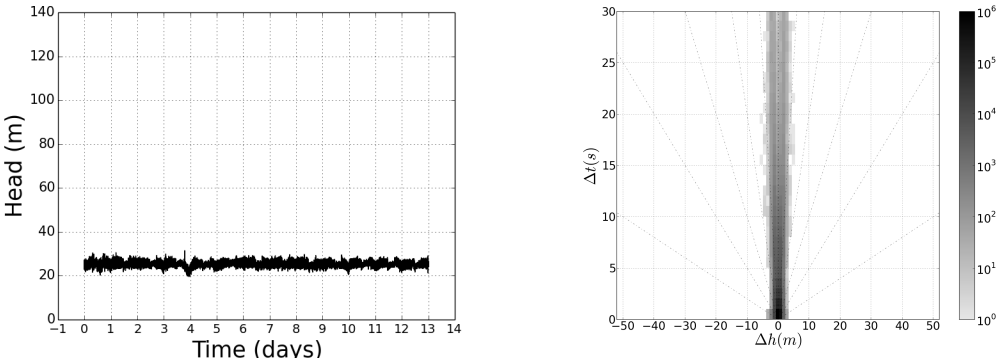
(a) **(b)**

Figure A.69: Zone 34, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



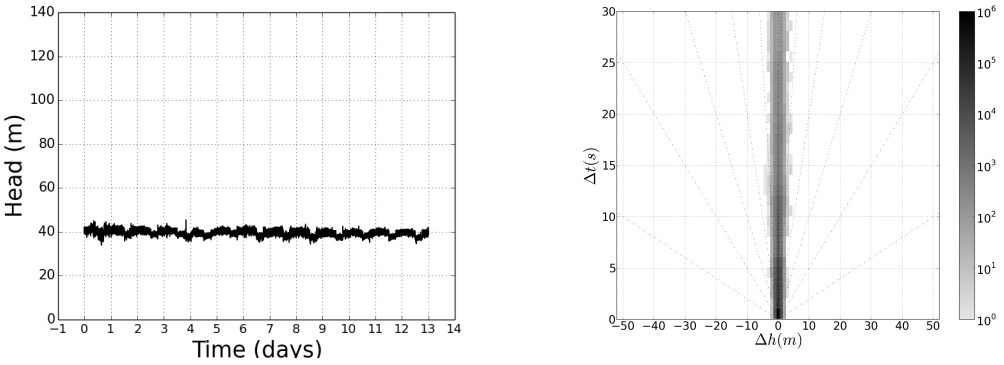
(a) **(b)**

Figure A.70: Zone 35, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



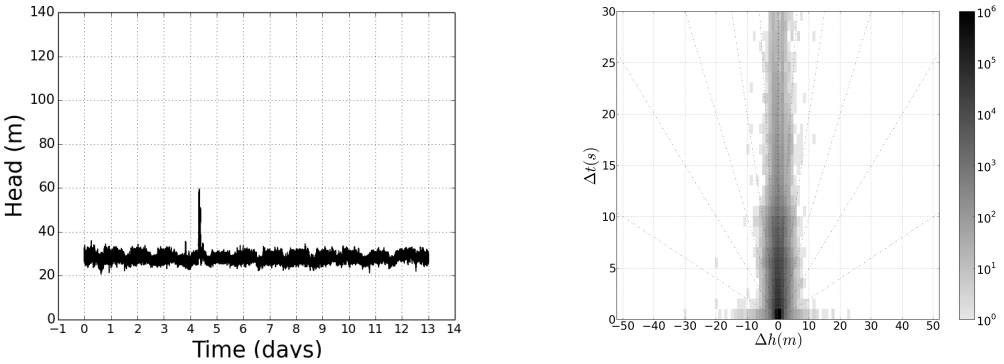
(a) **(b)**

Figure A.71: Zone 35, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



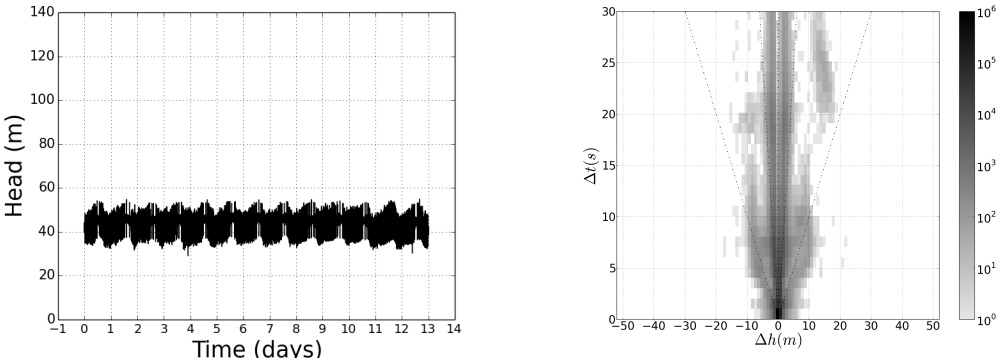
(a) **(b)**

Figure A.72: Zone 36, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



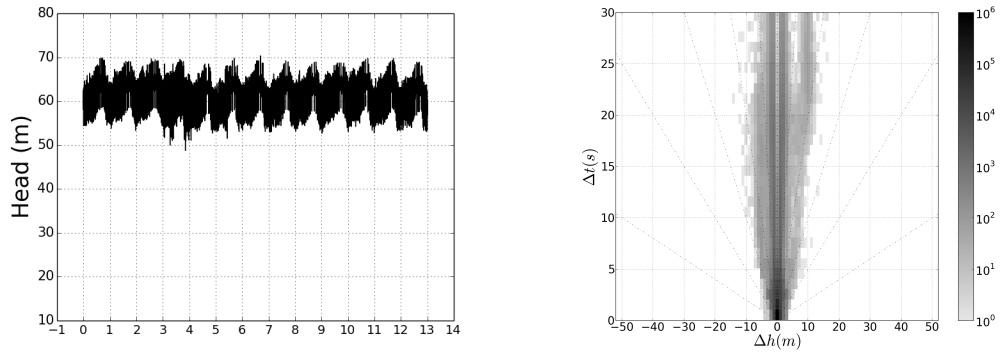
(a) **(b)**

Figure A.73: Zone 36, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



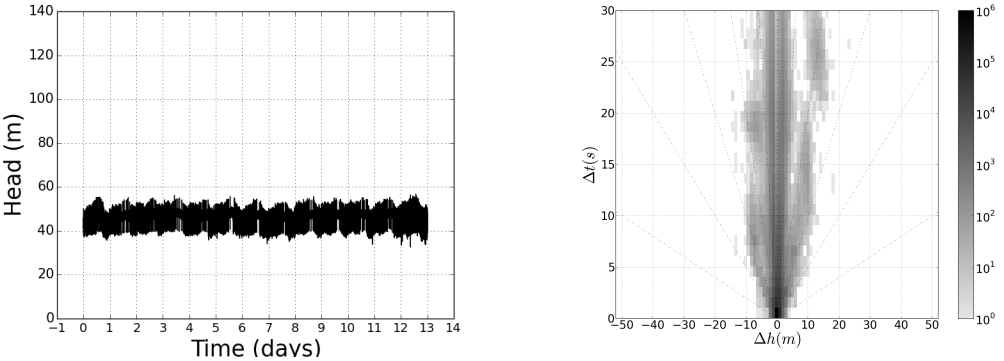
(a) **(b)**

Figure A.74: Zone 37, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



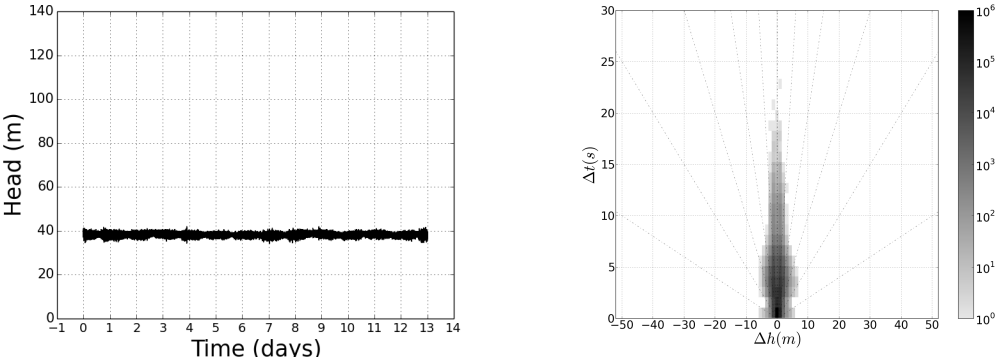
(a) **(b)**

Figure A.75: Zone 37, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



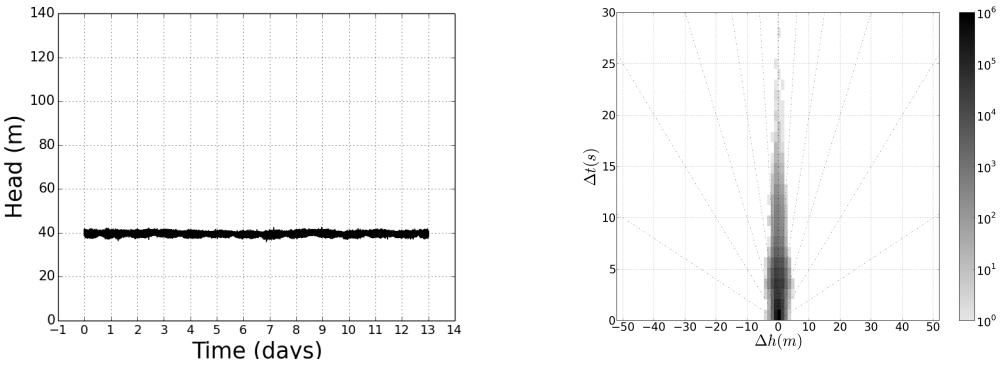
(a) **(b)**

Figure A.76: Zone 37, location 3, (a) pressure transients recorded at 100Hz and (b) its transient fingerprint.



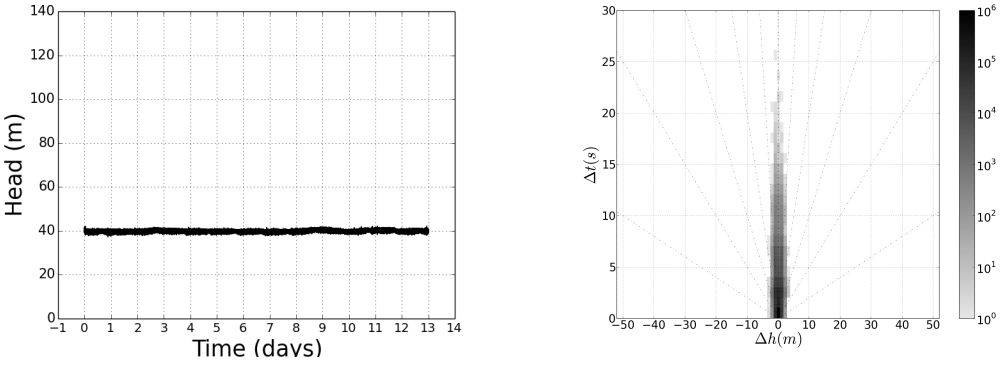
(a) **(b)**

Figure A.77: Zone 38, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



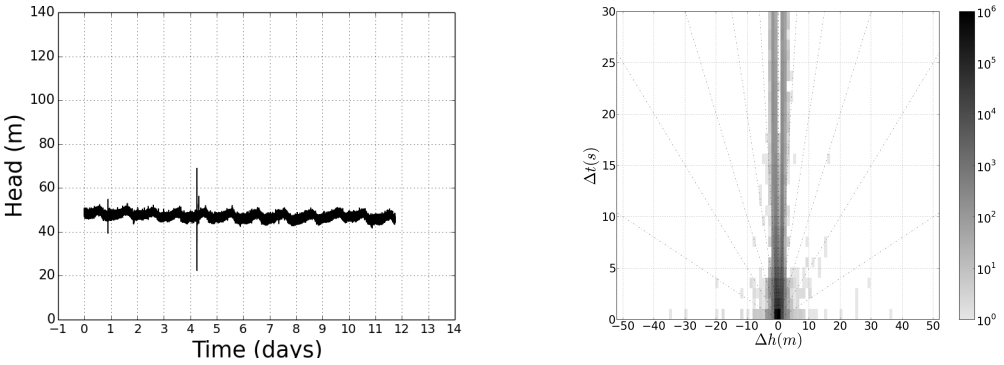
(a) **(b)**

Figure A.78: Zone 38, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



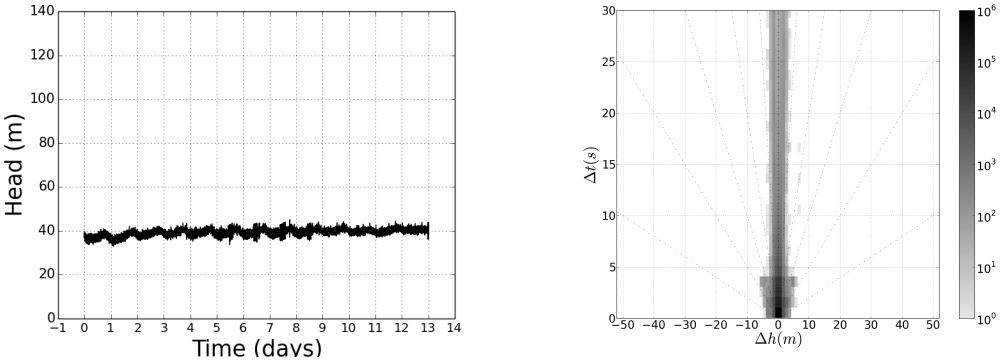
(a) **(b)**

Figure A.79: Zone 38, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



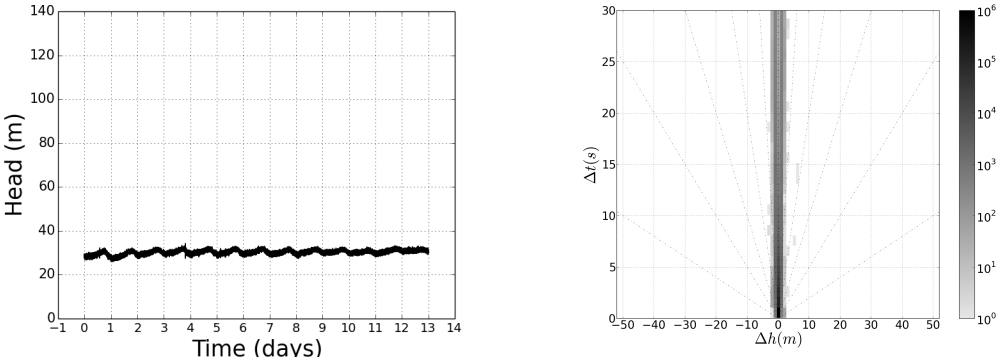
(a) **(b)**

Figure A.80: Zone 39, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



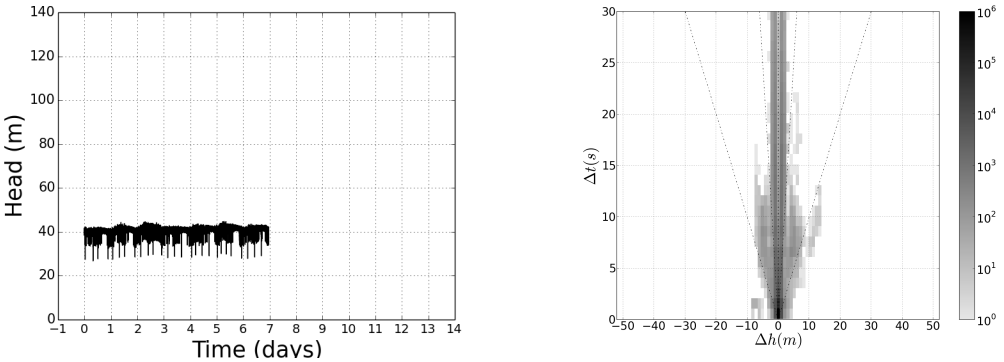
(a) **(b)**

Figure A.81: Zone 39, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



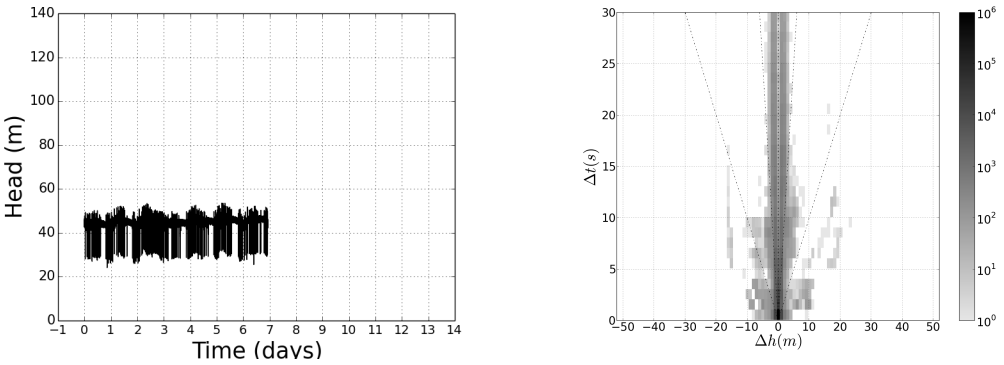
(a) **(b)**

Figure A.82: Zone 40, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



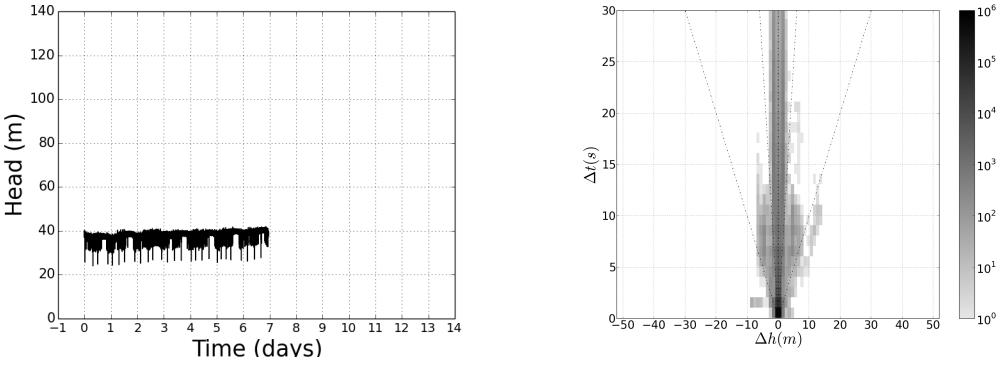
(a) **(b)**

Figure A.83: Zone 40, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



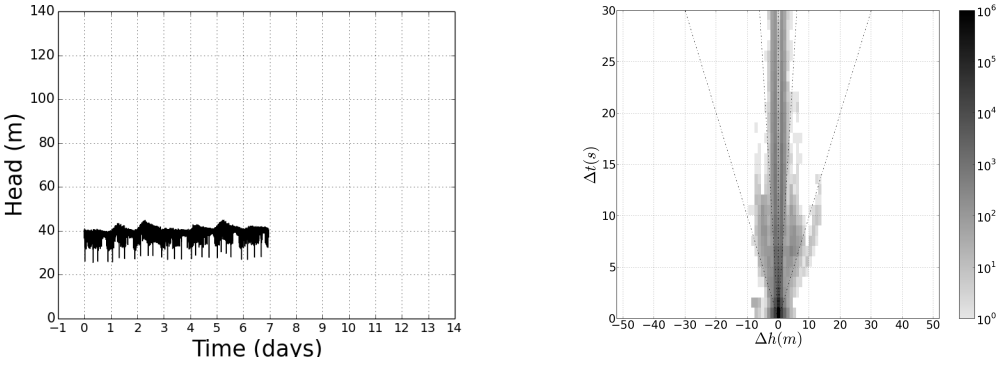
(a) **(b)**

Figure A.84: Zone 40, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



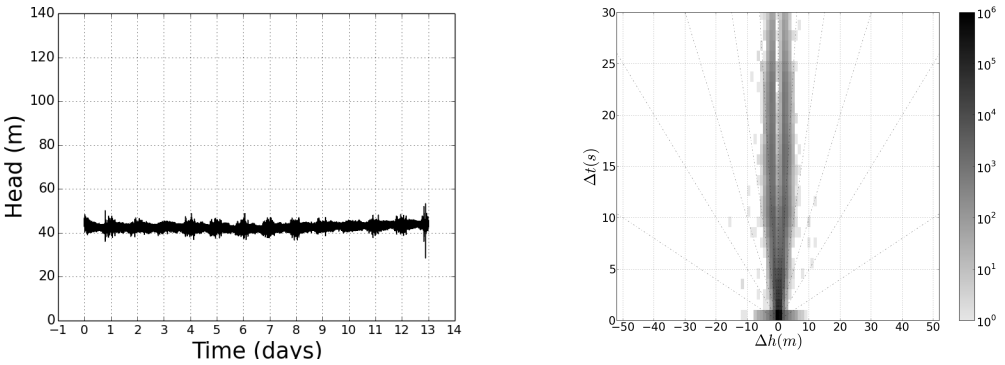
(a) **(b)**

Figure A.85: Zone 40, location 4, (a) pressure transients recorded at 100 Hz and its transient fingerprint.



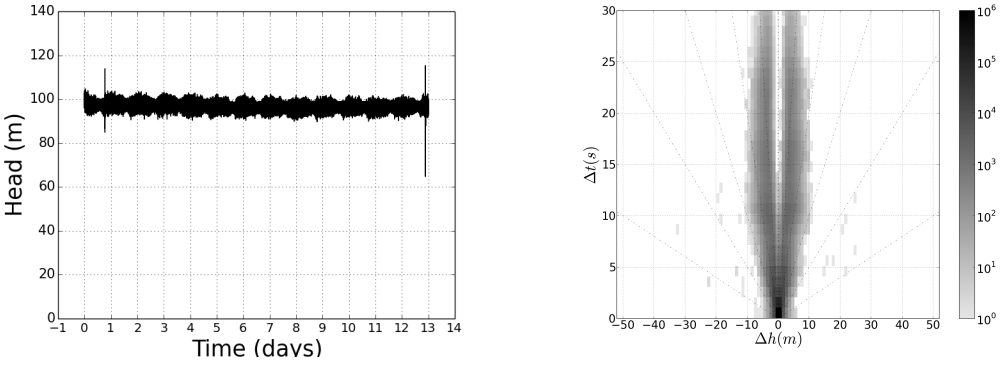
(a) **(b)**

Figure A.86: Zone 40, location 5, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



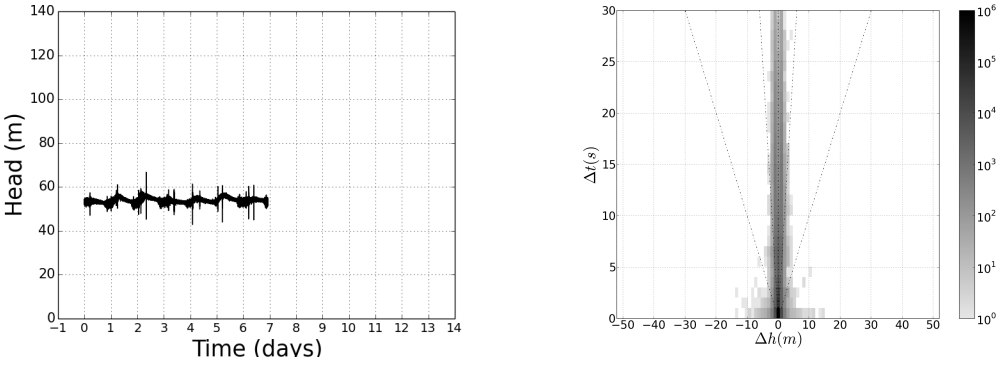
(a) **(b)**

Figure A.87: Zone 41, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



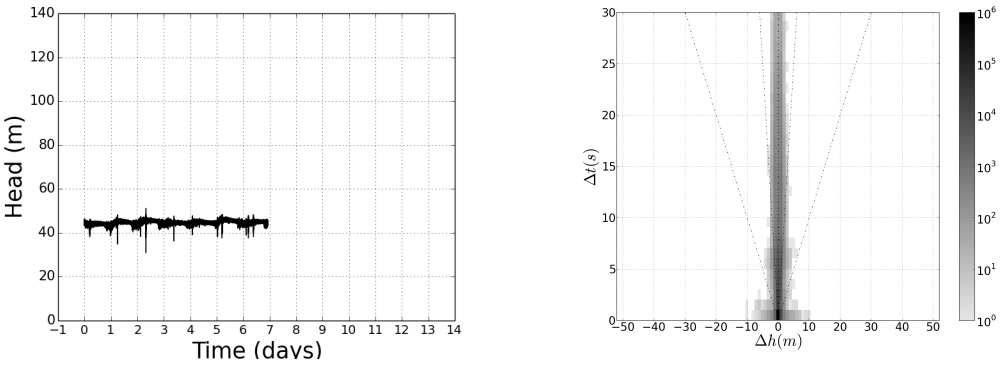
(a) **(b)**

Figure A.88: Zone 41, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



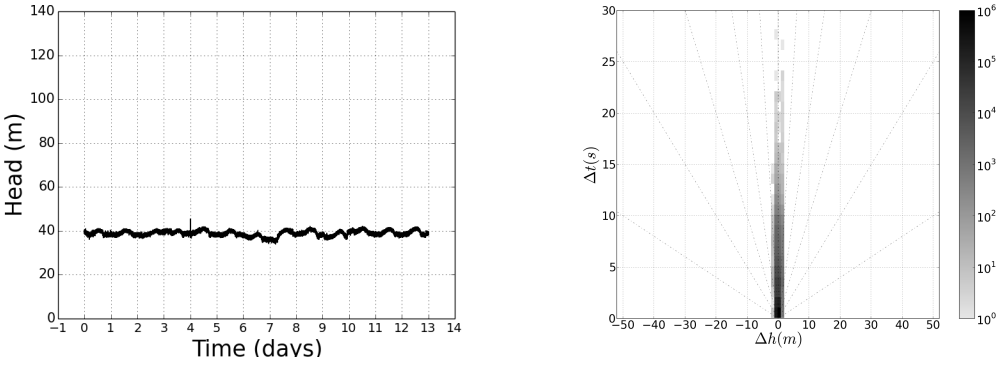
(a) **(b)**

Figure A.89: Zone 42, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



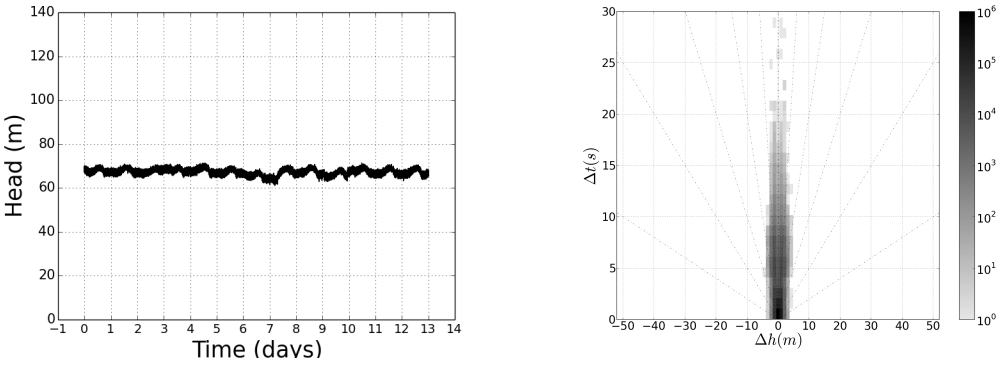
(a) **(b)**

Figure A.90: Zone 43, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



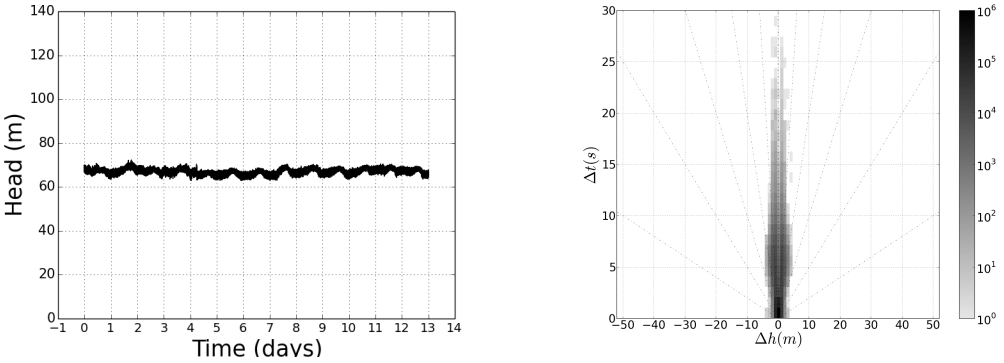
(a) **(b)**

Figure A.91: Zone 44, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



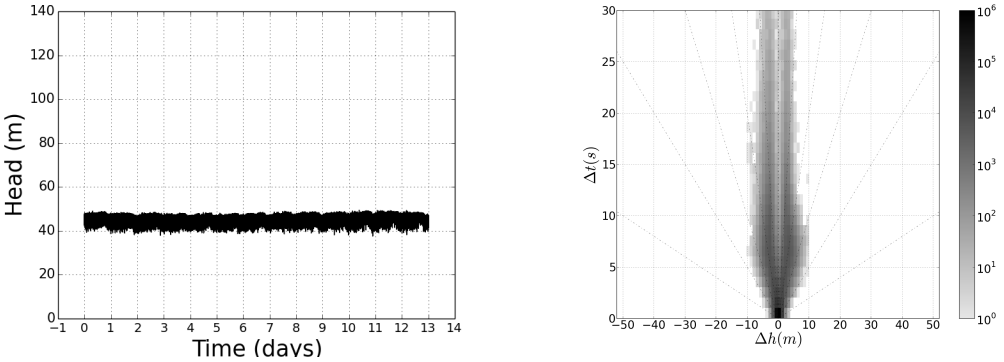
(a) **(b)**

Figure A.92: Zone 44, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



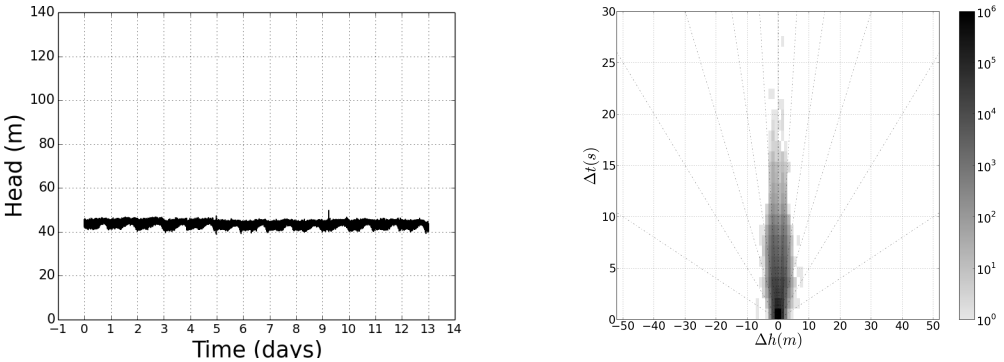
(a) **(b)**

Figure A.93: Zone 44, location 3, (a) pressure transients recorded at 100Hz and (b) its transient fingerprint.



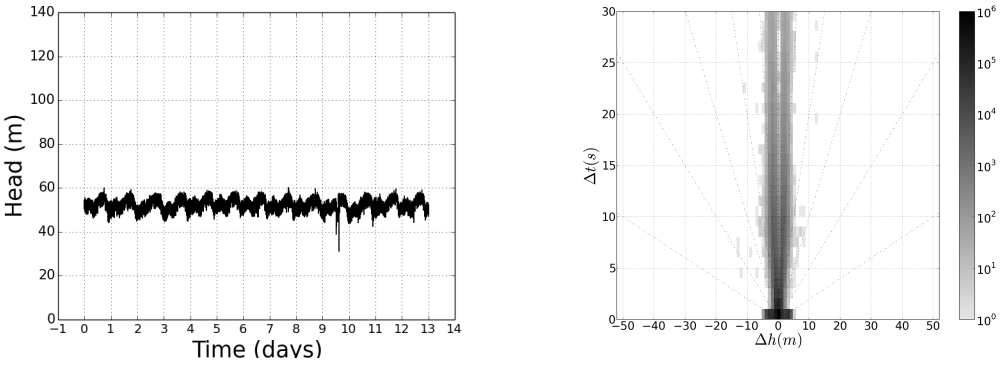
(a) **(b)**

Figure A.94: Zone 45, location 1, (a) pressure transients recorded at 100Hz and (b) its transient fingerprint.



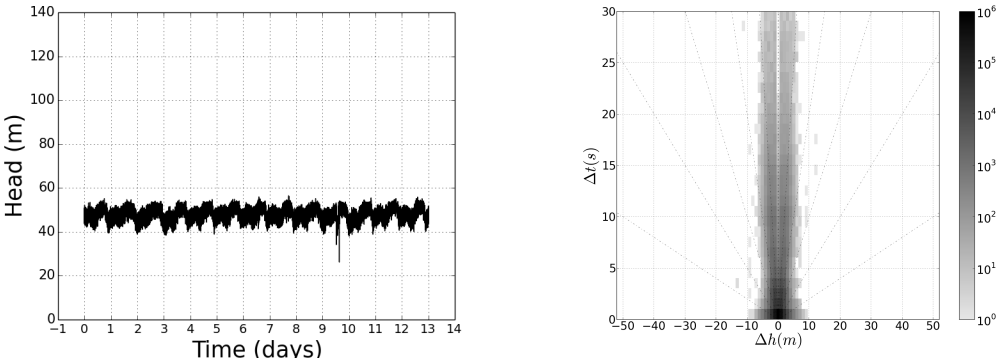
(a) **(b)**

Figure A.95: Zone 45, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



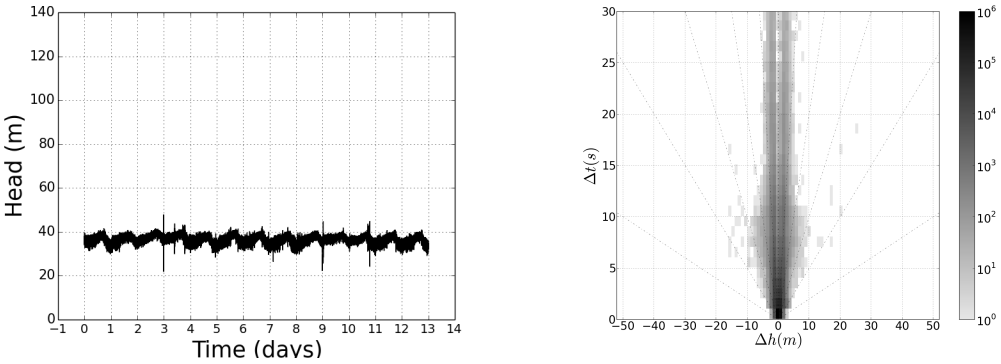
(a) **(b)**

Figure A.96: Zone 46, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



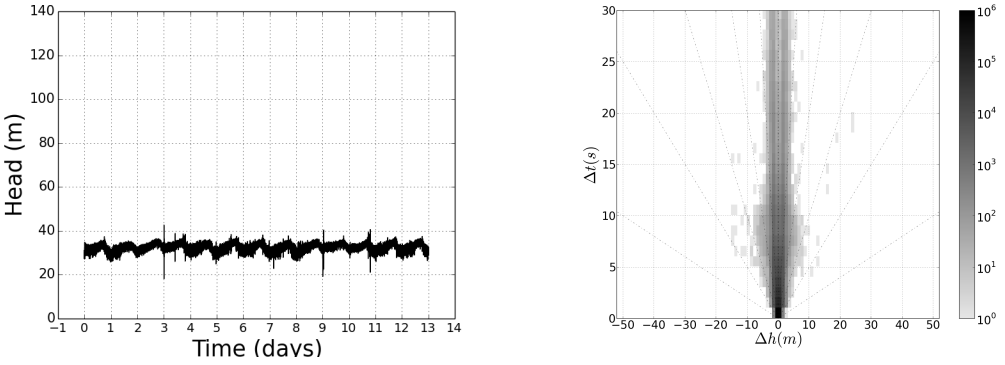
(a) **(b)**

Figure A.97: Zone 46, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



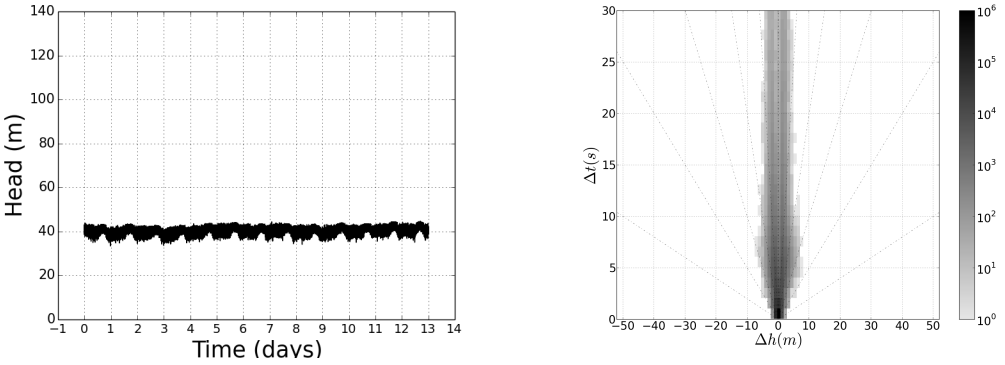
(a) **(b)**

Figure A.98: Zone 47, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



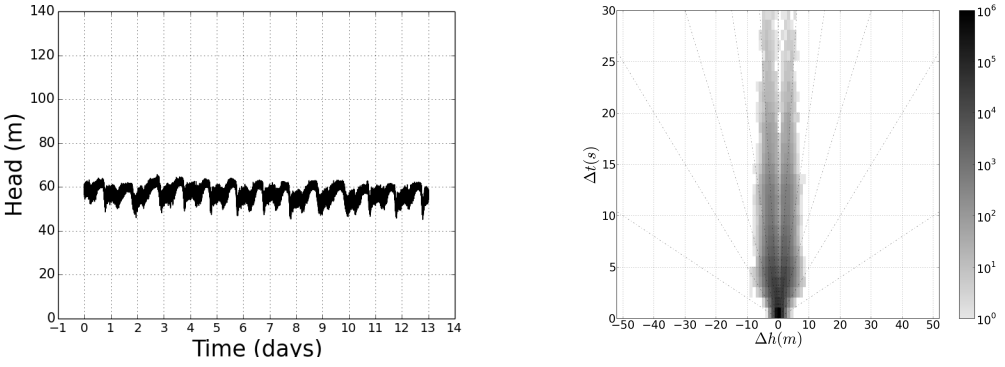
(a) **(b)**

Figure A.99: Zone 48, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



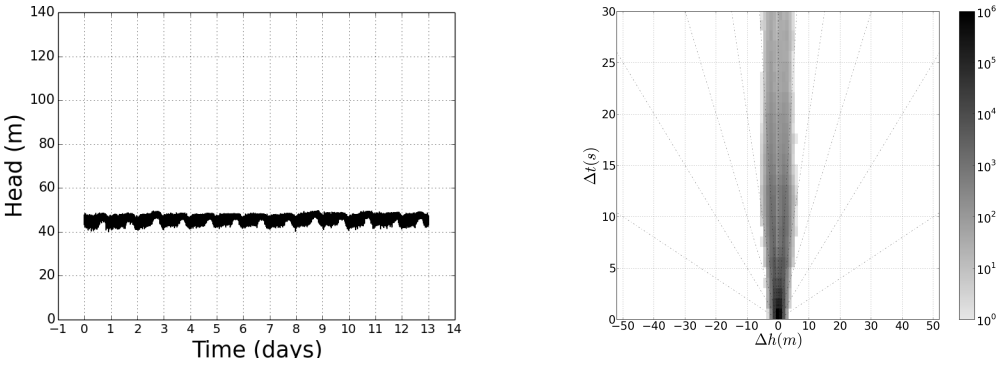
(a) **(b)**

Figure A.100: Zone 49, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



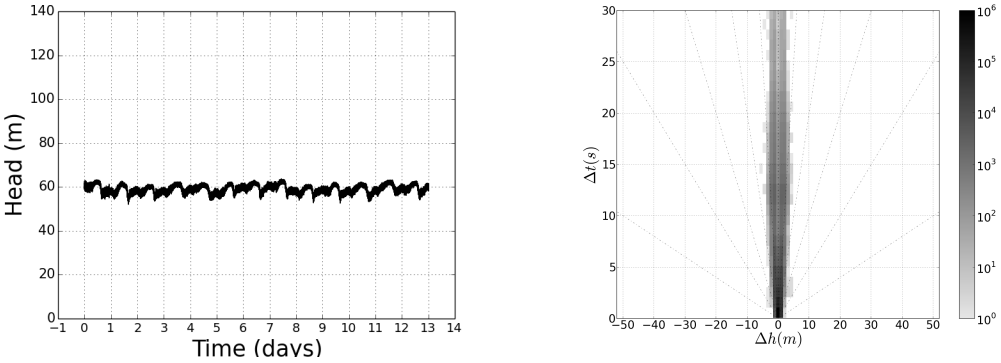
(a) **(b)**

Figure A.101: Zone 50, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



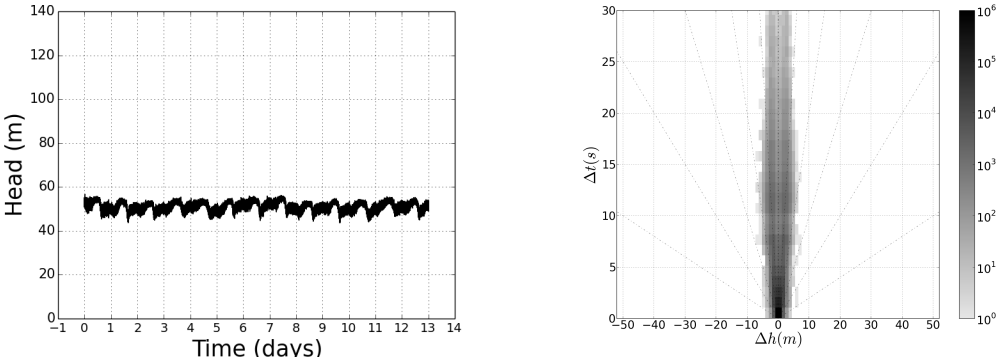
(a) **(b)**

Figure A.102: Zone 51, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



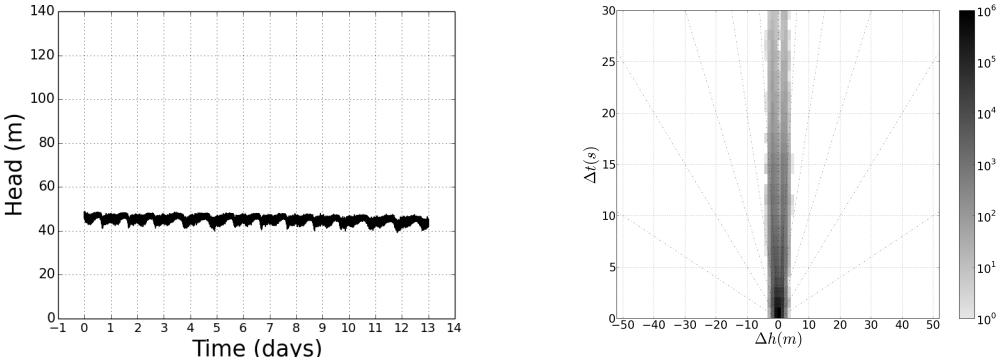
(a) **(b)**

Figure A.103: Zone 52, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



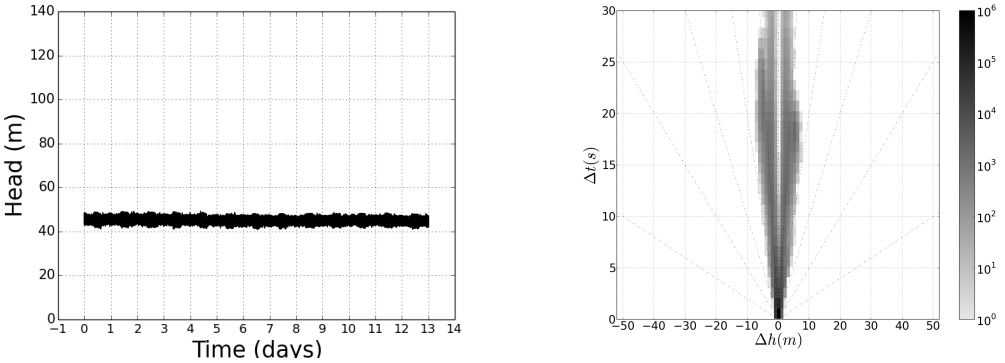
(a) **(b)**

Figure A.104: Zone 52, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



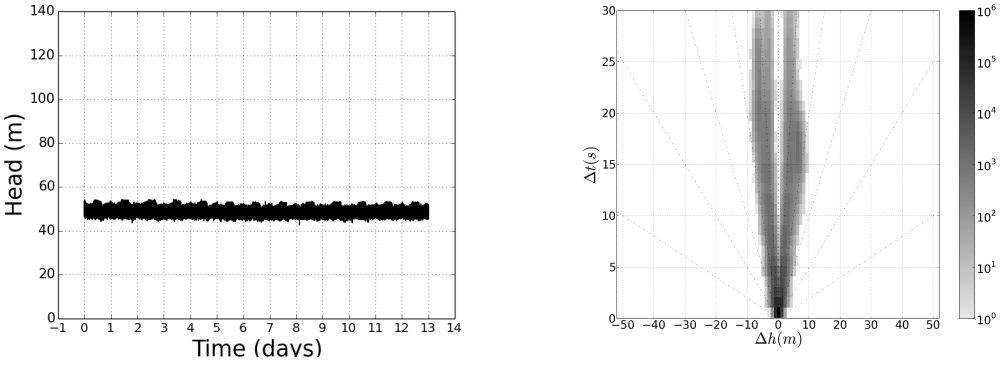
(a) **(b)**

Figure A.105: Zone 53, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



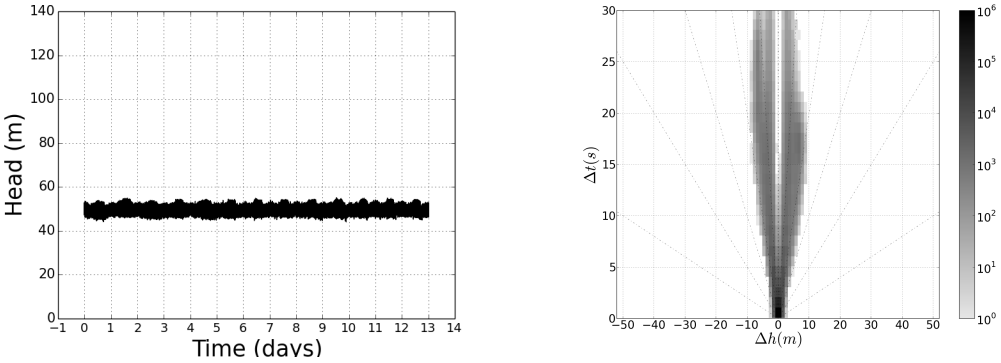
(a) **(b)**

Figure A.106: Zone 54, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



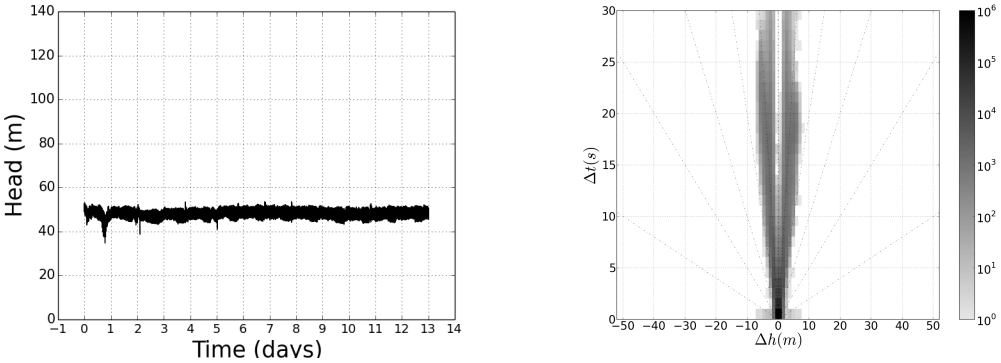
(a) **(b)**

Figure A.107: Zone 54, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



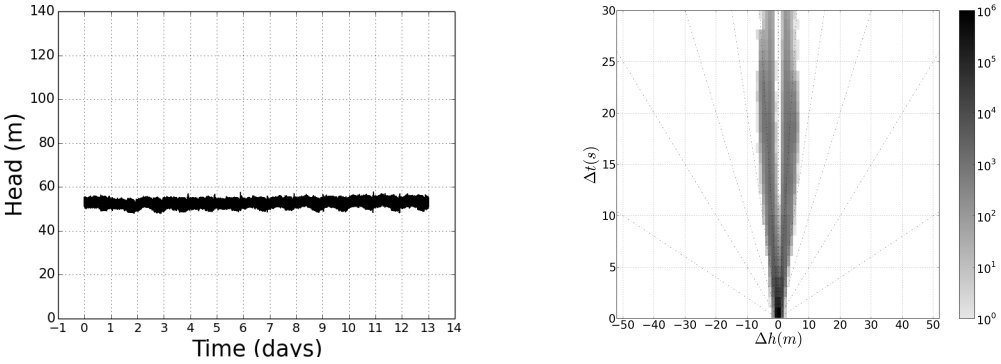
(a) **(b)**

Figure A.108: Zone 54, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



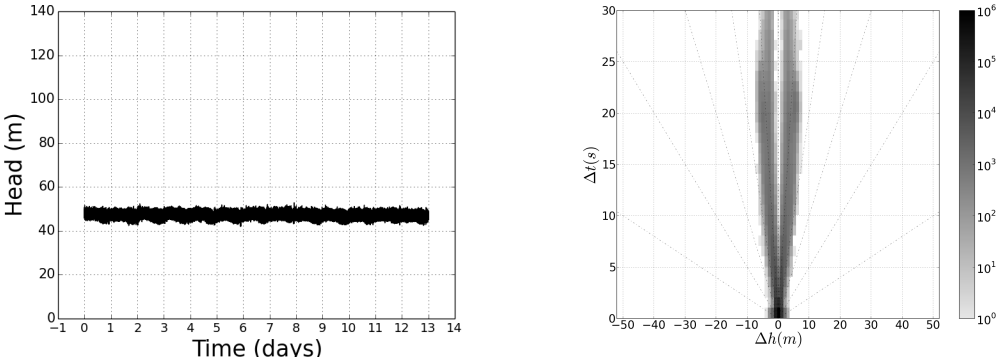
(a) **(b)**

Figure A.109: Zone 55, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



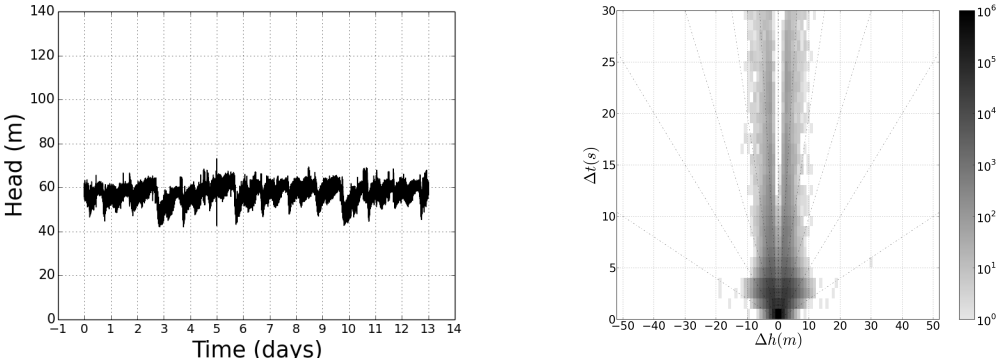
(a) **(b)**

Figure A.110: Zone 55, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



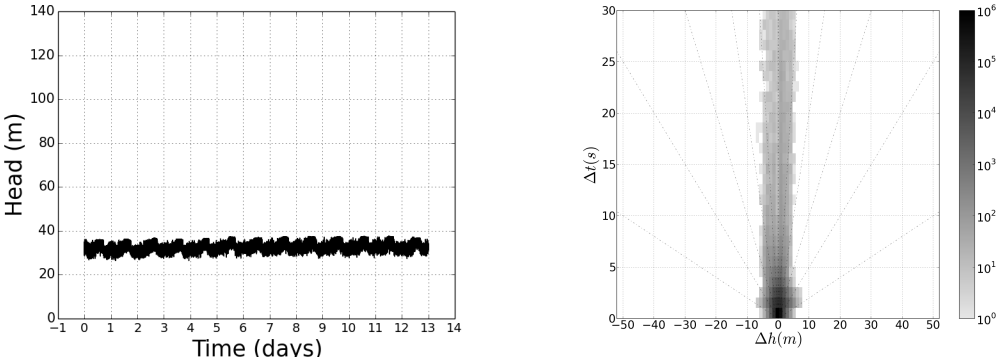
(a) **(b)**

Figure A.111: Zone 56, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



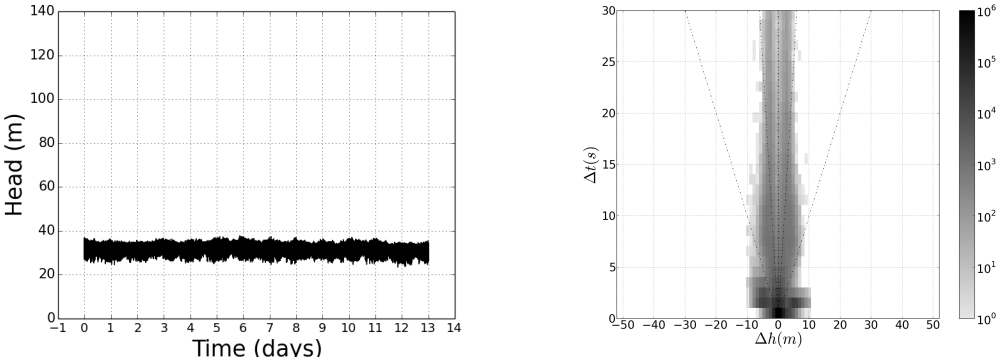
(a) **(b)**

Figure A.112: Zone 57, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



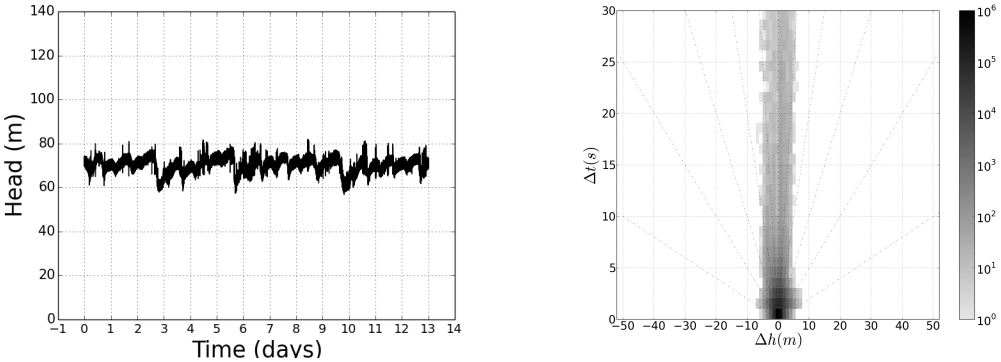
(a) **(b)**

Figure A.113: Zone 58, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



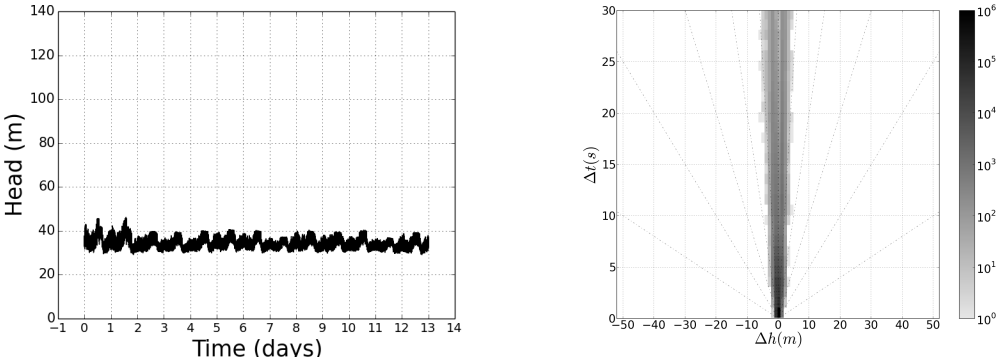
(a) **(b)**

Figure A.114: Zone 59, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



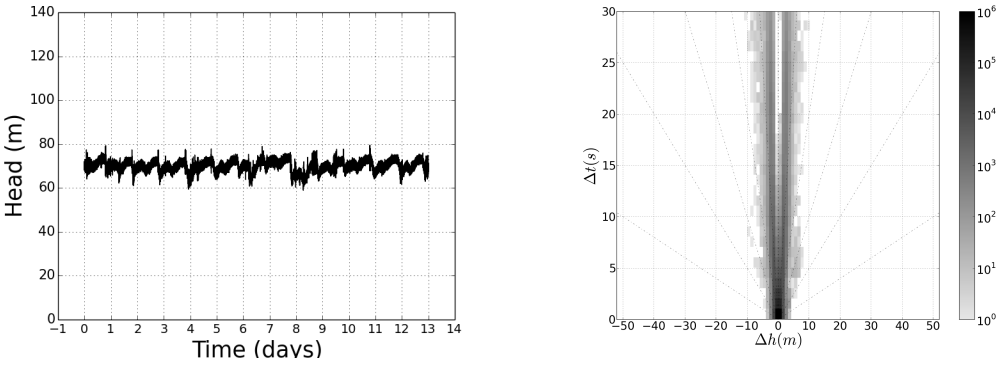
(a) **(b)**

Figure A.115: Zone 60, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



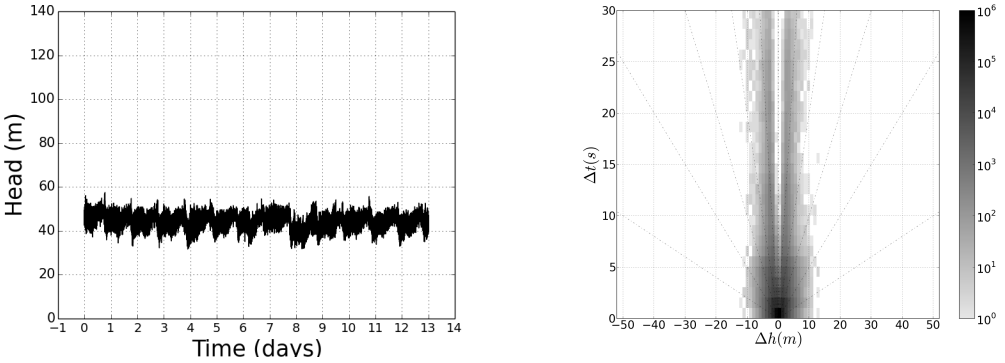
(a) **(b)**

Figure A.116: Zone 61, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



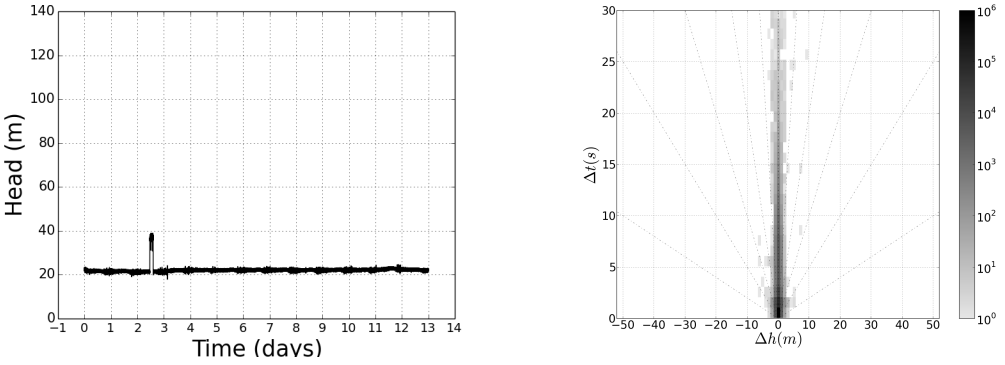
(a) **(b)**

Figure A.117: Zone 62, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



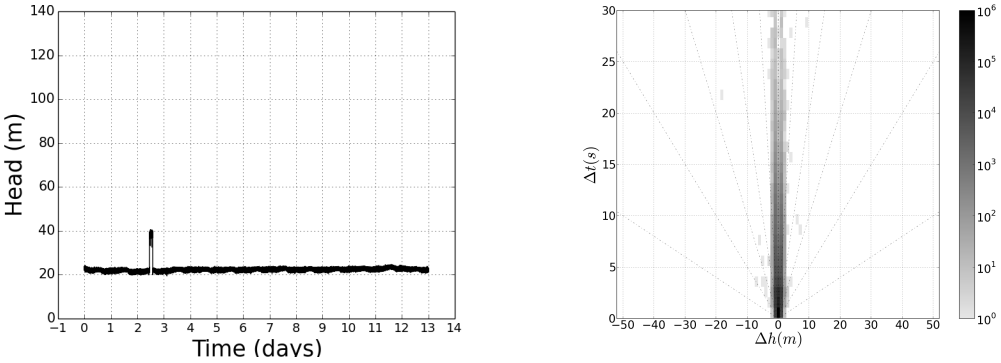
(a) **(b)**

Figure A.118: Zone 62, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



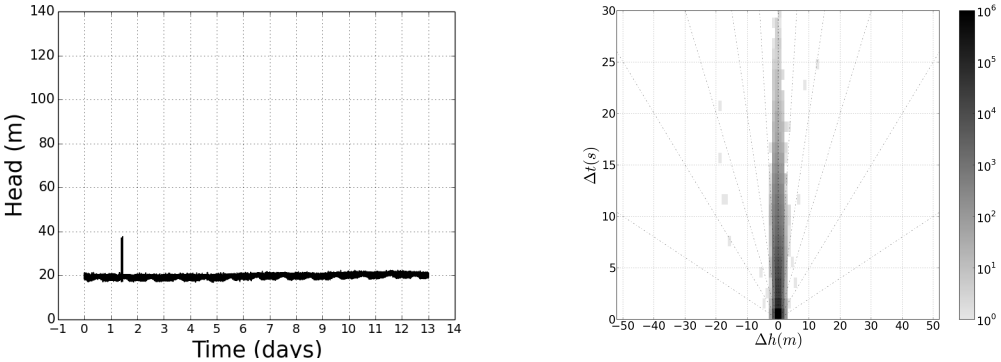
(a) **(b)**

Figure A.119: Zone 63, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



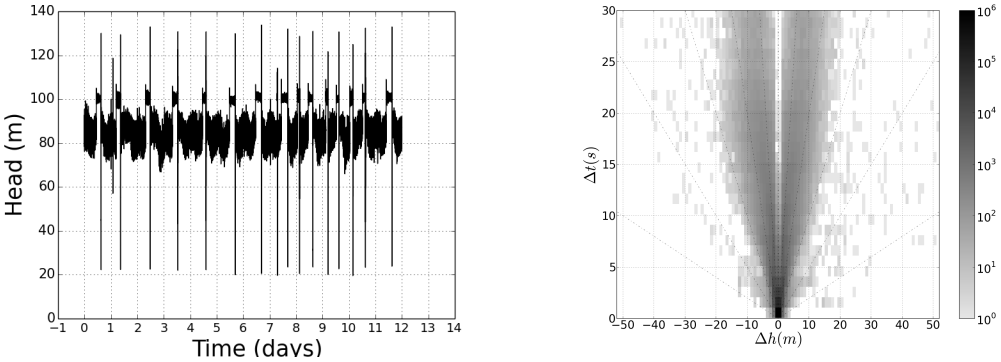
(a) **(b)**

Figure A.120: Zone 63, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



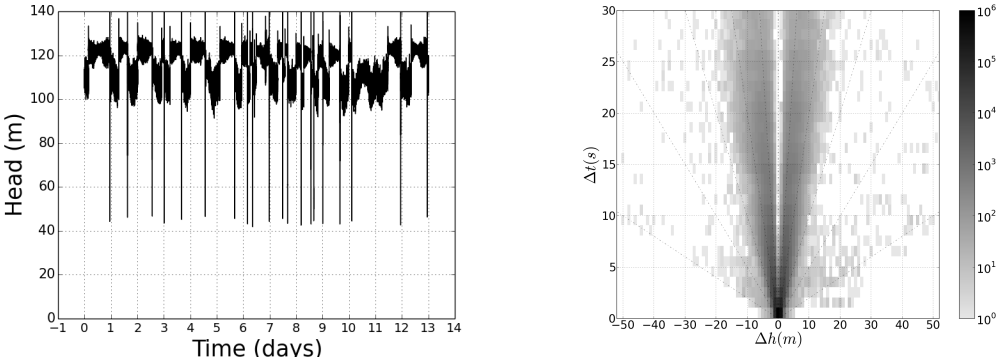
(a) **(b)**

Figure A.121: Zone 63, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



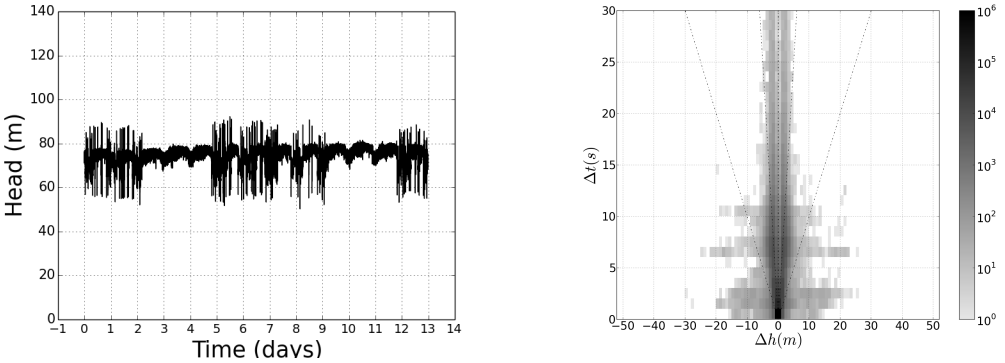
(a) **(b)**

Figure A.122: Zone 64, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



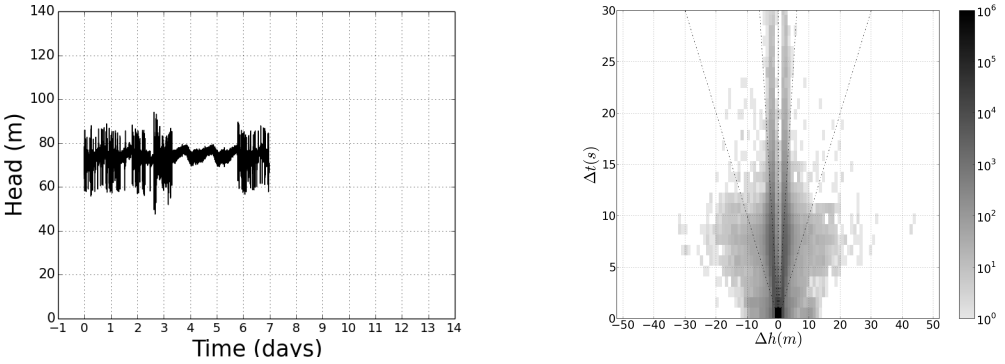
(a) **(b)**

Figure A.123: Zone 64, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



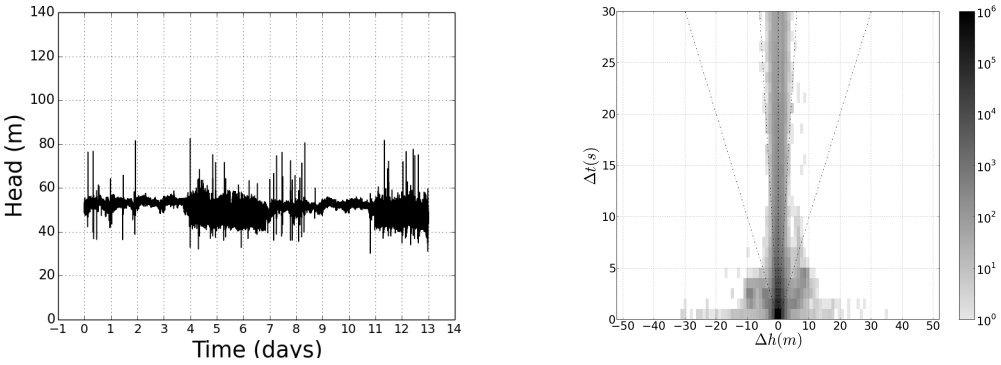
(a) **(b)**

Figure A.124: Zone 65, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



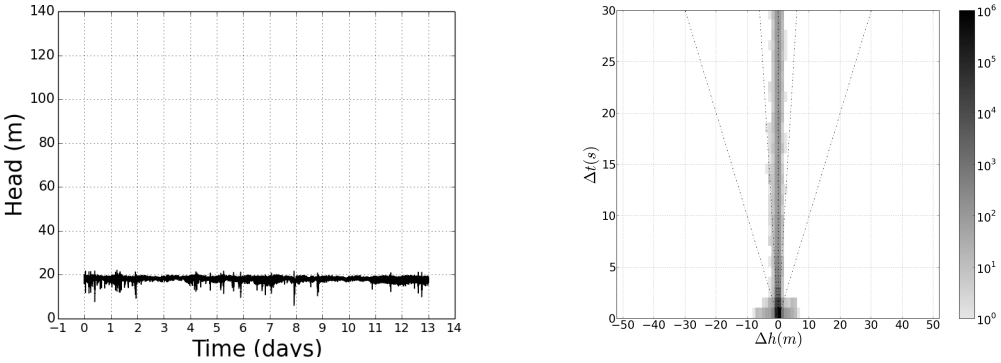
(a) **(b)**

Figure A.125: Zone 65, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



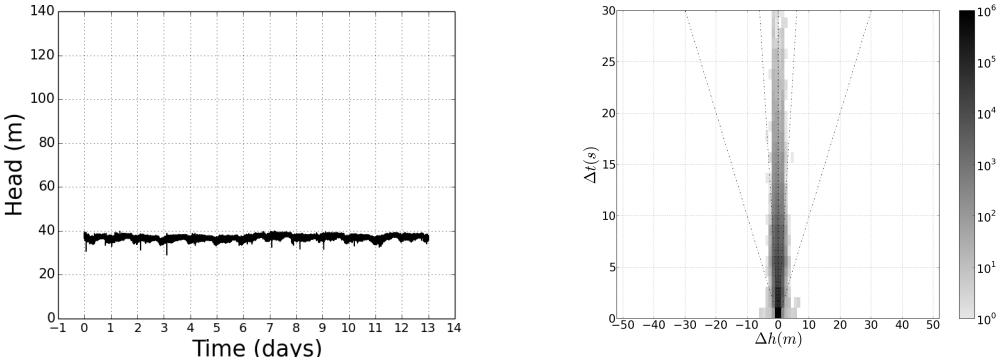
(a) **(b)**

Figure A.126: Zone 66, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



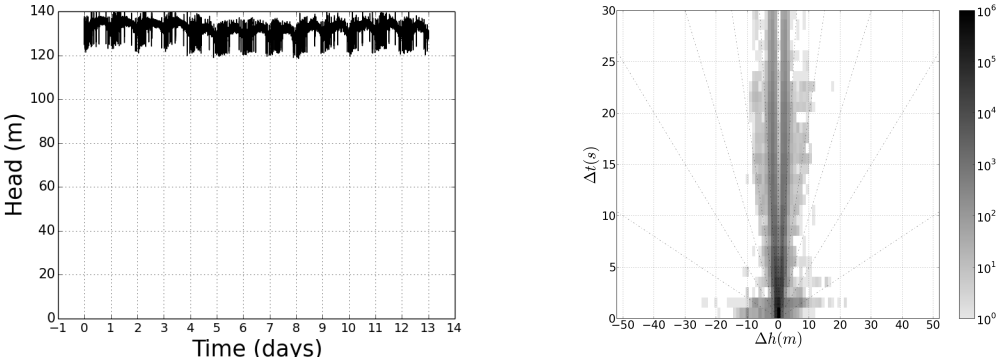
(a) **(b)**

Figure A.127: Zone 66, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



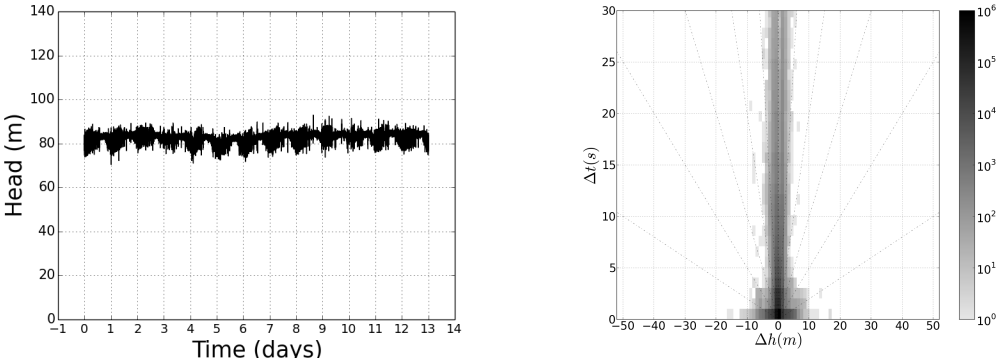
(a) **(b)**

Figure A.128: Zone 67, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



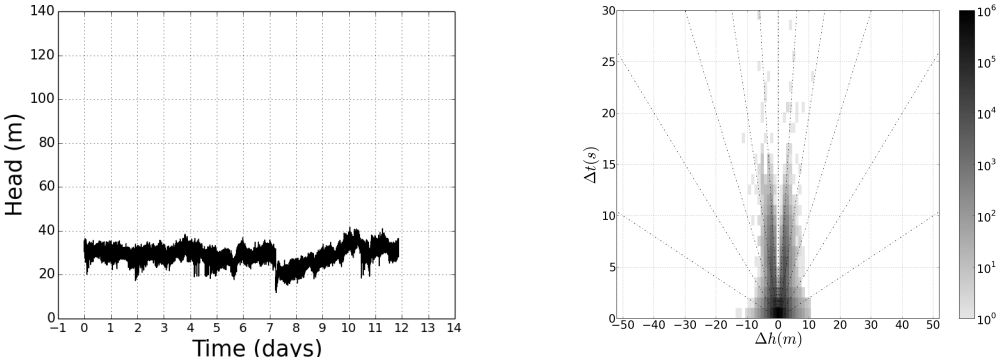
(a) **(b)**

Figure A.129: Zones 68, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



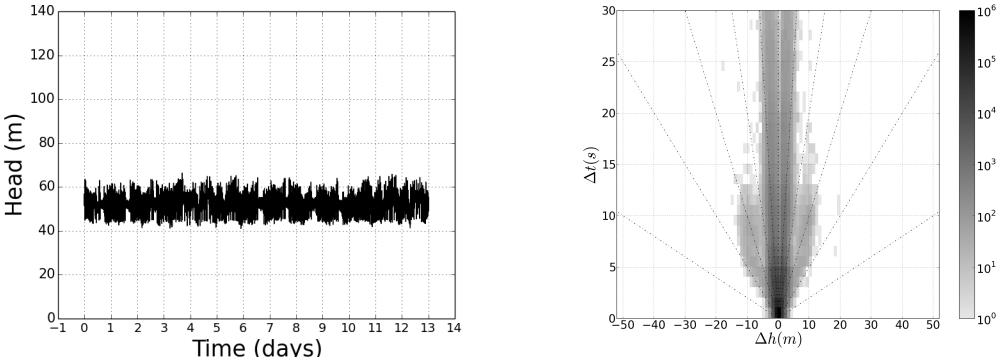
(a) **(b)**

Figure A.130: Zone 69, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



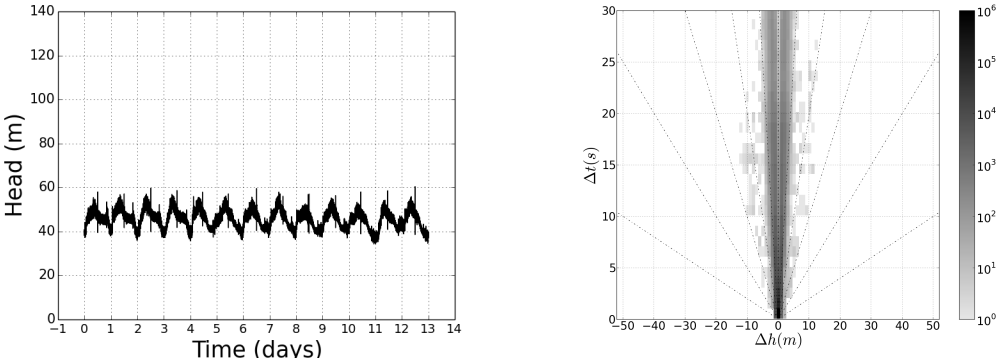
(a) **(b)**

Figure A.131: Zone 70, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



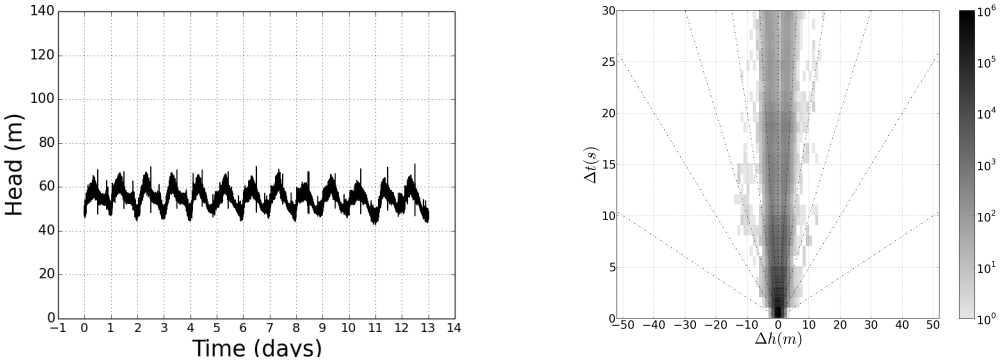
(a) **(b)**

Figure A.132: Zone 71, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



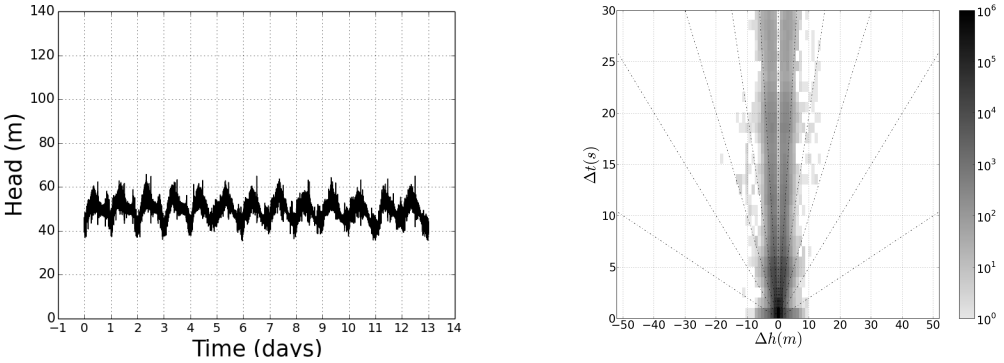
(a) **(b)**

Figure A.133: Zone 72, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



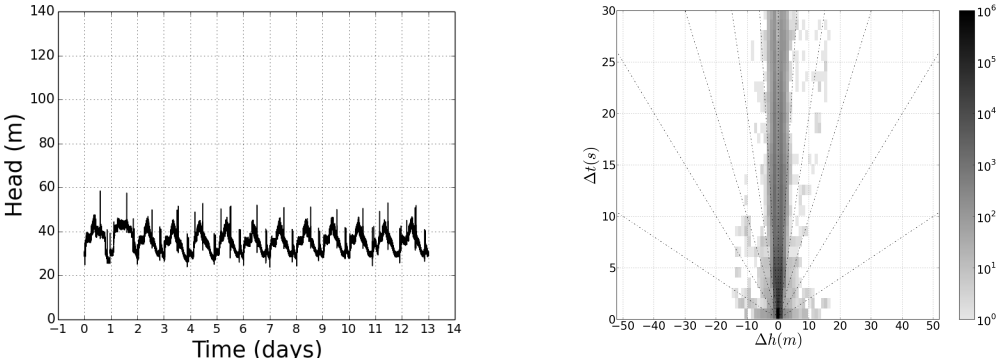
(a) **(b)**

Figure A.134: Zone 72, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



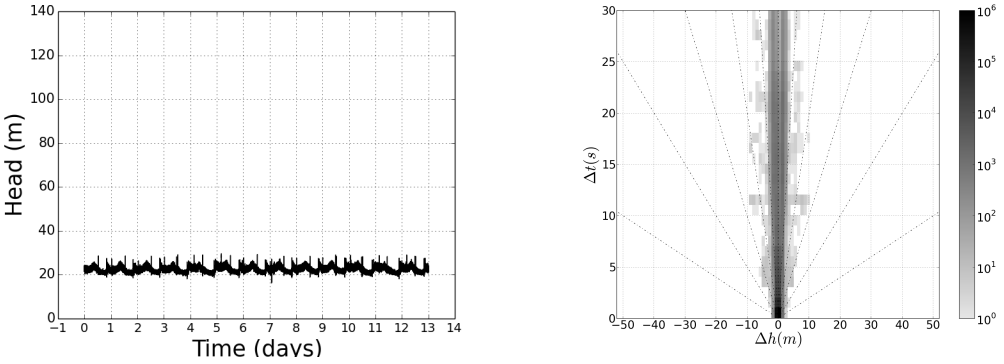
(a) **(b)**

Figure A.135: Zone 72, location 3, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



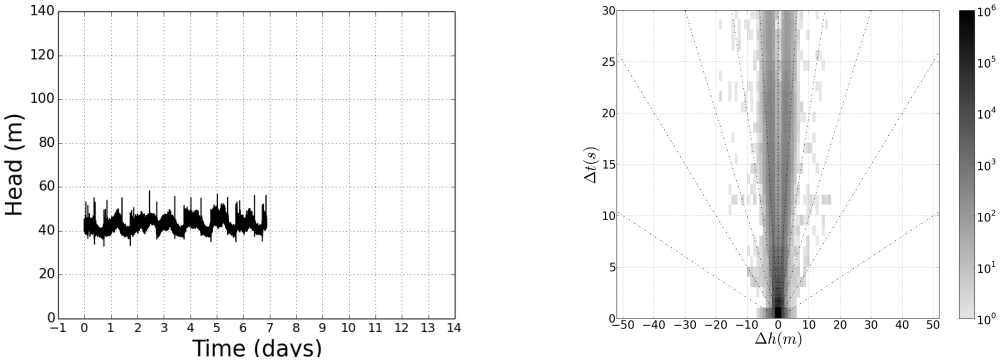
(a) **(b)**

Figure A.136: Zone 72, location 4, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



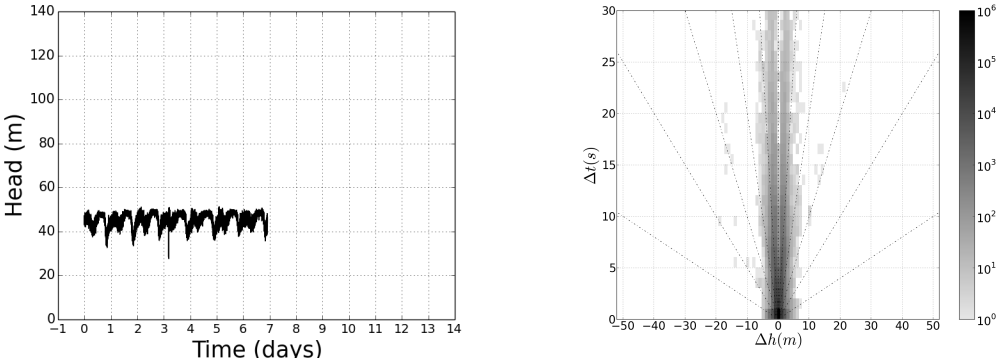
(a) **(b)**

Figure A.137: Zone 73, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



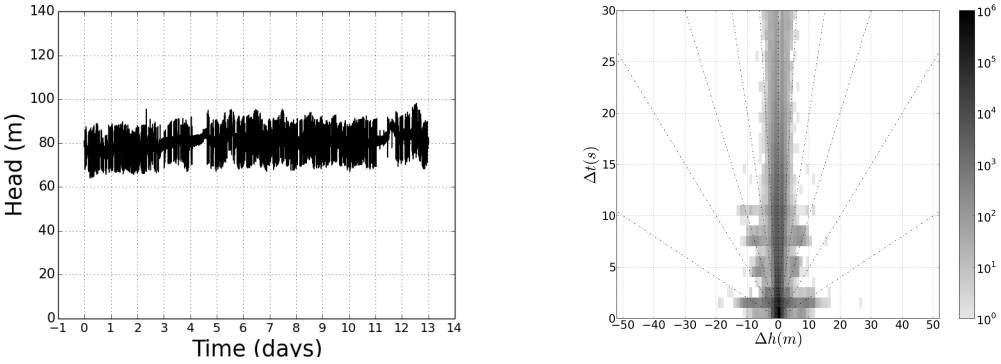
(a) **(b)**

Figure A.138: Zone 73, location 2, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



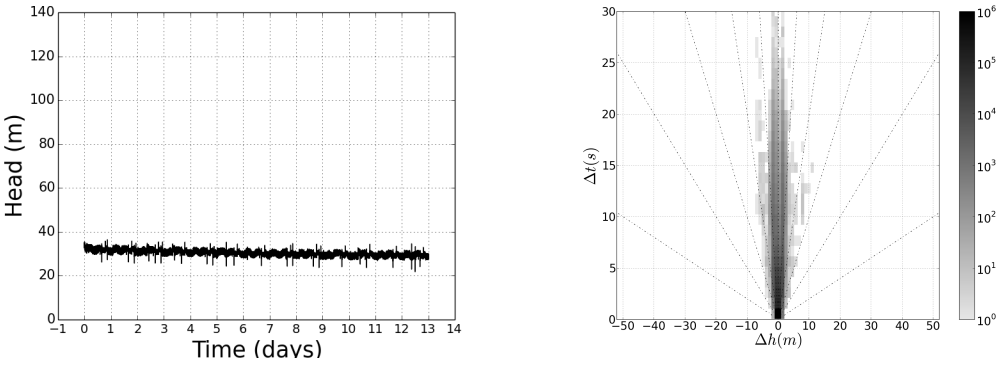
(a) **(b)**

Figure A.139: Zone 74, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



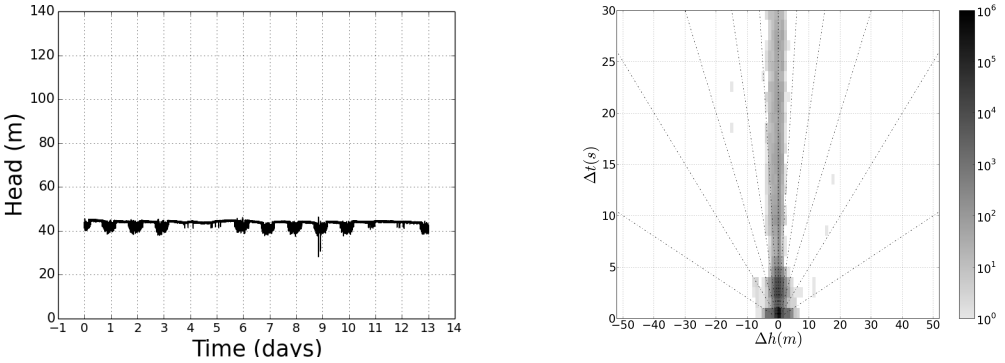
(a) **(b)**

Figure A.140: Zone 75, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



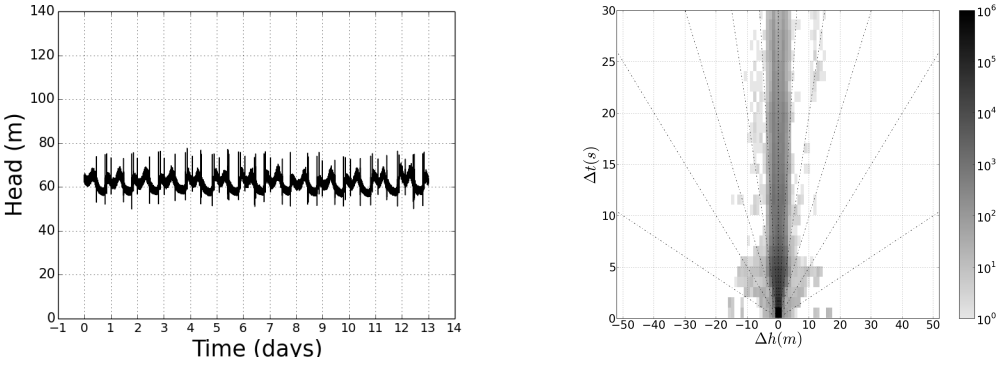
(a) **(b)**

Figure A.141: Zone 76, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



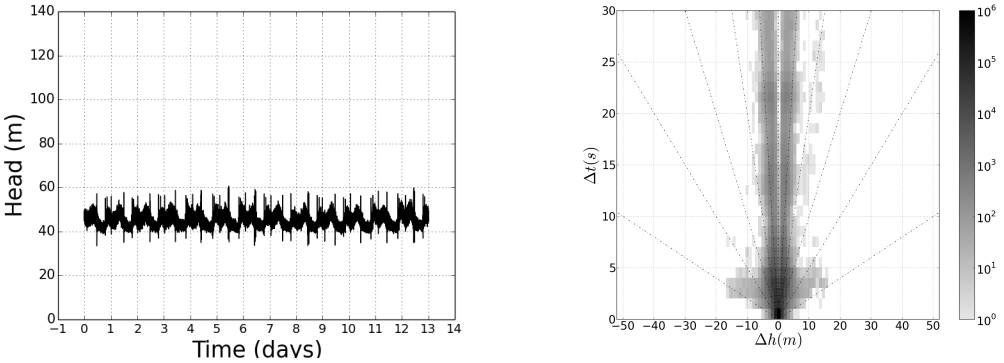
(a) **(b)**

Figure A.142: Zone 77, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



(a) **(b)**

Figure A.143: Zone 78, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.



(a) **(b)**

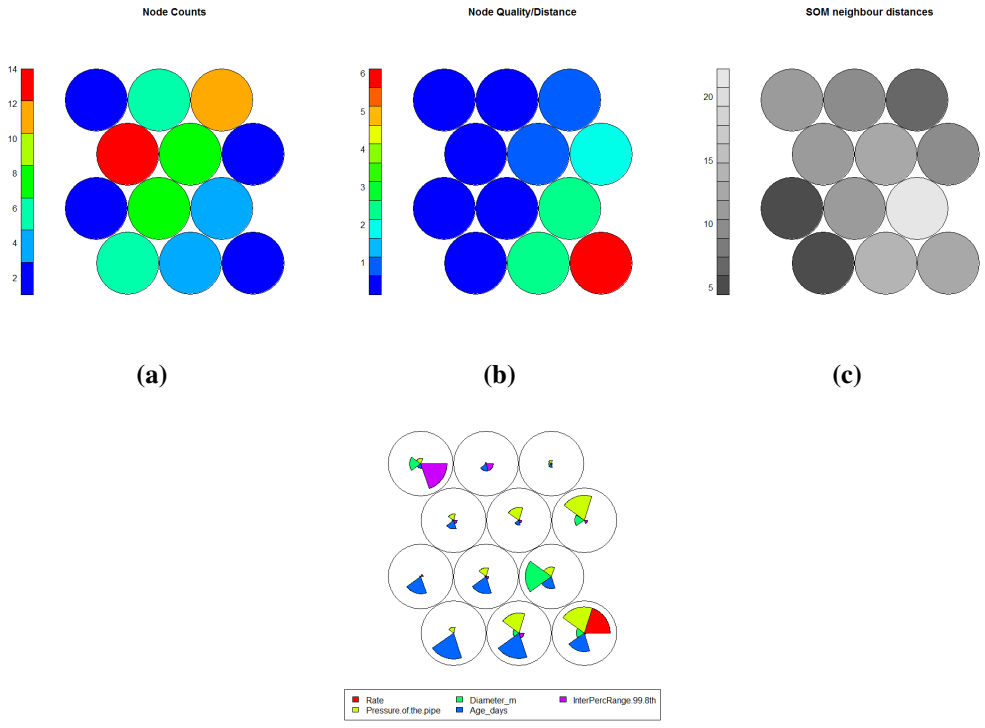
Figure A.144: Zone 79, location 1, (a) pressure transients recorded at 100 Hz and (b) its transient fingerprint.

Appendix B

SOMs

B.1 SOMs rate of change of pressure

Figures B.1 and Figures B.2 show preliminary SOMs variables assessment for the space size using the the 99.8th interpercentile range.



(a) **(b)** **(c)**

(d)

Figure B.1: Preliminary pipe-measured SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

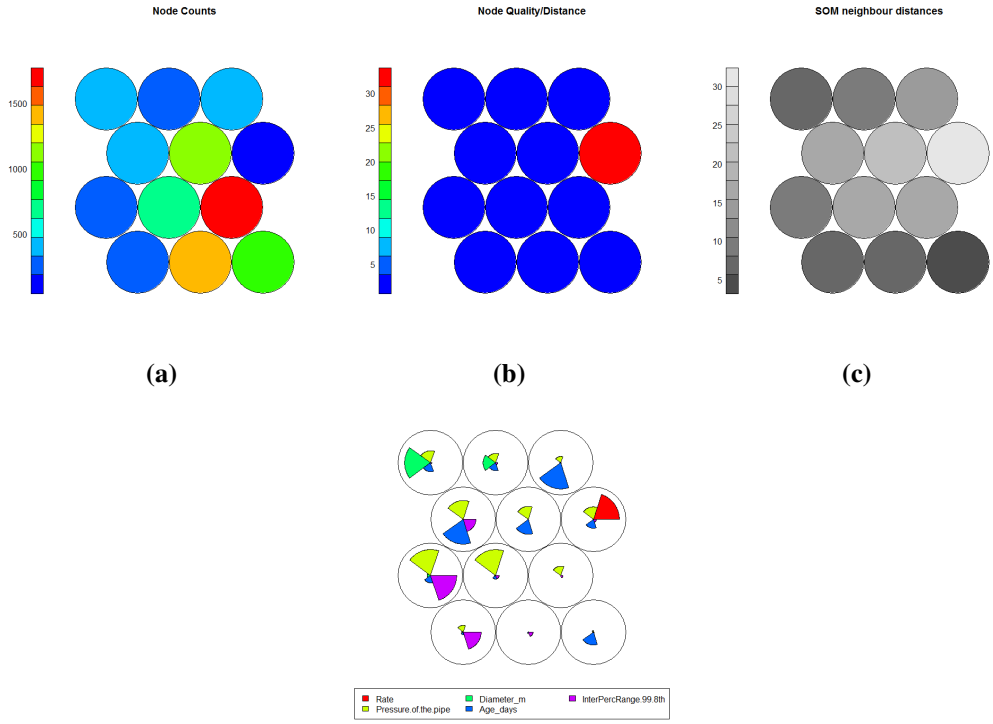


Figure B.2: Preliminary zone-level SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

Figure B.3 and Figure B.4 present the 99.8th interpercentile range for the pipe-measured and the zone-level data respectively. There is also no apparent variability and correlation with pipe repair rates. Therefore, likewise the 98th interpercentile range, this variable is rejected.

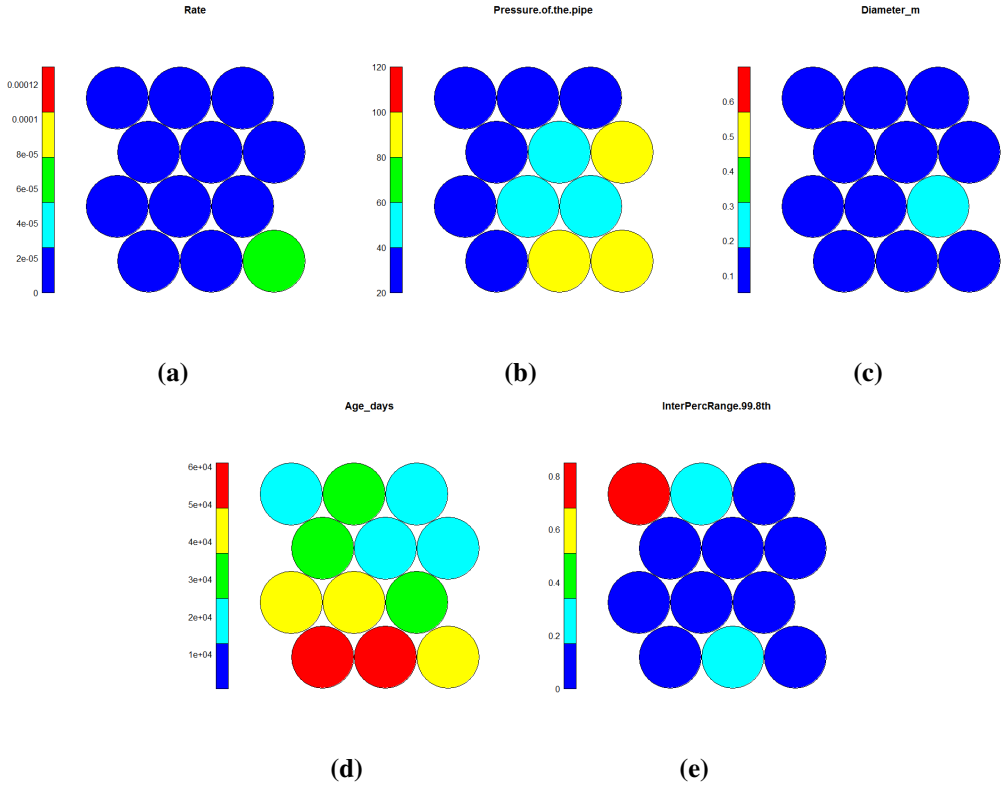


Figure B.3: Pipe measured SOMs output (a) pipe repair rate, (b) hydraulic static pressure, (c) diameter, (d) age, (e) the 99.8th interpercentile range of the rate of change of pressure.

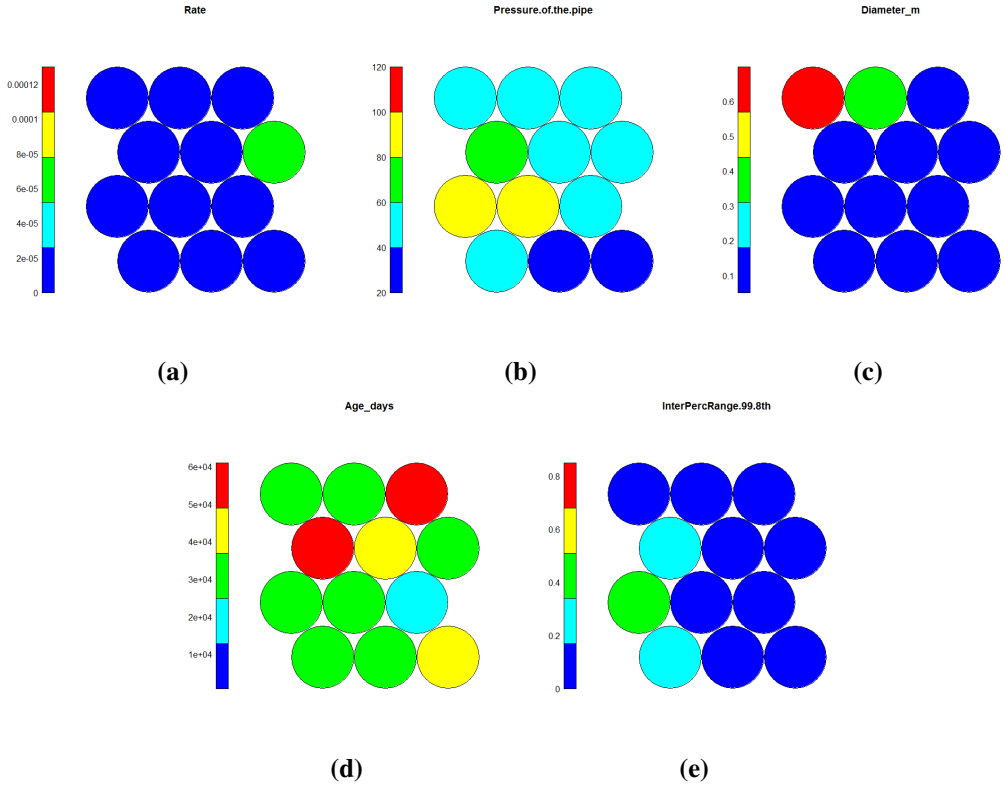
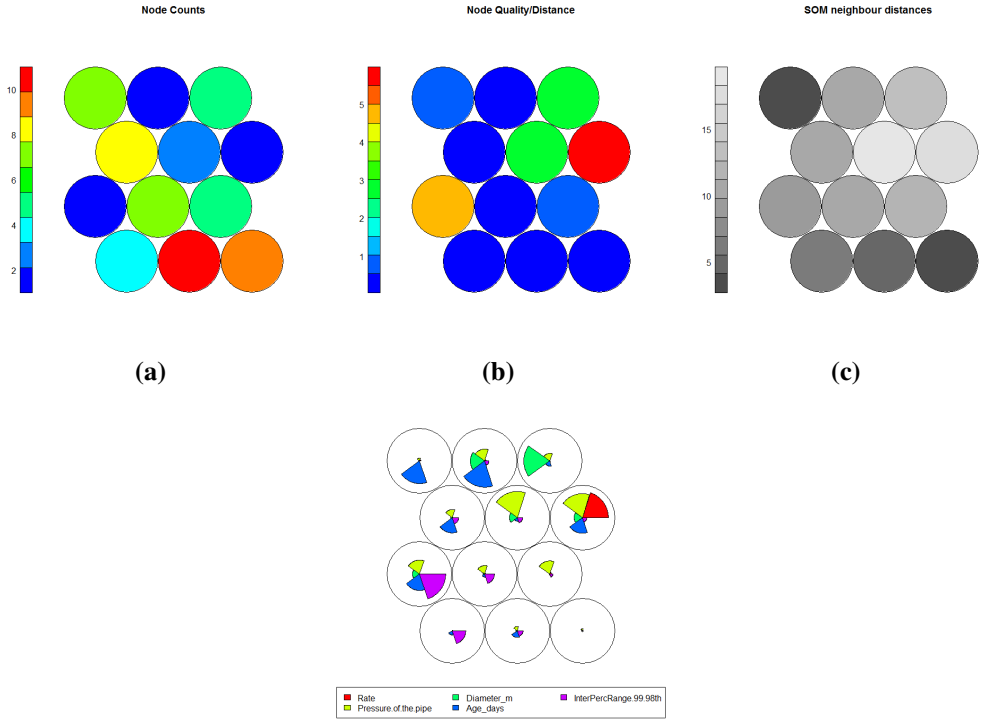


Figure B.4: Zone level pipe SOMs output (a) pipe repair rate, (b) hydraulic static pressure, (c) diameter, (d) age, (e) the 99.8th interpercentile range of the rate of change of pressure.

Figures B.5 and Figures B.6 show preliminary SOMs variables assessment for the the 99.98th interpercentile range.



(a) **(b)** **(c)**

(d)

Figure B.5: Preliminary pipe-measured SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

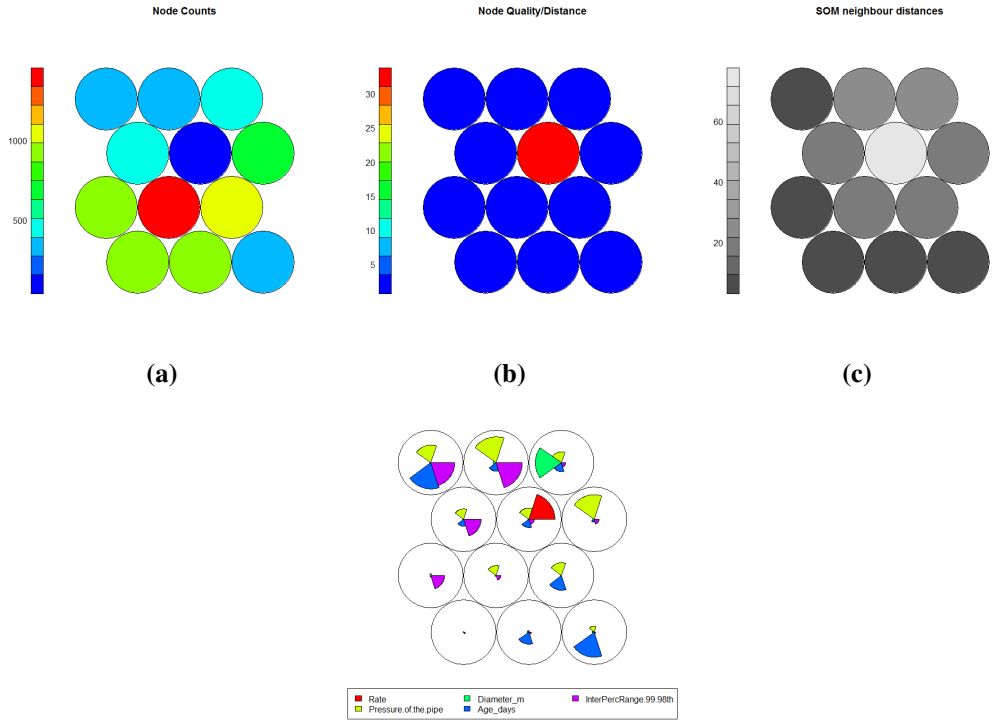


Figure B.6: Preliminary zone-level SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

Figure B.7 and Figure B.8 present the 99.98th interpercentile range for the pipe-measured and the zone-level datasets respectively. The variability is observed and the apparent correlation with pipe repairs.

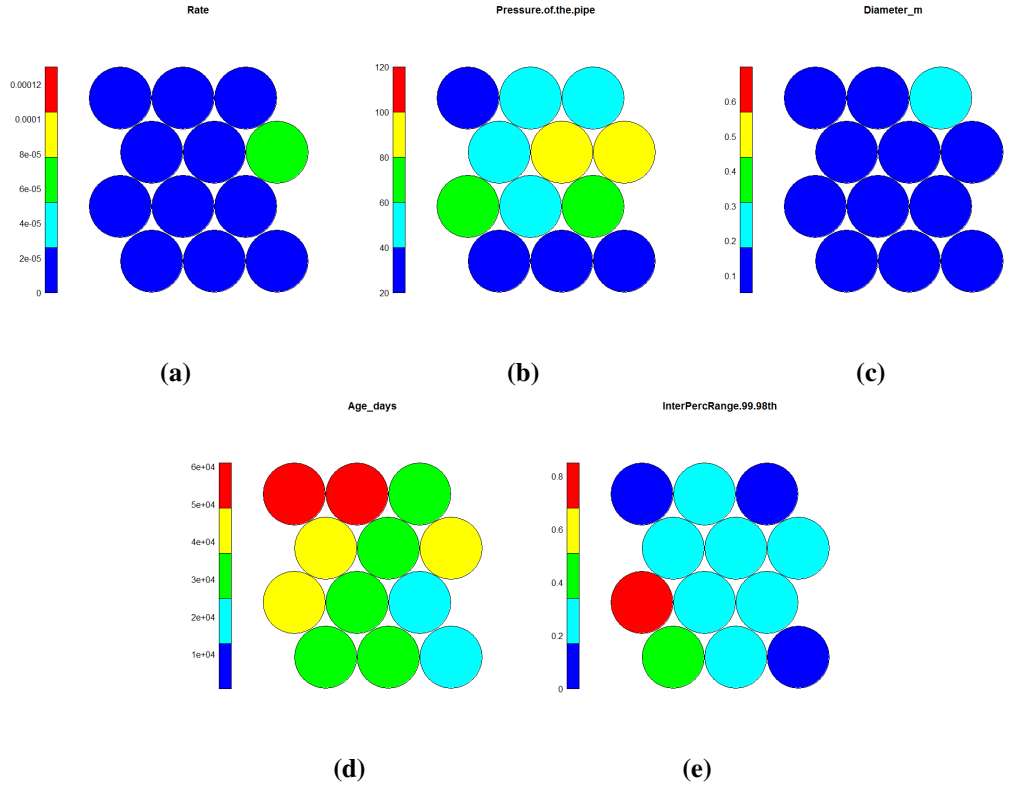


Figure B.7: Pipe measured SOMs output (a) pipe repair rate, (b) hydraulic static pressure, (c) diameter, (d) age, (e) the 99.98th interpercentile range of the rate of change of pressure.

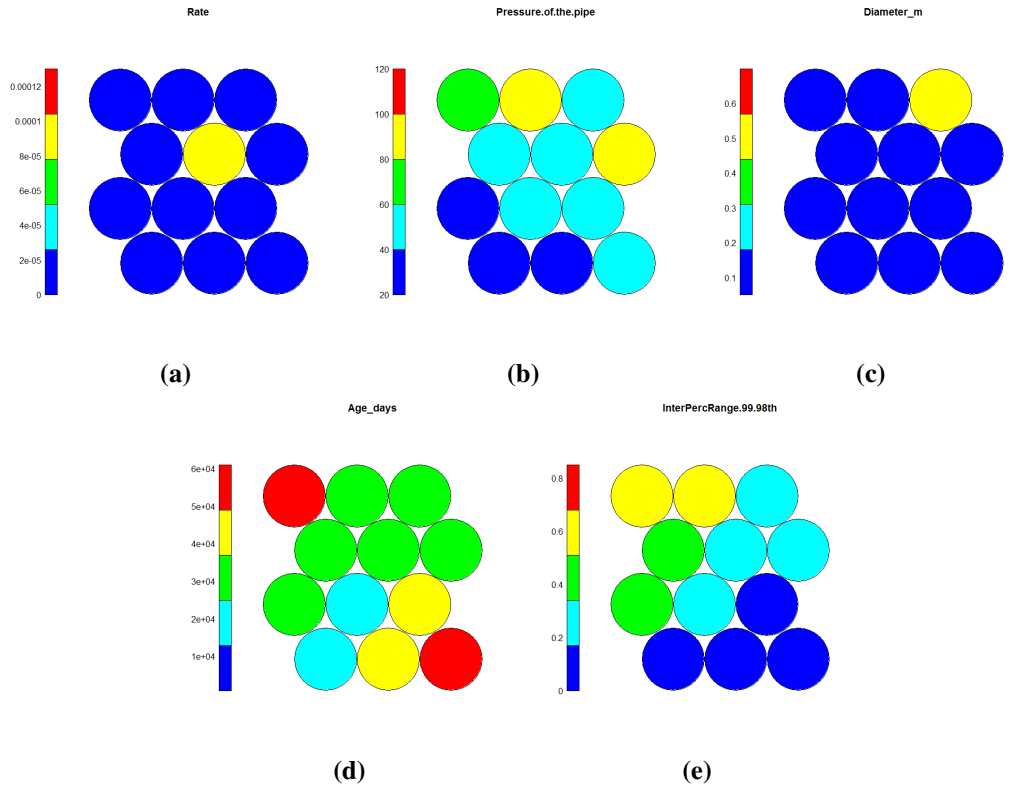
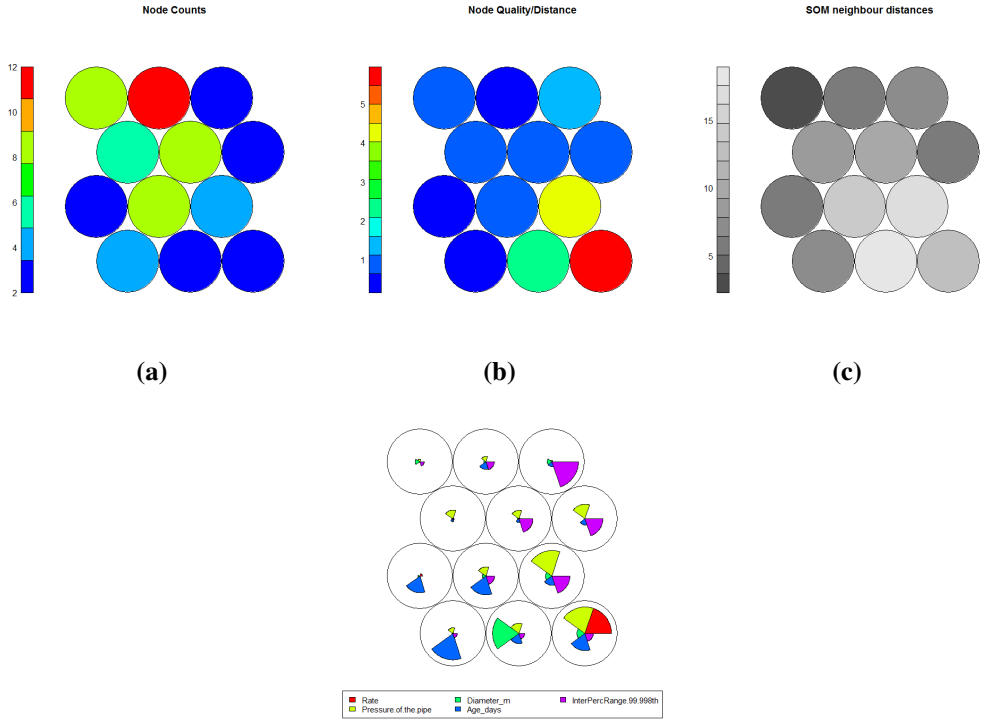


Figure B.8: Zone level SOMs output (a) pipe repair rate, (b) hydraulic static pressure, (c) diameter, (d) age, (e) the 99.98th interpercentile range of the rate of change of pressure.

Figures B.9 and Figures B.10 show preliminary SOMs variables assessment for the the 99.998th interpercentile range.



(a) **(b)** **(c)**

(d)

Figure B.9: Preliminary pipe-measured SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

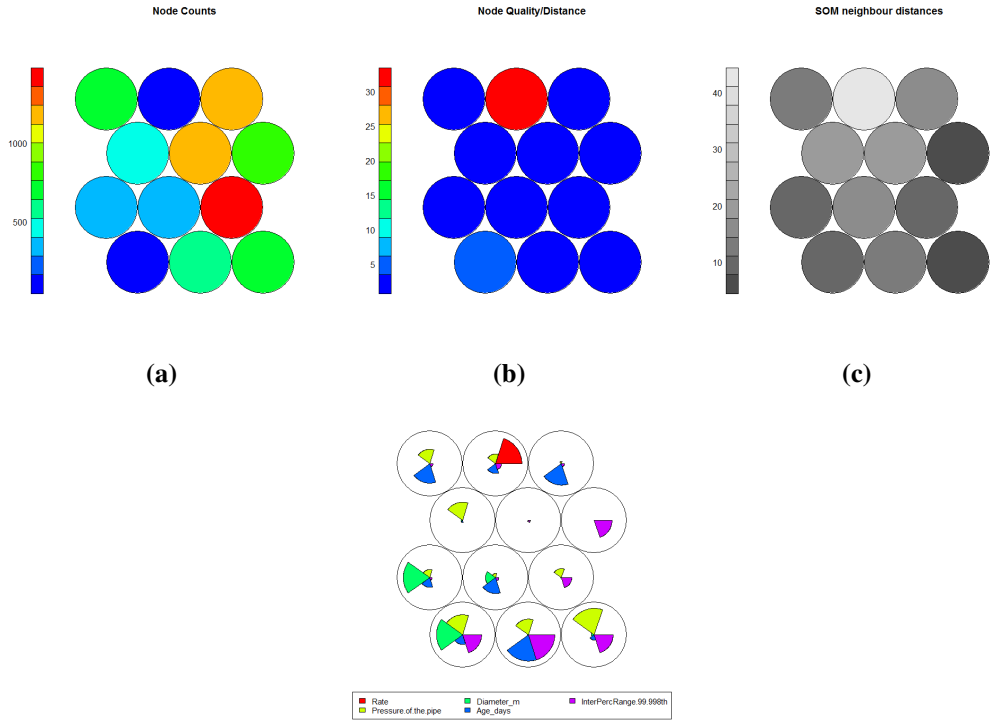


Figure B.10: Preliminary zone-level SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

Figure B.11 and Figure B.12 present the 99.998th interpercentile range for the pipe-measured and the zone-level data respectively. The further increase in the variability and correlation with pipe repairs is observed. According to the SOMs assessment criteria the smallest value for which variability in the SOMs map is observed should be retained. The variability was already observed in the previous variable, the 99.98th interpercentile range, therefore the 99.998th interpercentile range is not selected and the 99.98th interpercentile range presented in Figure B.7 and Figure B.8 are taken forward to EPR.

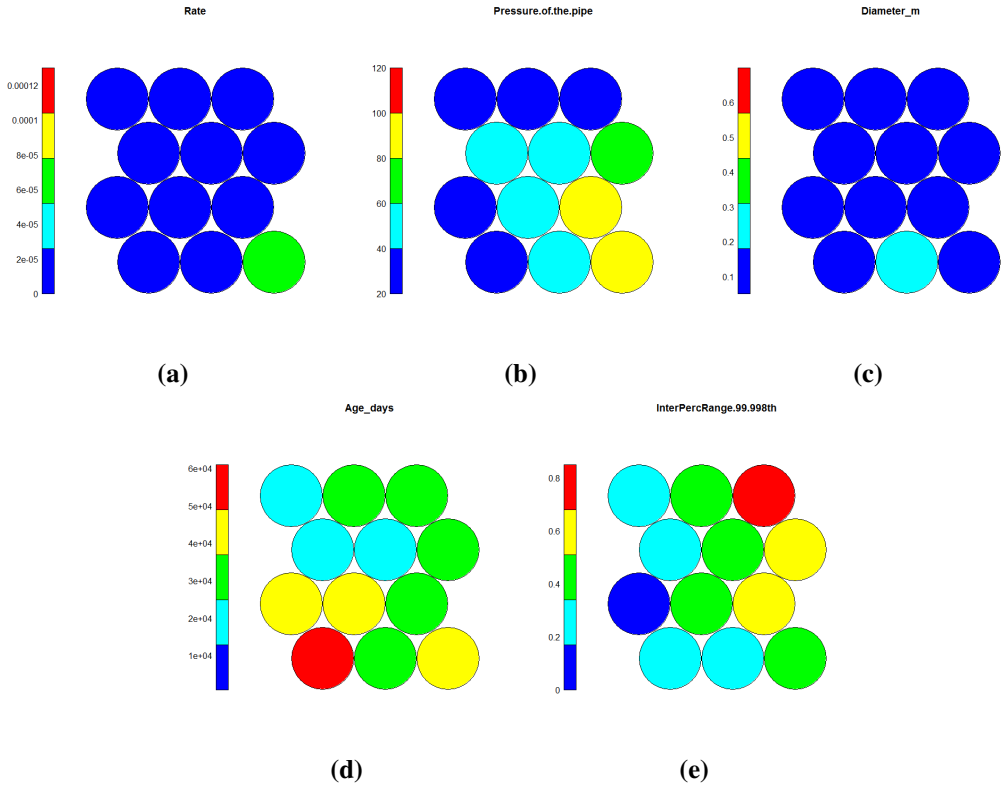


Figure B.11: Pipe measured SOMs output (a) pipe repair rate, (b) hydraulic static pressure, (c) diameter, (d) age, (e) the 99.998th interpercentile range of the rate of change of pressure.

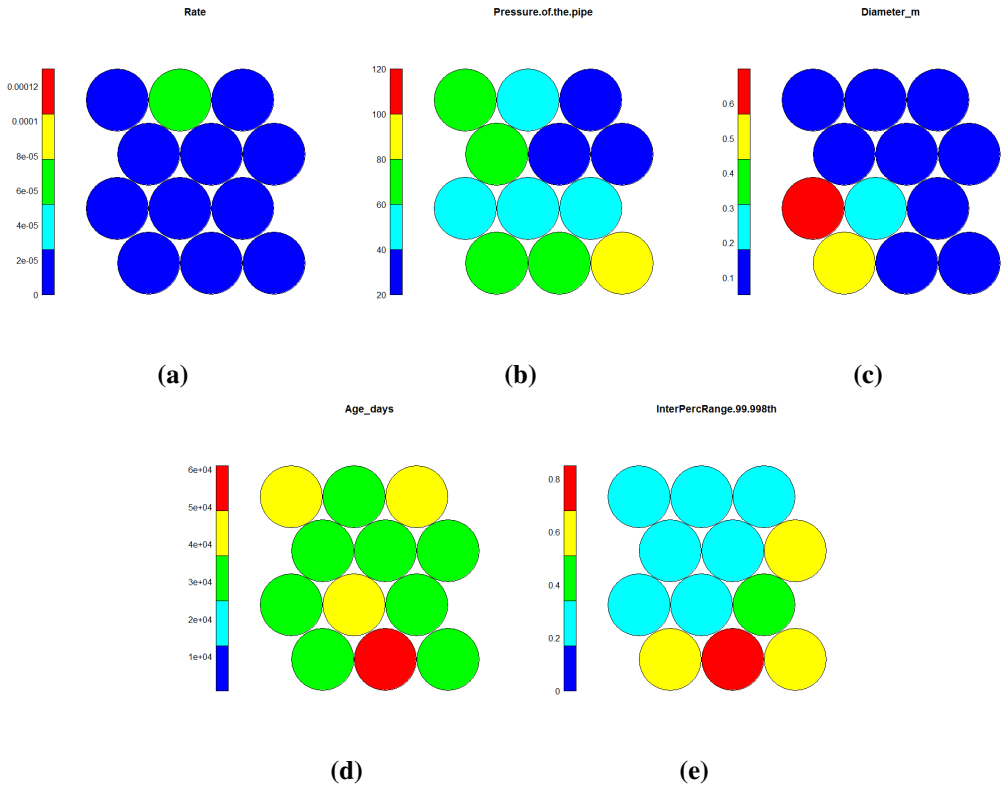


Figure B.12: Zone level SOMs output (a) pipe repair rate, (b) hydraulic static pressure, (c) diameter, (d) age, (e) the 99.998th interpercentile range of the rate of change of pressure.

B.2 SOMs transient fingerprint

Figures B.13 and Figures B.14 show preliminary SOMs variables assessment for the 99% transient categorisation to identify strong and distinct clusters that can be compared to a full SOMs map (see Figure B.15 and Figure B.16).

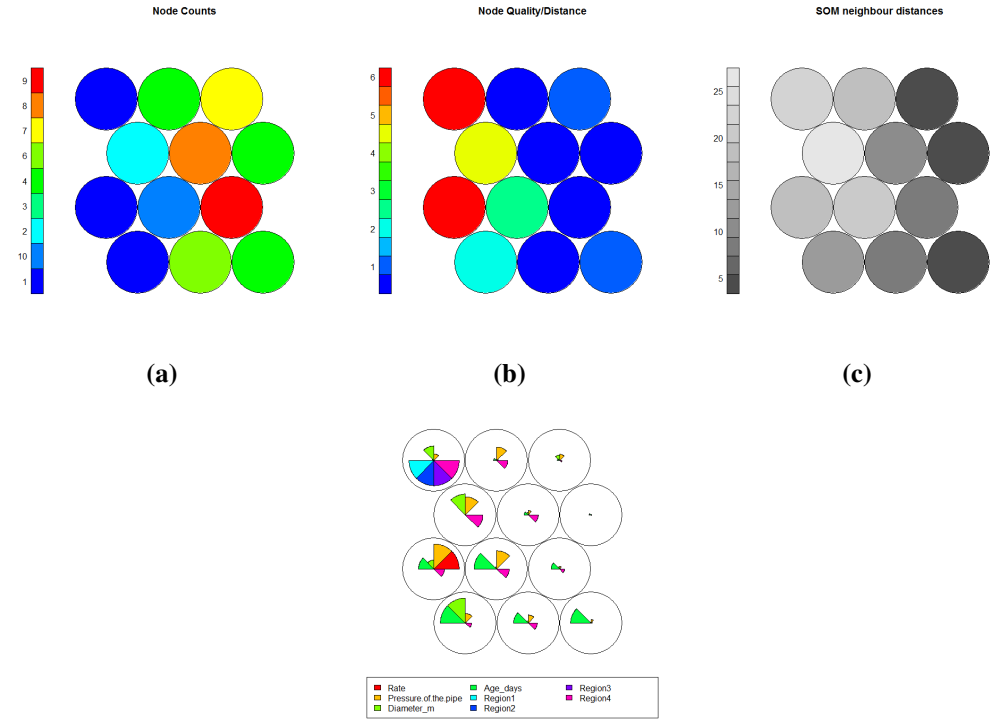


Figure B.13: Preliminary pipe-measured SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

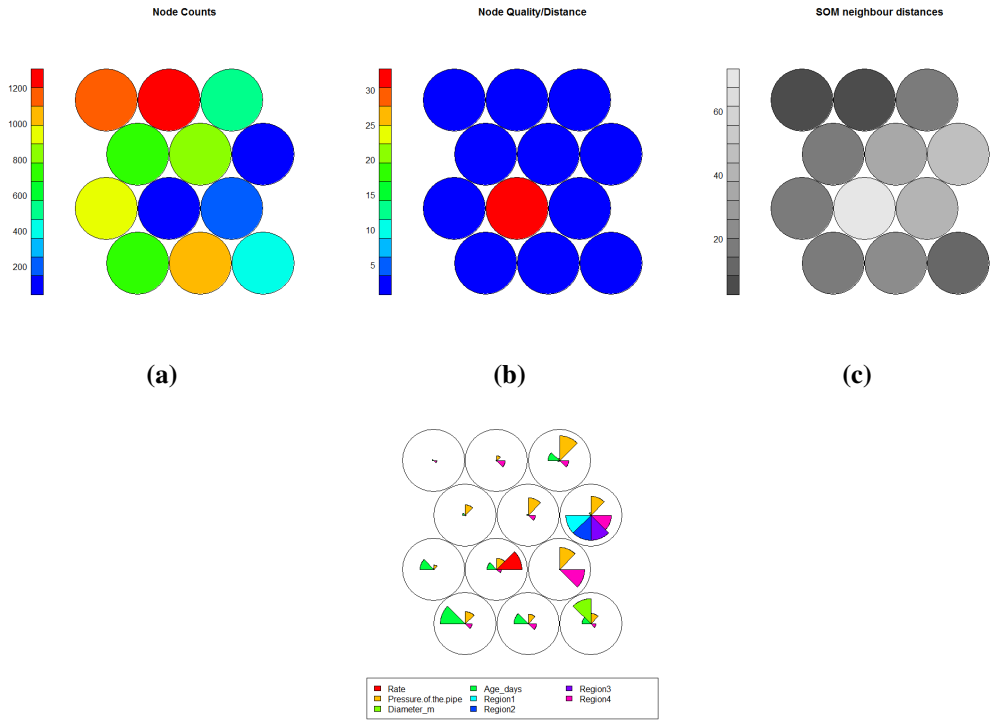


Figure B.14: Preliminary zone-level SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

Figure B.15 and Figure B.16 present the 99% transient categorisation for the pipe-measured and the zone-level data respectively. The figures show that the separations/ differences between

regions 1, 2 and 3 are not observed.

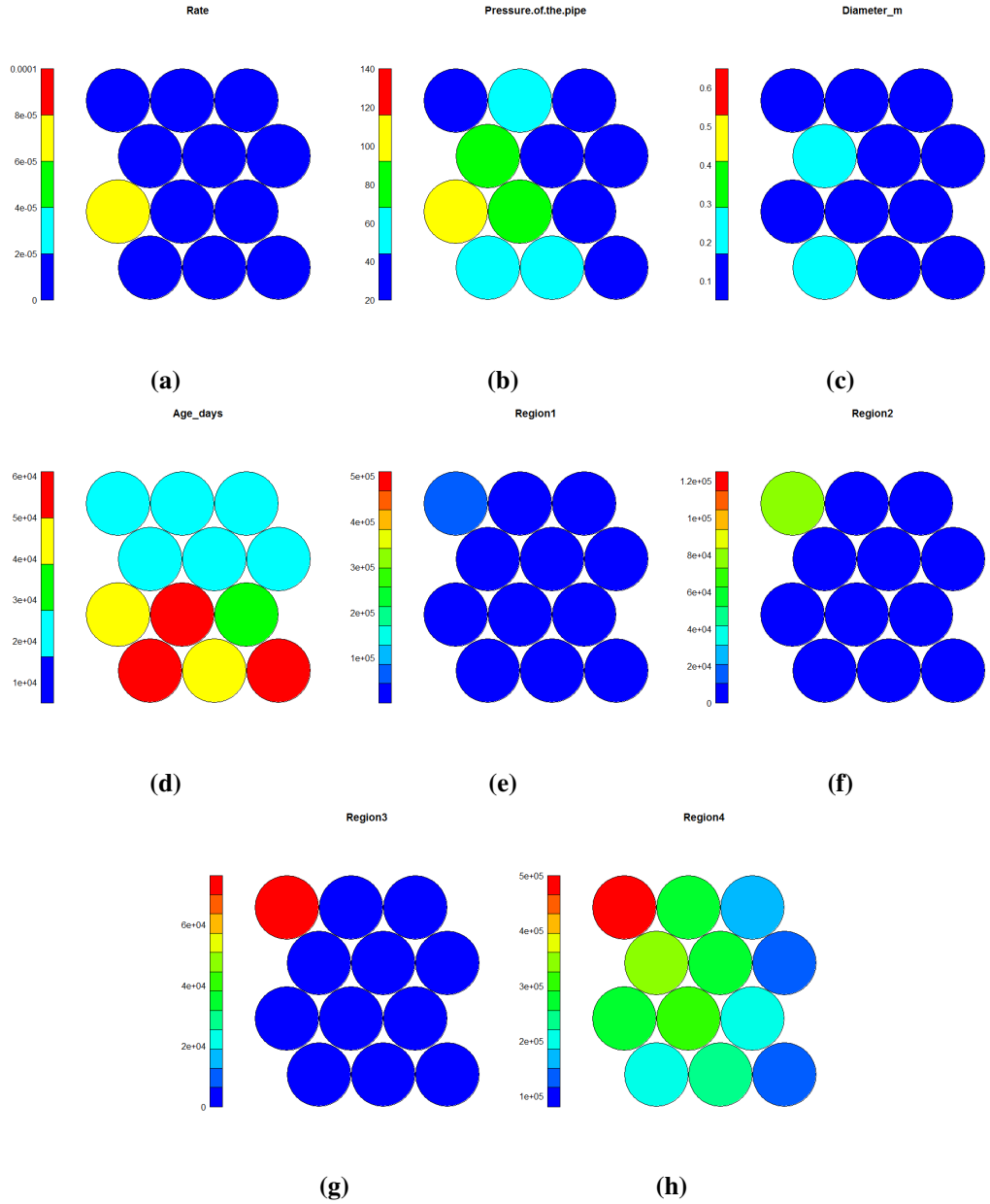


Figure B.15: Pipe measured SOMs output for a base case and the 99% transient categorisation, (a) pipe repair rate, (b) diameter, (c) age, (d) hydraulic static pressure, (e) counts - region 1, (f) counts - region 2, (g) counts - region 3, (h) counts - region 4.

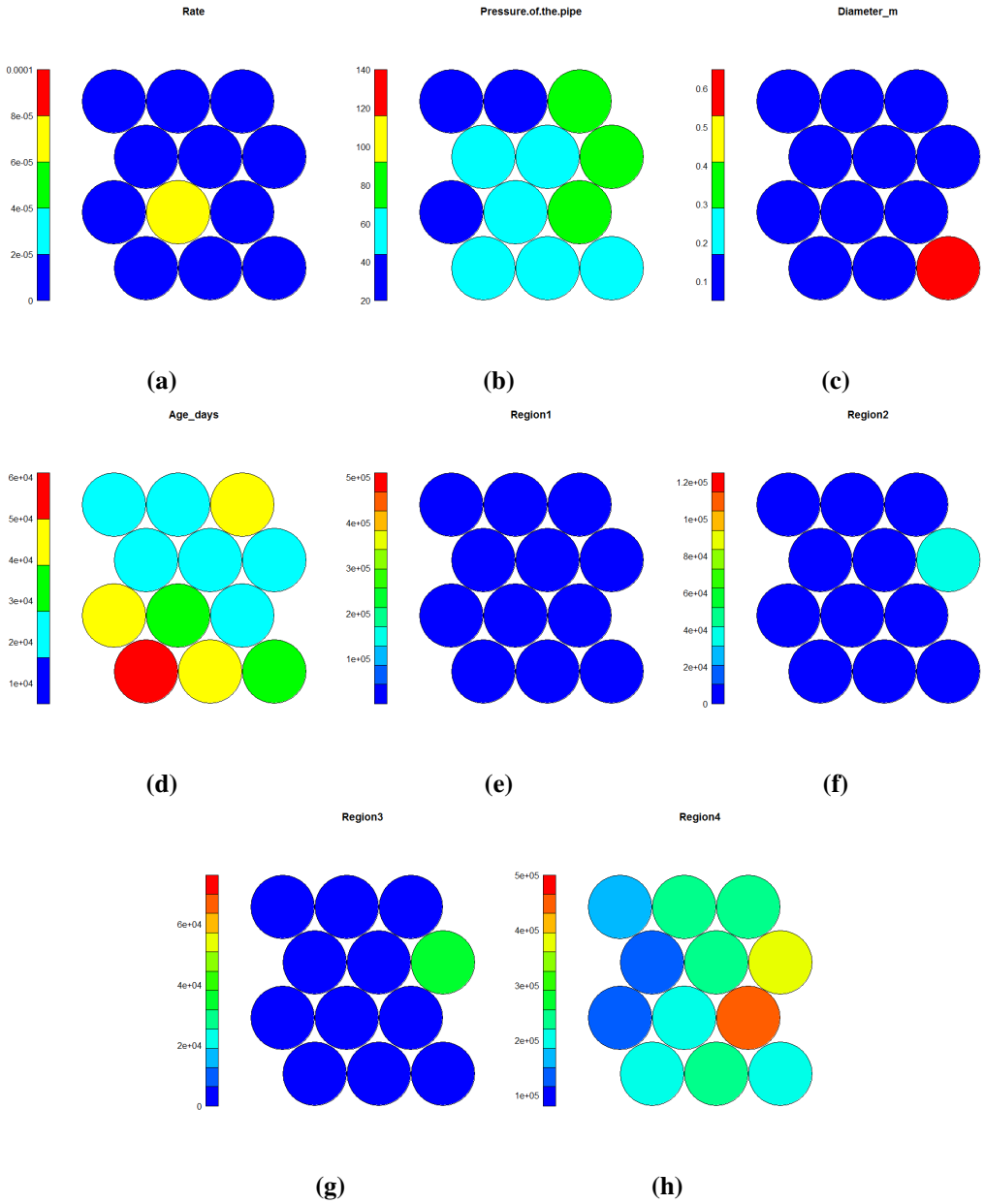
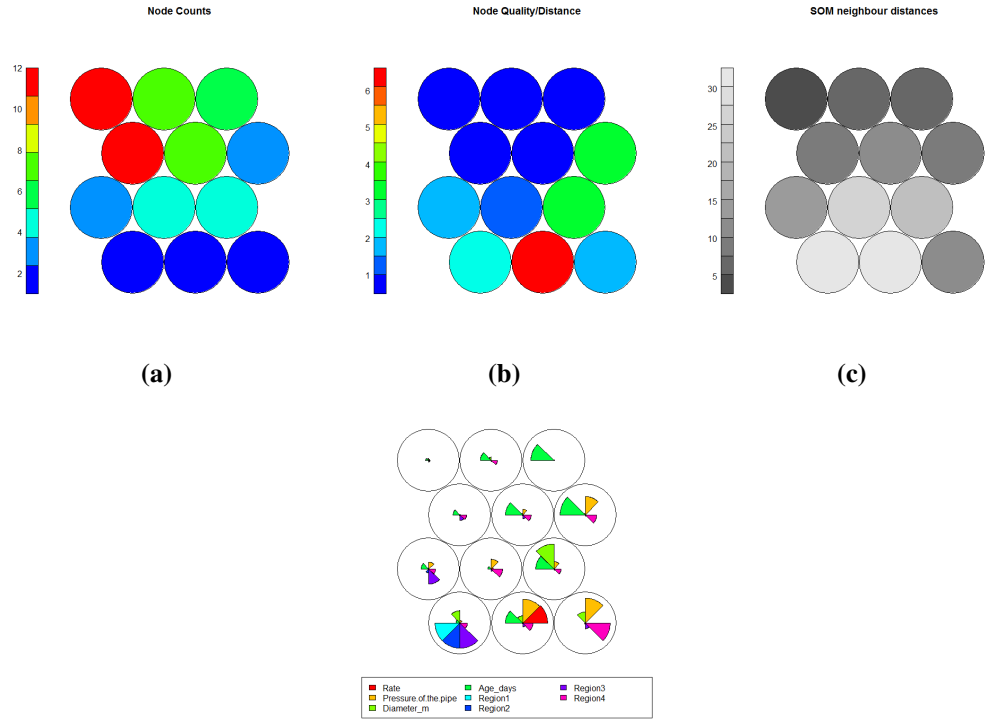


Figure B.16: Zone level SOMs output for a base case and the 99% transient categorisation, (a) pipe repair rate, (b) diameter, (c) age, (d) hydraulic static pressure, (e) counts- region 1, (f) counts- region 2, (g) counts- region 3, (h) counts- region 4.

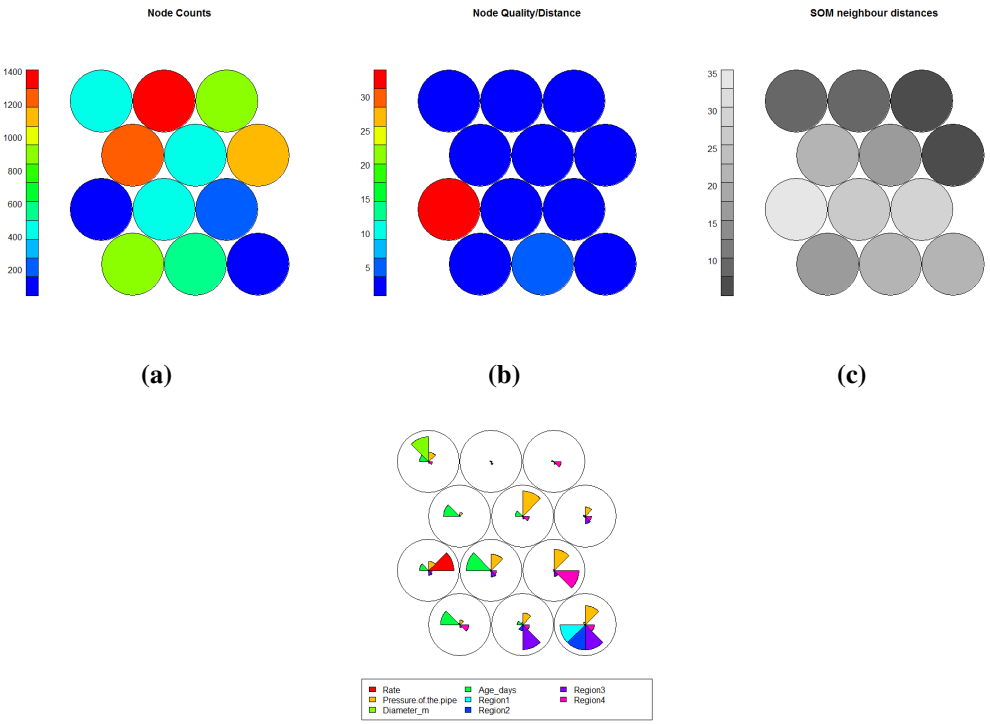
Figures B.17 and Figures B.18 show preliminary SOMs variables assessment for the 95% transient categorisation to identify strong and distinct clusters that can be compared to a full SOMs map.



(a) **(b)** **(c)**

(d)

Figure B.17: Preliminary pipe-measured SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.



(a) **(b)** **(c)**

(d)

Figure B.18: Preliminary zone-level SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

The 95% transient categorisation presented in Figure B.19 and Figure B.20 does not show differences between regions 1 and 2. Some differences, however, are noticeable in region 3.

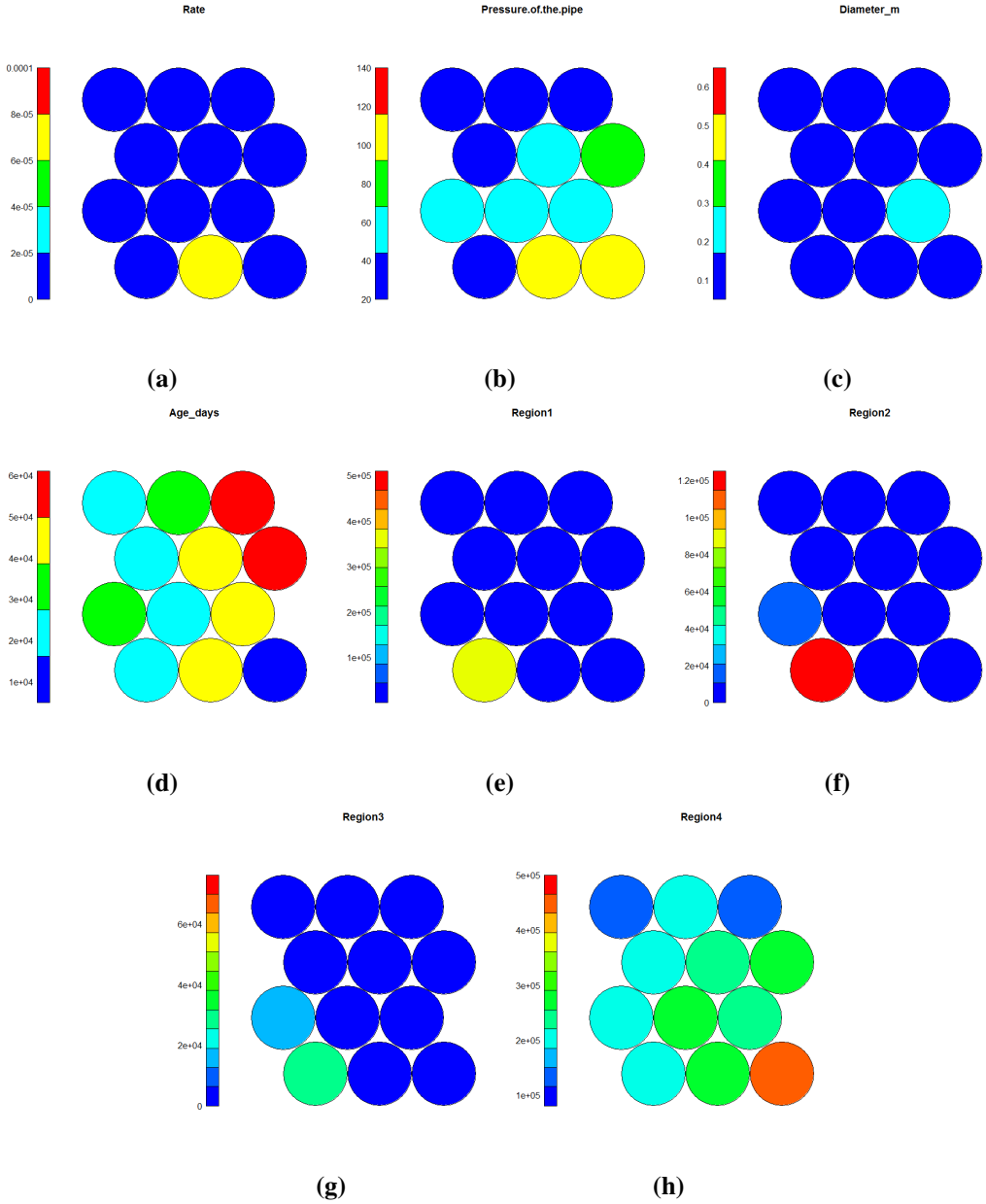


Figure B.19: Pipe measured SOMs output for a base case and the 95% transient categorisaiton, (a) pipe repair rate, (b) diameter, (c) age, (d) hydraulic static pressure, (e) counts - region 1, (f) counts - region 2, (g) counts - region 3, (h) counts - region 4.

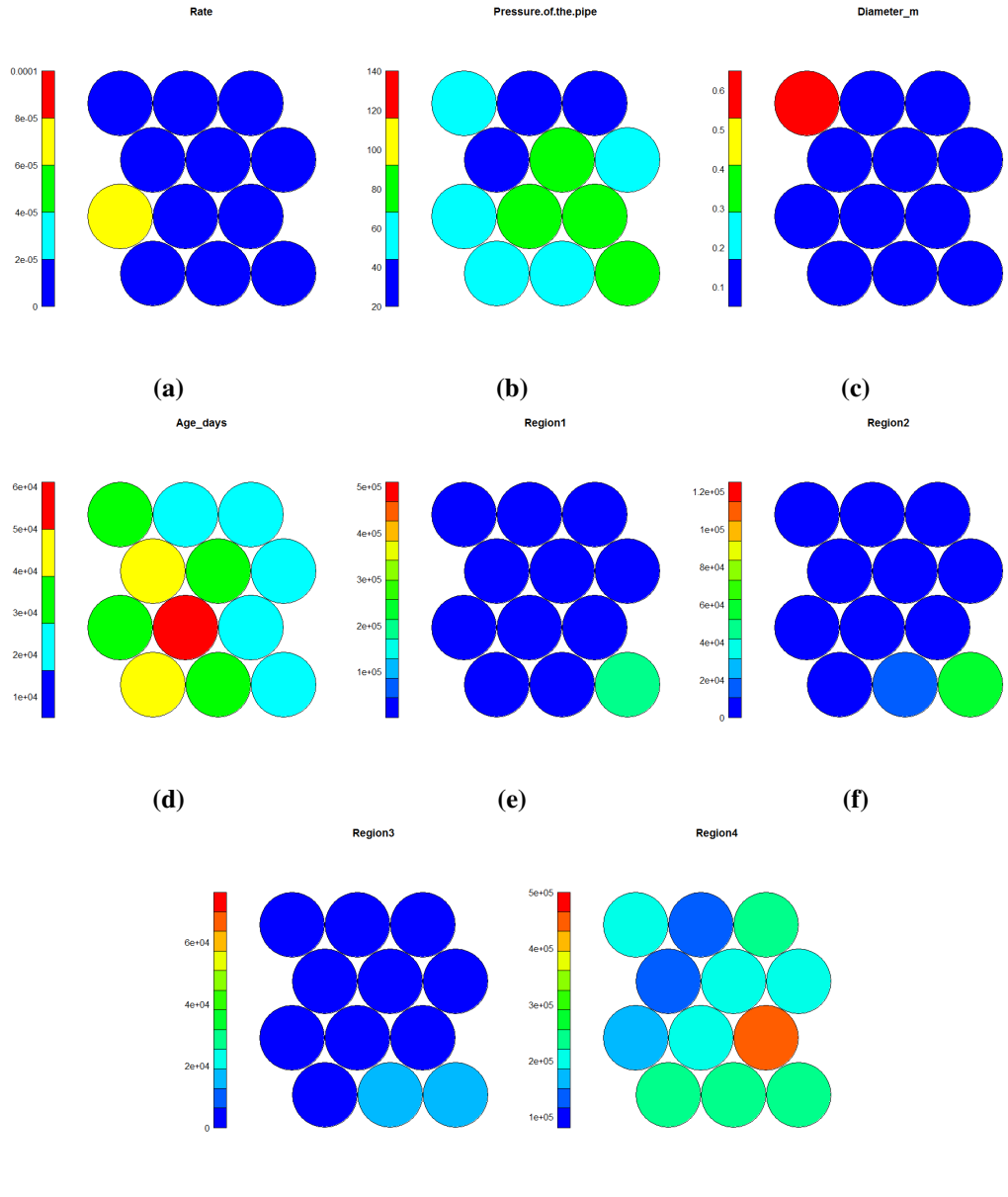
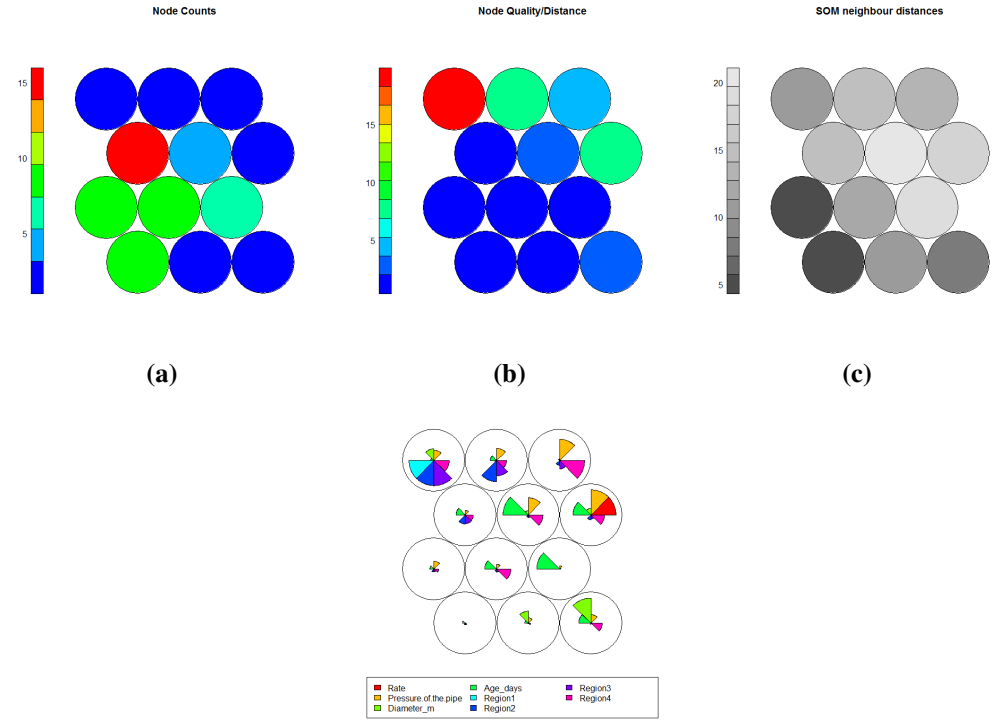


Figure B.20: Zone level SOMs output for a base case and the 95% transient categorisation, (a) pipe repair rate, (b) diameter, (c) age, (d) hydraulic static pressure, (e) counts- region 1, (f) counts- region 2, (g) counts- region 3, (h) counts- region 4.

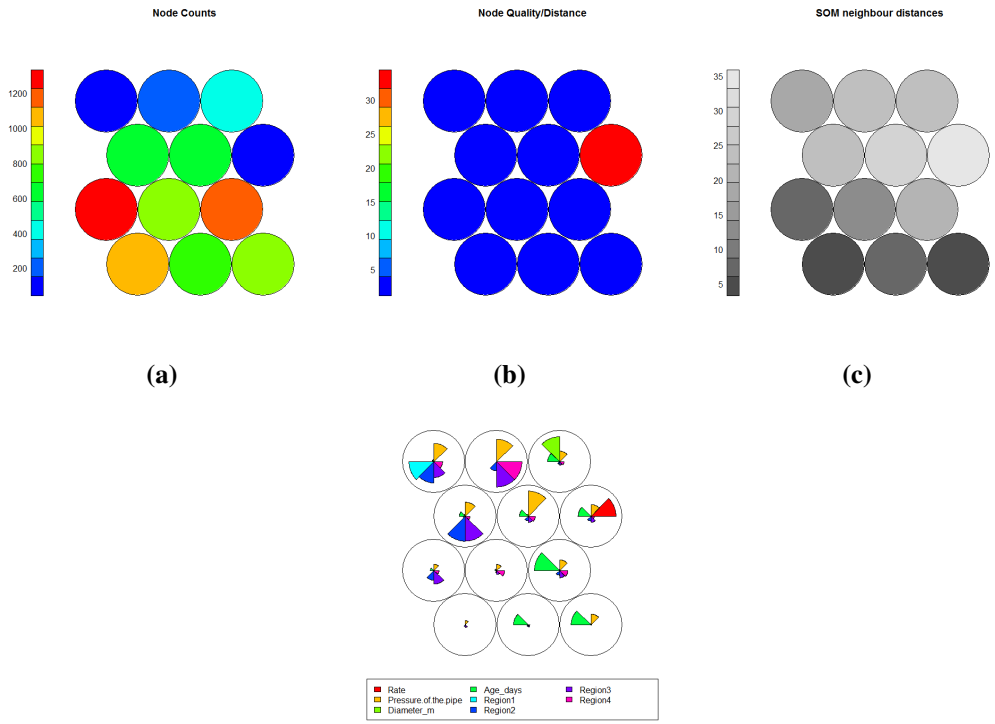
Figures B.21 and Figures B.22 show preliminary SOMs variables assessment for the 90% transient categorisation to identify strong and distinct clusters that can be compared to a full SOMs map.



(a) **(b)** **(c)**

(d)

Figure B.21: Preliminary pipe-measured SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.



(a) **(b)** **(c)**

(d)

Figure B.22: Preliminary zone-level SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

The 90% transient categorisation presented in Figure B.23 and Figure B.24 show differences in regions 1, 2, 3 and 4.

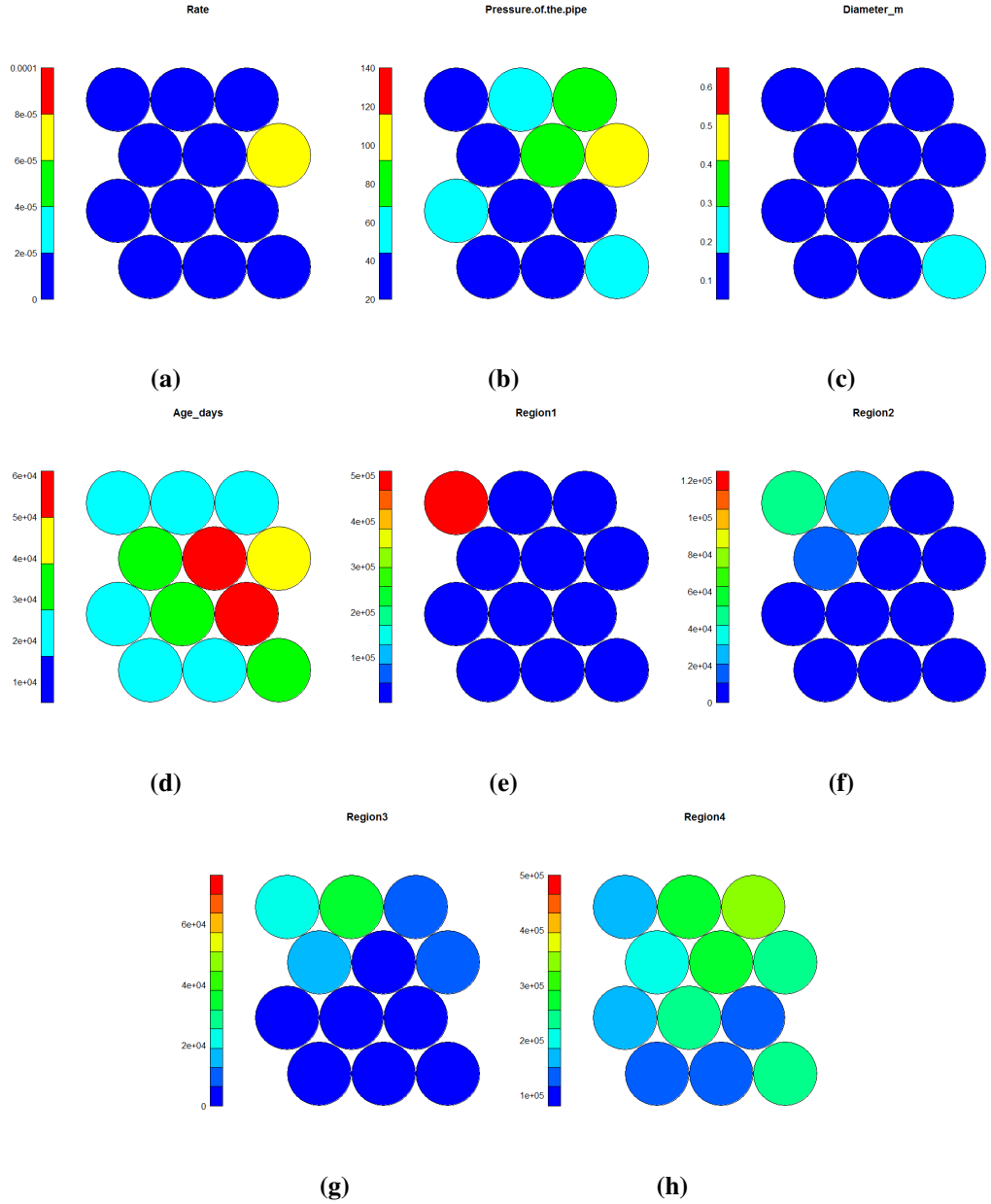


Figure B.23: Pipe measured SOMs output for a base case and the 90% transient categorisation, (a) pipe repair rate, (b) diameter, (c) age, (d) hydraulic static pressure, (e) counts - region 1, (f) counts - region 2, (g) counts - region 3, (h) counts - region 4.

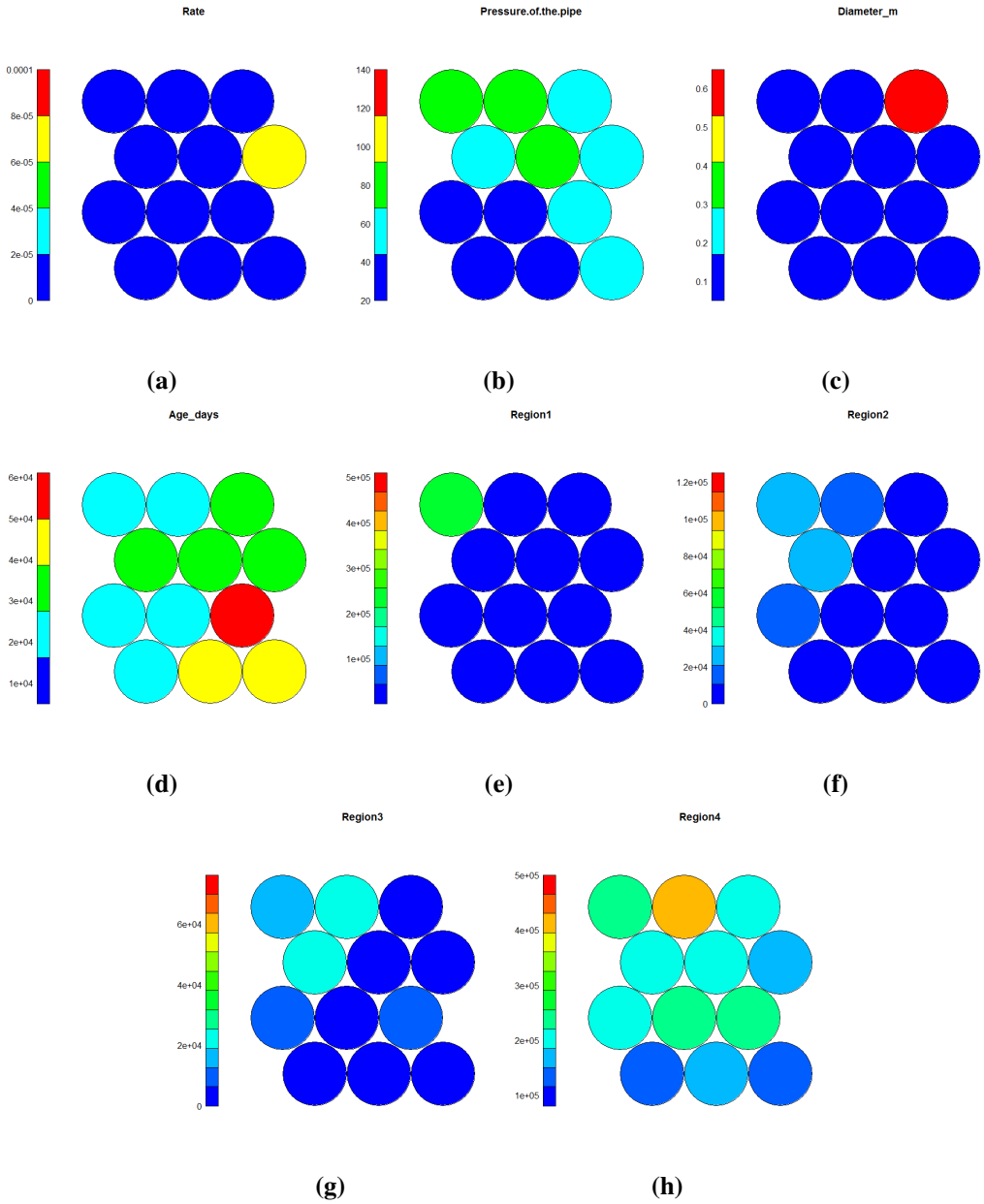
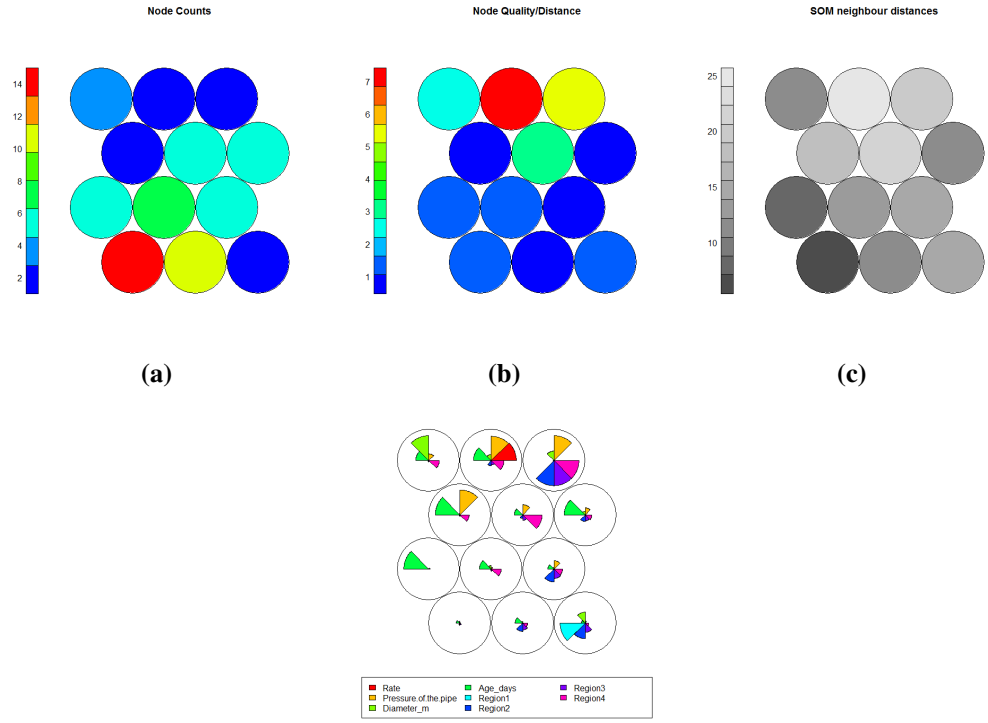


Figure B.24: Zone level SOMs output for a base case and the 90% transient categorisation, (a) pipe repair rate, (b) diameter, (c) age, (d) hydraulic static pressure, (e) counts- region 1, (f) counts- region 2, (g) counts- region 3, (h) counts- region 4.

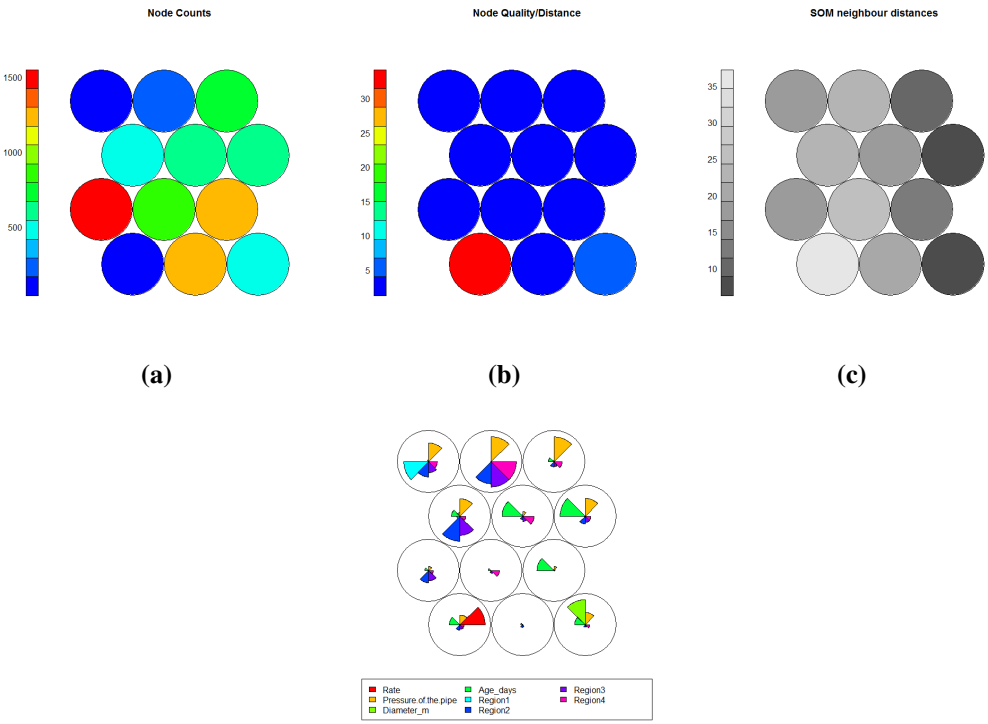
Figures B.25 and Figures B.26 show preliminary SOMs variables assessment for the 85% transient categorisation to identify strong and distinct clusters that can be compared to a full SOMs map.



(a) **(b)** **(c)**

(d)

Figure B.25: Preliminary pipe-measured SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.



(a) **(b)** **(c)**

(d)

Figure B.26: Preliminary zone-level SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

The 85% transient categorisation presented in Figure B.27 and Figure B.28 show further increase in separations between four regions respectively.

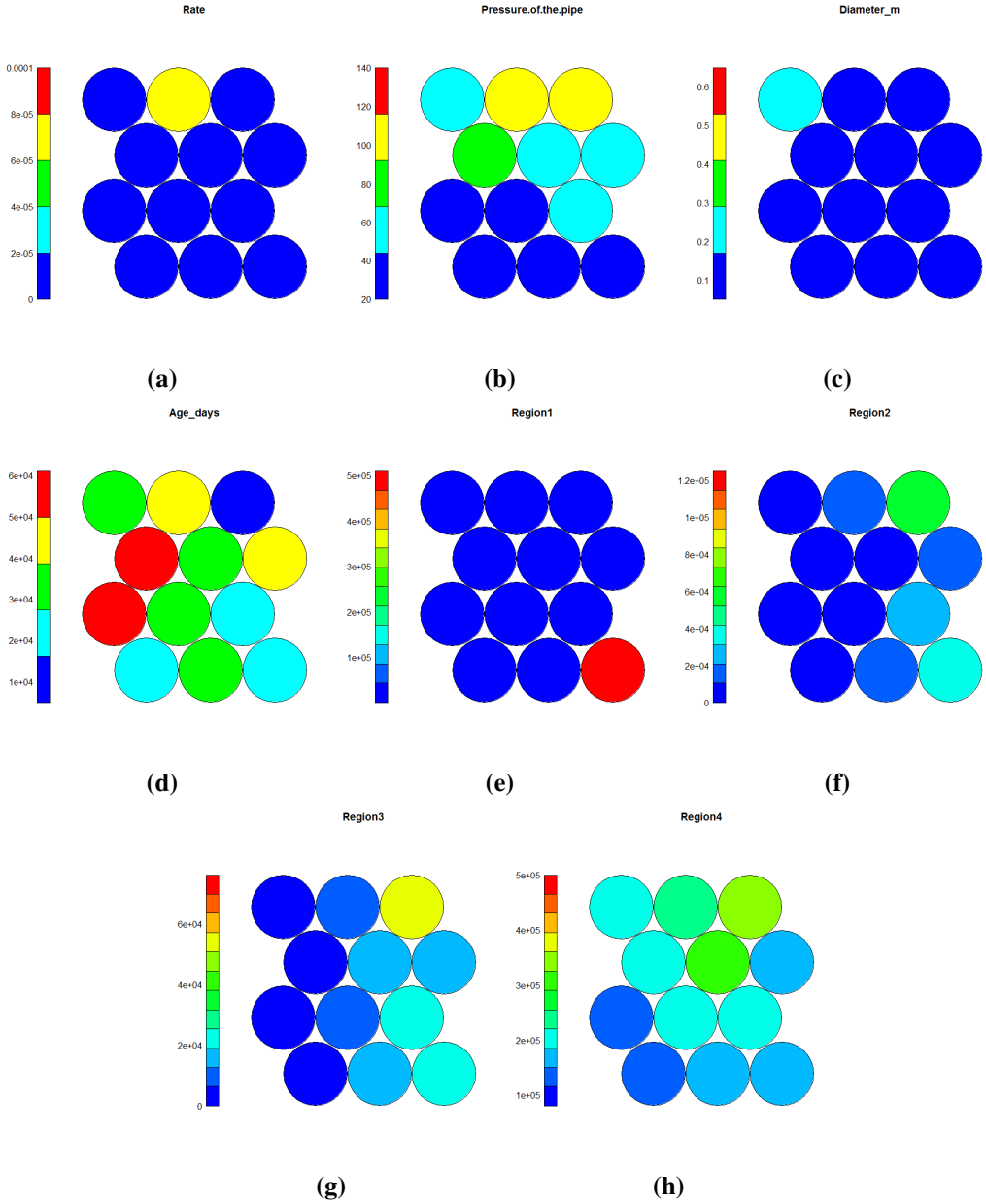


Figure B.27: Pipe measured SOMs output for a base case and the 85% transient categorisation, (a) pipe repair rate, (b) diameter, (c) age, (d) hydraulic static pressure, (e) counts - region 1, (f) counts - region 2, (g) counts - region 3, (h) counts - region 4.

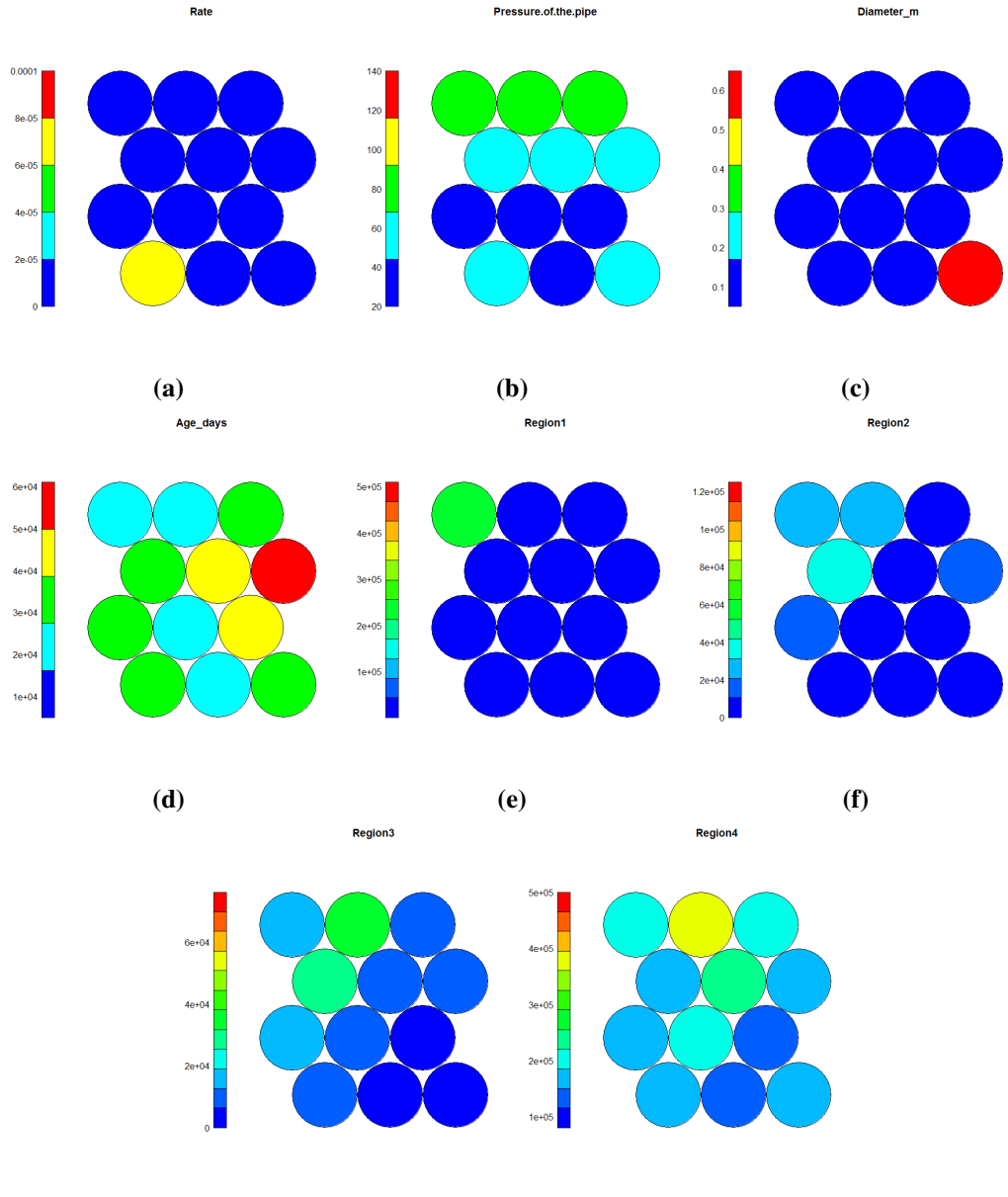


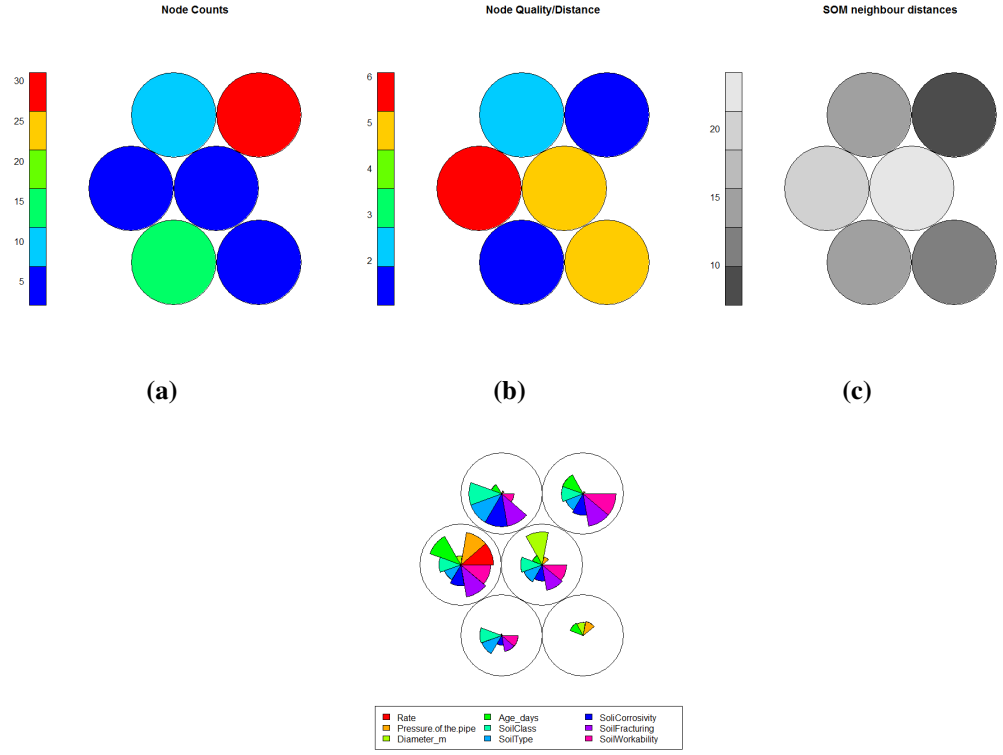
Figure B.28: Zone level SOMs output for a base case and the 85% transient categorisation, (a) pipe repair rate, (b) diameter, (c) age, (d) hydraulic static pressure, (e) counts- region 1, (f) counts- region 2, (g) counts- region 3, (h) counts- region 4.

As can be seen in the presented figures neither the 99% transient categorisation or the 95% transient categorisation produced the separation between regions of the data nor provided any additional information. Therefore both categorisations are rejected.

B.3 SOMs soil data

Figure B.29 shows the initial SOMs variables assessment for the pipe-measured data. It shows that there are no empty clusters and pipe repair rates is a distinct variable.

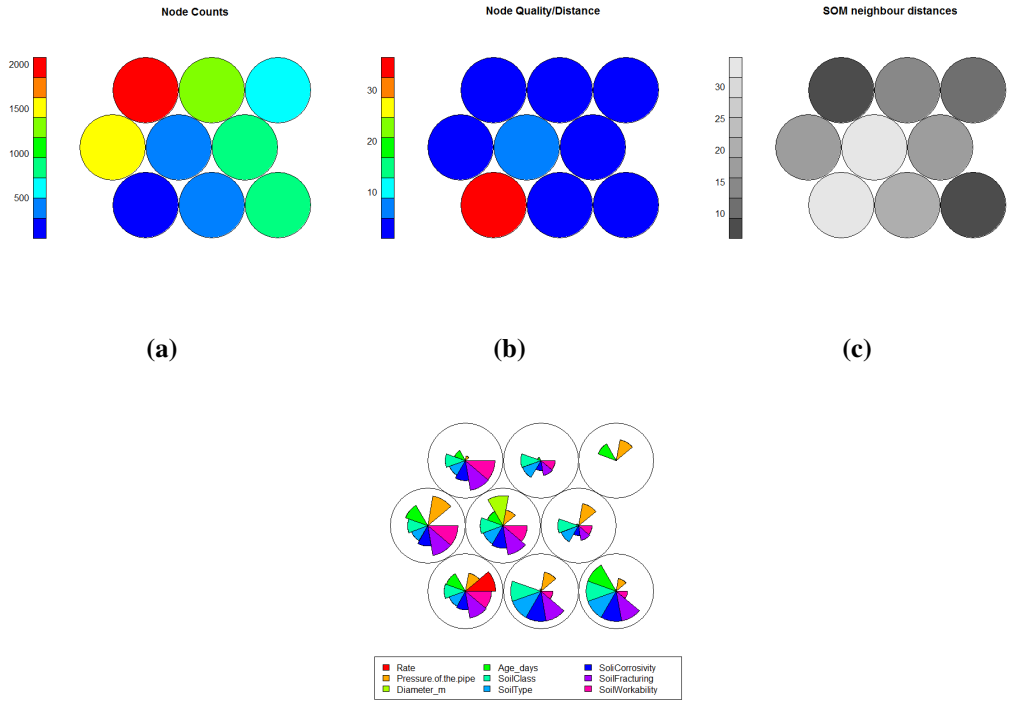
Figure B.30 shows the initial SOMs variables assessment for the zone-level data. A distinct cluster representing pipe repair rates is identified.



(a) **(b)** **(c)**

(d)

Figure B.29: Preliminary pipe level SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.



(a) **(b)** **(c)**

(d)

Figure B.30: Preliminary pipe level SOM variables assessment (a) node counts, (b) node quality/distance, (c) SOM neighbour distance, (d) Weight vectors.

References

Ahammed, M. and Melchers, R. Reliability of underground pipelines subject to corrosion. *Journal of transportation engineering*, 120(6):989–1002, 1994.

Ahlburg, A. A commentary on error measures. *International Journal of Forecasting*, 8:99–111, 1992.

Altomare, C., Gironella, X., and Laucelli, D. Evolutionary data-modelling of an innovative low reflective vertical quay. *Journal of Hydroinformatics*, 15(3):763–779, 2013. ISSN 1464-7141.

Anderson, M. and Woessner, W. *Applied groundwater modeling: simulation of flow and advective transport*. 1992. ISBN 0120594854. doi: 10.1016/0169-7722(92)90015-7.

Andreou, S. *Predictive models for pipe break failures and their implications on maintenance planning for deteriorating water distribution systems*. PhD thesis, Massachusetts Institute of Technology, 1986.

Andreou, S., Marks, D., and Clark, R. A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory. *Advances in Water Resources*, 10(1):2–10, mar 1987. ISSN 03091708. doi: 10.1016/0309-1708(87)90002-9.

Baracos, A. and Hurst, W. Effects of physical environment on cast-iron pipe. *Journal (American Water Works Association)*, 47(12):1195–1206, 1955.

Berardi, L., Savic, D., and Giustolisi, O. Investigation of burst-prediction formulas for water distribution systems by evolutionary computing. In *Proceeding of the 8th International Conference on Computing and Control for the Water Industry*, pages 275–280, 2005.

Berardi, L., Giustolisi, O., Kapelan, Z., and Savic, D. A. Development of pipe deterioration models for water distribution systems using EPR. *Journal of Hydroinformatics*, 10(2):113, 2008. ISSN 1464-7141.

Berardi, L., Laucelli, D., and Giustolisi, O. A decision support tool for a sustainable management of water distribution networks. *Plurimondi*, (7):49–74, 2010.

Bergant, A., Simpson, A., and Tijsseling, A. Water hammer with column separation: A historical review. *Journal of Fluids and Structures*, 22(2):135–171, feb 2006. ISSN 08899746.

Bergant, A., Tijsseling, A. S., Vítkovský, J. P., Covas, D. I., Simpson, A. R., and Lambert, M. F. Parameters affecting water-hammer wave attenuation, shape and timing—Part 2: Case studies. *Journal of Hydraulic Research*, 46(3):382–391, may 2008a. ISSN 0022-1686. doi: 10.3826/jhr.2008.2847.

Bergant, A., Tijsseling, A. S., Vítkovský, J. P., Covas, D. I., Simpson, A. R., and Lambert, M. F. Parameters affecting water-hammer wave attenuation, shape and timing—Part 1: Mathematical tools. *Journal of Hydraulic Research*, 46(3):373–381, may 2008b. ISSN 0022-1686. doi: 10.3826/jhr.2008.2848.

Besner, M., Broséus, R., Lavoie, J., Giovanni, G., Payment, P., and Prévost, M. Pressure monitoring and characterization of external sources of contamination at the site of the payment drinking water epidemiological studies. *Environmental science & technology*, 44(1):269–77, jan 2010. ISSN 0013-936X. doi: 10.1021/es901988y.

Blokker, E., Furnass, W., Machell, J., Mounce, S., Schaap, P., and Boxall, J. Relating water quality and age in drinking water distribution systems using self-organising maps. *Environments*, 3(2): 10, 2016. ISSN 2076-3298.

Box, G., Jenkins, G., and Reinsel, G. *Time series analysis: forecasting and control*. Englewood Cliffs, N.J. : Prentice Hall, 1994, 3rd ed. edition, 1994.

Boxall, J., O'Hagan, A., Pooladsaz, S., Saul, A., and Unwin, D. Pipe level estimation of burst rates in water distribution mains. In *Proceeding of the 8th International Conference on Computing and Control for the Water Industry, CCWI2005*, pages 33–38, Exeter, 2005. University of Exeter, UK.

Boxall, J., O'Hagan, A., Pooladsaz, S., Saul, A., and Unwin, D. Estimation of burst rates in water distribution mains. *Proceedings of the ICE - Water Management*, 160(2):73–82, jan 2007. ISSN 1741-7589. doi: 10.1680/wama.2007.160.2.73.

Brandon, T. *Water distribution systems*. 1984.

Brevis, W., Susmel, L., and Boxal, J. Investigating in-service failures of water pipes from a multiaxial notch fatigue point of view: a conceptual study. *Proceeding of the Institution of Mechanical Engineering, PartC: Journal of Mechanical Engineering Science*, 10:1–20, 2014. ISSN 0954-4062.

Chaudhry, M. *Applied hydraulic transients*. Van Nostrand Reinhold Co, New York, 1987.

Clark, R., Stafford, C., and Goodrich, J. Water distribution systems: a spatial and cost evaluation. *Journal of the water ressources, planning and management.*, 108:243–256, 1982.

Constantine, G., Darroch, J., and Miller, R. Predicting underground pipeline failure. *Australian Water and Wastewater*, 23(2):9–10, 1996.

Coombes, K., Fritsche, H., Clarke, C., Chen, J., Baggerly, K., Morris, J., Xiao, L., Hung, M., and Kuerer, H. Quality Control and Peak Finding for Proteomics Data Collected from Nipple Aspirate Fluid by Surface-Enhanced Laser Desorption and Ionization. *Clinical Chemistry*, 49 (10):1615–1623, 2003.

Cooper, G., Mcgechan, M., and Vinten, A. The Influence of a Changed Climate on Soil Workability and Available Workdays in Scotland. *Journal of Agricultural Engineering Resources*, 68:253–269, 1997. ISSN 00218634.

Cooper, N., Blakey, G., Sherwin, C., Ta, T., Whiter, J., and Woodward, C. The use of GIS to develop a probability-based trunk mains burst risk model. *Urban Water*, 2(2):97–103, 2000. ISSN 14620758.

Covas, D., Ramos, H., and Almeida, A. B. D. Standing Wave Difference Method for Leak Detection in Pipeline Systems. (December):1106–1116, 2005.

Creasey, J. and Garrow, D. Investigation of instances of low or negative pressures in UK drinking water systems - Final report. Technical Report October, WRc, UK, 2011.

Davis, P., Burn, S., Moglia, M., and Gould, S. A physical probabilistic model to predict failure rates in buried PVC pipelines. *Reliability Engineering & System Safety*, 92(9):1258–1266, sep 2007. ISSN 09518320. doi: 10.1016/j.ress.2006.08.001.

Debón, A., Carrión, A., Cabrera, E., and Solano, H. Comparing risk of failure models in water supply networks using ROC curves. *Reliability Engineering & System Safety*, 95(1):43–48, jan 2010. ISSN 09518320. doi: 10.1016/j.ress.2009.07.004.

Dehghan, A., Gad, E. F., and McManus, K. J. Statistical analysis of structural failures of water pipes. *Proceedings of the ICE - Water Management*, 161(4):207–214, aug 2008. ISSN 1741-7589. doi: 10.1680/wama.2008.161.4.207.

D'Souza, A. *Mechanics of materials*. The Pitman Press, Bath, 1966.

Du, P., Kibbe, W. a., and Lin, S. M. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics (Oxford, England)*, 22 (17):2059–65, sep 2006. ISSN 1367-4811.

Dyachkov, A. Rehabilitation of the water disribution network in the city of Moscow. *Water Supply -Oxford-*, 12(3/4):89–94, 1994.

Ebacher, G., Besner, M., Lavoie, J., Jung, B., Karney, B., and Prévost, M. Transient Modeling of a Full-Scale Distribution System: Comparison with Field Data. *Journal of Water Resources Planning and Management*, (4):173–182, 2011. doi: 10.1061/(ASCE)WR.1943-5452.

Economou, T., Kapelan, Z., and Bailey, T. An aggregated hierarchical Bayesian model for the prediction of pipe failures. *Proceedings of the combined international conference of computing and control for the water industry SUWM2007*, pages 1–7, 2007.

Edwards, J. and Collins, R. The Effect of Demand Uncertainty on Transient Propagation in Distribution Systems. *Computer and Contol in the Water Industry, CCWI. Perugia*, 2013.

Farley, B., Boxall, J., and Mounce, S. Optimal locations of pressure meters for burst detection. *Proceeding of 10th Annual International Symposium on Water Distribution Systems Analysis*, pages 1–11, 2008. ISSN 08950563.

Farley, B., Mounce, S., and Boxall, J. Development and Field Validation of a Burst Localization Methodology. *Journal of Water Resources Planning and Management*, 139(6):604–613, 2013.

Ferrante, M., Brunone, B., and Meniconi, S. Wavelets for the analysis of transient pressure signal for leak detection. *Journal of Hydraulic Engineering*, 133(11):1274–1282, 2007.

Ferrante, M., Brunone, B., and Meniconi, S. Leak detection in branched pipe systems coupling wavelet analysis and a Lagrangian model. *Journal of Water Supply: Research and Technology - AQUA*, 58(2):95–106, 2009. ISSN 00037214. doi: 10.2166/aqua.2009.022.

Friedman, M., Radder, L., Harrison, S., Howie, D., and Britton, M. Verification and Control of Pressure Transients and Intrusion in Distribution Systems. Technical report, AWWA Research Foundation, 2005.

Gartenhaus, S. *Physics Basic principles*. Holt, Rinehart and Winston, 1977.

Ghorbanian, V., Karney, B., and Guo, Y. Pressure standards in water distribution systems: reflection on current practice with consideration of some unresolved issues. *Journal of Water Resources Planning and Management*, pages 1–8, 2016.

Giustolisi, O. and Doglioni, A. Water distribution system failure analysis. In *Proceeding of the 8th International Conference on Computing and Control for the Water Industry*, number 1, pages 51–56, 2005.

Giustolisi, O. and Savic, D. A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics*, 8(3):207–222, 2006. ISSN 1464-7141.

Giustolisi, O. and Savic, D. Advances in data-driven analyses and modelling using EPR-MOGA. *Journal of Hydroinformatics*, 11(3-4):225–236, 2009. ISSN 1464-7141.

Giustolisi, O., Savic, D., and Kapelan, Z. Pressure-Driven Demand and Leakage Simulation for Water Distribution Networks. *Journal of Hydraulic Engineering*, 134(5):626–635, may 2008.

Gomes, R., Sá Marques, A., and Sousa, J. Estimation of the benefits yielded by pressure management in water distribution systems. *Urban Water Journal*, 8(2):65–77, 2011. ISSN 1573-062X. doi: 10.1080/1573062X.2010.542820.

Goulter, I. and Kazemi, A. Spatial and temporal groupings of water main pipe breakage in Winnipeg. *Canadian Journal of Civil Engineering*, 15:91–97, 1988.

Grenander, U. and Rosenblatt, M. *Statistical analysis of stationary time series*. 1984. ISBN 0828403201.

Gullick, R., LeChevalier, M., Svindland, R., and Friedman, M. Occurence of transient low and negative pressures in distribution systems. *Journal of American Water Works Association*, 96 (11):52–66, 2004.

Guo, X., Yang, K., Li, F., Wang, T., and Fu, H. Analysis of first transient pressure oscillation for leak detection in a single pipeline. *Journal of Hydrodynamics, Ser. B*, 24(3):363–370, jul 2012. ISSN 10016058.

Habibian, A. Effect of Temperature Changes on Water-Main Breaks. *Journal of Transportation Engineering*, 120(2):312–321, 1994.

Hager, W. Swiss contribution to water hammer theory. *Journal of Hydraulic Research*, 39(1): 37–41, 2001.

Hampson, W., Collins, R., Beck, S., and Boxall, J. Transient Source Localisation Methodology and Laboratory Validation. *Proceeding of the 12th International Conference of Computing and Contol for the Water Industry, CCWI. Perugia*, 00, 2013.

Harmer, K., Howells, G., Sheng, W., Fairhurst, M., and Deravi, F. A Peak-Trough Detection Algorithm Based on Momentum. *2008 Congress on Image and Signal Processing*, pages 454–458, 2008.

HMSO. Water Industry Act. 1991(2), 1991.

Hudak, P., Sadler, B., and Hunter, B. Analyzing underground water-pipe breaks in residual soils. *Water Engineering and Management*, 145(12):15–20, 1998.

Hyndman, R. and Koehler, A. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, oct 2006. ISSN 01692070.

Jaeger, C. Water hammer effects in power conduits. *Civil Engineering and Public Works Review*, 43(23):74–76, 138–140, 192–194, 244–246, 1948.

Jalil, B., Beya, O., Fauvet, E., and Laligant, O. Singularity detection by wavelet approach: application to electrocardiogram signal. *Processing of SPIE-IS&T Electronic Imaging*, 7535: 753506–753506–8, 2010. URL `3506/s1{&}Agg=doi`.

Jung, B., Karney, B., Boulos, P., and Wood, D. The need for comprehensive transient analysis of distribution systems. *Journal of American Water Works Association*, 99(1):112–123, 2007.

Kakoudakis, K., Behzadian, K., Farmani, R., and Butler, D. Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K -means clustering. *Urban Water Journal*, 9006(11):1–6, 2016.

Kaliatka, A., Uspuras, E., and Vaisnoras, M. Uncertainty and sensitivity analysis of parameters affecting water hammer pressure wave behaviour. *Kertechnik - Carl Hanser Verlag*, 71(5-6): 270–278, 2006.

Karim, M., Abbaszadegan, M., and Lechavellier, M. Pathogen intrusion into potable water distribution. *Abstracts of the General Meeting of the American Society for Microbiology*, 100: 561, 2000.

Karim, M., Abbaszadegan, M., and LeChevallier, M. Potential for pathogen intrusion during pressure transients. *Journal of American Water Works Association*, 95(5):134–146, 2003.

Karney, B. W. and McInnis, D. Transient analysis of water distribution systems. *Journal American Water Works Association*, 82(7):62–70, 1990.

Karney, B. and Mclnnis, D. Efficient calculation of transient flow in simple pipe networks. *Journal of Hydraulic Engineering*, 118(7):1014–1030, 1992.

Kennedy, C. and Turley, J. Time Series Analysis As Input for Predictive Modeling : Predicting Cardiac Arrest in a Pediatric Intensive Care Unit. *Theoretical Biology and Medical Modelling*, 8(40):1–25, 201.

Kettler, A. and Goulter, I. An analysis of pipe breakage in urban water distribution networks. *Canadian Journal of Civil Engineering*, 12(1):286–293, 1985.

Kirmeyer, G., Friedman, M., Martel, K., Howie, D., LeChevallier, M., Abbaszadegan, M., and Karim, M. Pathogen intrusion into the distribution system. Technical report, American Water Works Association, 2001.

Kleiner, Y. and Rajani, B. Considering Time-dependant Factors in the Statistical Prediction of Water Main Breaks. Technical report, Institute of Research in Construction, Ottawa, Canada, 2000.

Kleiner, Y. and Rajani, B. Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water*, 3(3):131–150, sep 2001. ISSN 14620758.

Kleiner, Y. and Rajani, B. Forecasting variations and trends in water-main breaks. *Journal of Infrastructure Systems*, 8(4):122–131, 2002.

Kohonen, T. Self-Organised formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

Kroon, J., Stoner, M., and Hunt, W. Water Hammer: Causes and Effects. *Journal of American Water Works Association*, 76(11):39–45, 1984.

Kumar, A., Meronyk, E., and Segan, E. Development of concepts for corrosion assessment and evaluation of underground pipelines. Technical report, 1984.

Kwon, H. and Lee, C. Probability of pipe breakage regarding transient flow in a small pipe network. *Annals of Nuclear Energy*, 38(2-3):558–563, feb 2011. ISSN 03064549.

Lambert, A. Losses from Water Supply Systems: Standard Terminology and Recommended Performance Measures. *IWA the blue pages*, (10):1–13, 2000.

Lambert, A. A realistic basis for an international comparison of real losses from public water supply systems. *The institution of Civil Engineers Conf., Water Environment 98-Maintaining the Flow*, 1998.

Lambert, A. International report. Technical Report October 2001, 2002.

Laucelli, D., Berardi, L., and Doglioni, A. Evolutionary polynomial regression toolbox: version 1.0, 2005. URL `http://www.hydroinformatics.it/`.

Laucelli, D., Rajani, B., Kleiner, Y., and Giustolisi, O. Study on relationships between climate-related covariates and pipe bursts using evolutionary-based modelling. *Journal of Hydroinformatics*, 16(4):743, 2014. ISSN 1464-7141.

Le Gat, Y. and Eisenbeis, P. Using maintenance records to forecast failures in water networks. *Urban Water*, 2(3):173–181, sep 2000. ISSN 14620758. doi: 10.1016/S1462-0758(00)00057-1.

Lei, J. and Saegrov, S. Statistical approach for describing failures and lifetimes of water mains.pdf. *Water Science & Technology*, 38(6):209–217, 1998.

Lemaitre, J. How to use damage mechanics. *Nuclear Engineering and Design*, 80(2):233–245, 1984.

Lewicki, P. and Hill, T. Statistics: methods and applications. *Tulsa, OK. Statsoft*, 2006.

Lighthill, J. *Waves in Fluids*. Cambridge University Press, 1978.

Little, R. and Jebe, E. *Statistical design of fatigue experiments*. Applied Science Publishers Ltd, Essex, UK, 1975.

Liu, B., Sera, Y., Matsubara, N., Otsuka, K., and Terabe, S. Signal denoising and baseline correction by discrete wavelet transform for microchip capillary electrophoresis. *Electrophoresis*, 24:3260–3265, 2003. ISSN 01730835. doi: 10.1002/elps.200305548.

Liu, Z., Yu, X., Tao, J., and Sun, Y. Multiphysics extension to physically based analyses of pipes with emphasis on frost actions. *Journal of Zhejiang University-Science A*, 13(11):877–887, 2012. ISSN 1673-565X.

Ljung, L. System Identification: Theory for User, 1999. ISSN 00051098.

Massey, B. *Mechanics of Fluid*. Taylor and Francis, 2006. ISBN 0203413520.

Matsuishi, M. and Endo, T. Fatigue of metals subjected to varying stress. *Japan Society of Mechanical Engineers, Fukuoka*, 1968.

McInnis, D. and Karney, B. Transients in distribution networks: Field tests and demand models. *Journal of Hydraulic Engineering*, 121(3):218–231, 1995.

Misiunas, D., Vítkovský, J., and Olsson, G. Pipeline burst detection and location using a continuous monitoring technique. *Advances in Water Supply*, pages 89–96, 2003.

Misiunas, D., Lambert, M., Simpson, A., and Olsson, G. Burst detection and location in water distribution networks. *Water Science & Technology*, 5(3-4):71–80, 2004.

Misiunas, D., Vítkovský, J., Olsson, G., Simpson, A., and Lambert, M. Pipeline Break Detection Using Pressure Transient Monitoring. *Journal of Water Resources Planning and Management - ASCE*, 131(4):316–325, 2005.

Morris, R. Principal causes and remedies of water main breaks. *Journal of American Water Works Association*, 59(7):782–798, 1967.

Morrison, J. Managing leakage by District Metered Areas: a practical approach. *Water*, 21(2): 44–46, 2004.

Mounce, S., Boxall, J., and Machell, J. Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. *Journal of Water Resources Planning and Management*, 136(3):309–318, 2010. ISSN 0733-9496.

Mounce, S., Sharpe, R., Speight, V., Holden, B., and Boxall, J. Knowledge discovery from large disparate corporate databases using self-organising maps to help ensure supply of high quality potable water. *11th International Conference on Hydroinformatics, HIC2014*, 2014.

Mounce, S., Husband, S., Furnass, W., and Boxall, J. B. Multivariate data mining for estimating the rate of discoloration material accumulation in drinking water systems. *Journal of Hydroinformatics*, 18(1):96–114, 2016. ISSN 18777058.

National Soil Resources Institute, U. Soil Site Report. Technical Report 1, 2010.

O'Day, D. Organizing and analyzing leak and break data for making main replacement decisions. *Journal of American Water Works Association*, 74(11):589–596, 1983.

O'Day, D. External corrosion in distribution systems. *Journal of American Water Works Association*, 81(10):45–52, 1989. ISSN 0003150X.

Ofwat. Updating the overall performance assessment (OPA) - Conclusions and methodology for 2004-05 onwards March 2004, 2004.

Ofwat. The guaranteed standards scheme (GSS). (4):1–12, 2008.

Olhede, S. and Walden, A. Noise reduction in quadrature doppler ultrasound using multiple Morse wavelets. *IEEE Transactions on Biomedical Engineering*, 50(1):51–57, 2003.

Pearn, W. and Laboratories, B. Assessing the statistical characteristics mean absolute error or forecasting. *International Journal of Forecasting*, 7(3):335–337, 1991.

Placidi, G., Alecci, M., and Sotgiu, A. Post-processing noise removal algorithm for magnetic resonance imaging based on edge detection and wavelet analysis. *Physics in medicine and biology*, 48:1987–1995, 2003. ISSN 0031-9155.

Pugsley, A. *The safety of structures*. The Pitman Press, Bath, Bath, edward arn edition, 1966.

Rachid, F. and Mattos, H. Model for structural failure of elasto-viscoplastic pipelines. *Meccanica*, 29(3):293–304, sep 1994. ISSN 0025-6455.

Rachid, F. and Mattos, H. Modelling the damage induced by pressure transients in elasto-plastic pipes. *Meccanica*, 33:139–160, 1998.

Rachid, F., Gama, R., and Mattos, H. Modelling of hydraulic transient in damageable elasto-viscoplastic piping systems. *Applied mathematical modelling mathematical modelling*, 18(4):207–215, 1994.

Rajani, B. and Kleiner, Y. Comprehensive review of structural deterioration of water mains: physically based models. *Urban Water*, 3(3):151–164, sep 2001. ISSN 14620758. doi: 10.1016/S1462-0758(01)00032-2.

Rajani, B. and Makar, J. A methodology to estimate remaining service life of grey cast iron water mains. *Canadian Journal of Civil Engineering*, 27(6):1259–1272, 2000. ISSN 0315-1468.

Rajani, B. and Tesfamariam, S. Estimating time to failure of ageing cast iron water mains under uncertainties. In *Proceeding of the 8th International Conference on Computing and Control for the Water Industry*, pages 57–64, 2005.

Rajani, B. and Zhan, C. On the estimation of frost loads. *Canadian geotechnical journal*, 33(4): 629–641, 1996.

Rajani, B., Zhan, C., and Kuraoka, S. Pipe soil interaction analysis of jointed water mains. *Canadian Geotechnical Journal*, 33:393–404, 1996.

Ramalingam, D., Lingireddy, S., and Wood, D. Using the WCM for transient modelling for water distribution networks. *American Water Works Association*, (2):1–15, 2009.

Ramos, H., Covas, D., Borga, A., and Loureiro, D. Surge damping analysis in pipe systems : modelling and experiments. *Jurnal of Hydraulic Research*, 42(4):413–425, 2004a.

Ramos, H., Covas, D., Borga, A., and Loureiro, D. Surge damping analysis in pipe systems: modelling and experiments. *Journal of Hydraulic Research*, 42(4):413–425, 2004b.

Røstum, J. *Statistical Modelling of Pipe Failures in Water Networks*. PhD thesis, 2000.

Royer, M. White Paper on Improvement of Structural Integrity Monitoring for Drinking Water Mains. Technical Report March, EPA, National Risk Management Research Laboratory, Cincinnati, Ohio 45268, 2005.

Sadiq, R., Rajani, B., and Kleiner, Y. Fuzzy-based method to evaluate soil corrosivity for prediction of water Main deterioration. *Journal of Infrastructure Systems*, 10(12):149–156, 2004. ISSN 1076-0342.

Sargaonkar, A., Kamble, S., and Rao, R. Model study for rehabilitation planning of water supply network. *Computers, Environment and Urban Systems*, 39(5):172–181, may 2013. ISSN 01989715. doi: 10.1016/j.compenvurbsys.2012.08.002.

Savic, D. and Banyard, J. *Water distribution systems*. ICE Publishing, London, 2011. ISBN 978-0-7277-4112-7.

Savić, D., Giustolisi, O., Berardi, L., Shepherd, W., Djordjević, S., and Saul, A. Modelling sewer failure by evolutionary computing. *Proceedings of the ICE - Water Management*, 159(June): 111–118, 2006. ISSN 1741-7589.

Savic, D., Giustolisi, O., and Laucelli, D. Asset deterioration analysis using multi-utility data and multi-objective data mining. *Journal of Hydroinformatics*, 11(3–4):211, 2009.

Seica, M. V., Packer, J. A., and Asce, F. Mechanical Properties and Strength of Aged Cast Iron Water Pipes. *Journal of Materials in Civil Engineering*, 16(1):69–77, 2004.

Shamir, U. and Howard, C. An analytic approach to scheduling pipe replacement. *Journal of American Water Works Association*, 71(7):248–158, 1979.

Skipworth, P., Cashman, A., Saul, A., Walters, G., Savic, D., and Engelhardt, M. Whole life costing for water distribution network management, 2002.

Smith, S. The scientist and engineer's guide to digital signal processing. pages 1–10, 1997.

Soares, A., Covas, D., and Ramos, H. Damping analysis of hydraulic transients in pump-rising main systems. *Journal of Hydraulic Engineering*, 139(2):233–243, 2013.

Srirangarajan, S., Iqbal, M., Lim, H., Allen, M., Preis, A., and Whittle, A. Water main burst event detection and localization. *Proceedings of the 12th Water Distribution Systems Analysis Conference*, (10), dec 2010.

Srirangarajan, S., Allen, M., Preis, A., Iqbal, M., Lim, H., and Whittle, A. Wavelet-based Burst Event Detection and Localization in Water Distribution Systems. *Journal of Signal Processing Systems*, 72(1):1–16, sep 2012. ISSN 1939-8018.

Stacha, J. Criteria for Pipeline Replacement. *Journal of American Water Works Association*, 70 (5):256–258, 1978.

Starczewska, D., Collins, R., and Boxall, J. Occurence of transients in water distribution networks. *Proceeding of the 13th International Conference of Computing and Contol for the Water Industry, CCWI. Leicester*, 00:1–11, 2015.

Stephens, M., Lambert, M., Simpson, A., and Vitkovsky, J. Calibrating the water-hammer response of a field pipe network by using a mechanical damping model. *Journal of Hydraulic Engineering*, 137(10):1225–1237, 2011.

Stoianov, I. and Nachman, L. PIPENET: A wireless sensor network for pipeline monitoring. *Proceedings of the Sixth International Symposium on Information Processing in Sensor Networks*, (4):264–273, apr 2007.

Stoianov, I., Nachman, L., Whittle, A., Madden, S., and Kling, R. Sensor networks for monitoring water supply and sewer systems: lessons from Boston. *Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium*, pages 100–100, 2006.

Streeter, V. and Wylie, E. *Hydraulic transient*. McGraw-Hill Inc., New York, 1967.

Tesfamariam, S. and Rajani, B. Estimating time to failure of cast-iron water mains. *Proceedings of the ICE - Water Management*, 160(2):83–88, jan 2007. ISSN 1741-7589. doi: 10.1680/wama.2007.160.2.83. URL `http://www.icevirtuallibrary.com/content/article/10.1680/wama.2007.160.2.83`.

Tesfamariam, S., Rajani, B., and Sadiq, R. Possibilistic approach for consideration of uncertainties to estimate structural capacity of ageing cast iron water mains. 1064(December):1050–1064, 2006. doi: 10.1139/L06-042.

Thomson, J. and Wang, L. State of Technology Review Report on Condition Assessment of Ferrous Water Transmission and Distribution Systems. Technical Report June, 2009.

Thorley, A. *Fluid Transients in Pipeline Systems*. D&L George Ltd, Herts, England, 1991.

Thornton, J. Managing leakage by managing pressure: a practical approach. *Water 21*, 10:1–2, 2003. ISSN 15619508.

Tijsseling, A. and Heinsbroek, A. The influence of bend motion on waterhammer pressures and pipe stresses. *Proceeding of the 3rd ASME/JSME Joint Fluids Engineeting Conference*, pages 1–7, 1999.

Ulanicki, B., AbdelMeguid, H., Bounds, P., and Patel, R. Pressure control in district metering areas with boundary and internal pressure reducing valves. *Proceedings of the 10th Annual Water Distribution System Analysis Conference, WDSE2008*, pages 691–703, 2008. ISSN 08950563.

Walski, T. and Pelliccia, A. Economic analysis of water main breaks. *Journal American Water Works Association*, 73(3):140–147, 1982.

Walski, T. and Lutes, T. Hydraulic transients cause low-pressure problems. *Journal of American Water Works Association*, 86(12):24–32, 1994.

Walski, T., Wade, R., Sharp, R., Sjostrom, J., and Schlessinger, D. Conducting a pipe break analysis for a large city. *AWWA Conference Symposium*, pages 387–402, 1986.

Watson, T. and Mason, A. A hierarchical Bayesian model for the maintenance of water pipe networks. *Proceedings of the 7th International Conference on Hydroinformatics, HIC2006, Nice, Franc*, pages 1–8, 2006. URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.75.8181{&}rep=rep1{&}type=pdf`.

Wehrens, R. and Buydens, L. Self- and Super-organizing Maps in R: The kohonen package. *Journal of statistical software*, 21(5), 2007.

Wood, D., Dorsch, R., and Ughtner, C. Wave-plan analysis of unsteady flow in closed conduits. *Journal of the Hydraulics Division*, 92(2):83–110, 1966.

Wylie, E. *Resonance in pressurised piping system*. PhD thesis, University of Michigan, 1965.

Wylie, E. and Streeter, V. *Fluid transients*. McGraw-Hill Inc., 1978.

Wylie, E. and Streeter, V. *Fluid transients in systems*. Prentice Hall, Englewood Cliffs, USA, 1993.

Yamijala, S., Guikema, S., and Brumbelow, K. Statistical models for the analysis of water distribution system pipe break data. *Reliability Engineering & System Safety*, 94(2):282–293, feb 2009. ISSN 09518320.

Young, P., Jakeman, A., and R., M. An instrumental variable method for model order identification. *Automatica*, 16(3):281–294, 1980.

Zhan, C. and Rajani, B. Estimation of frost load in a trench: theory and experiment. *Canadian geotechnical journal*, 34(4):568–579, aug 1997. ISSN 0008-3674.