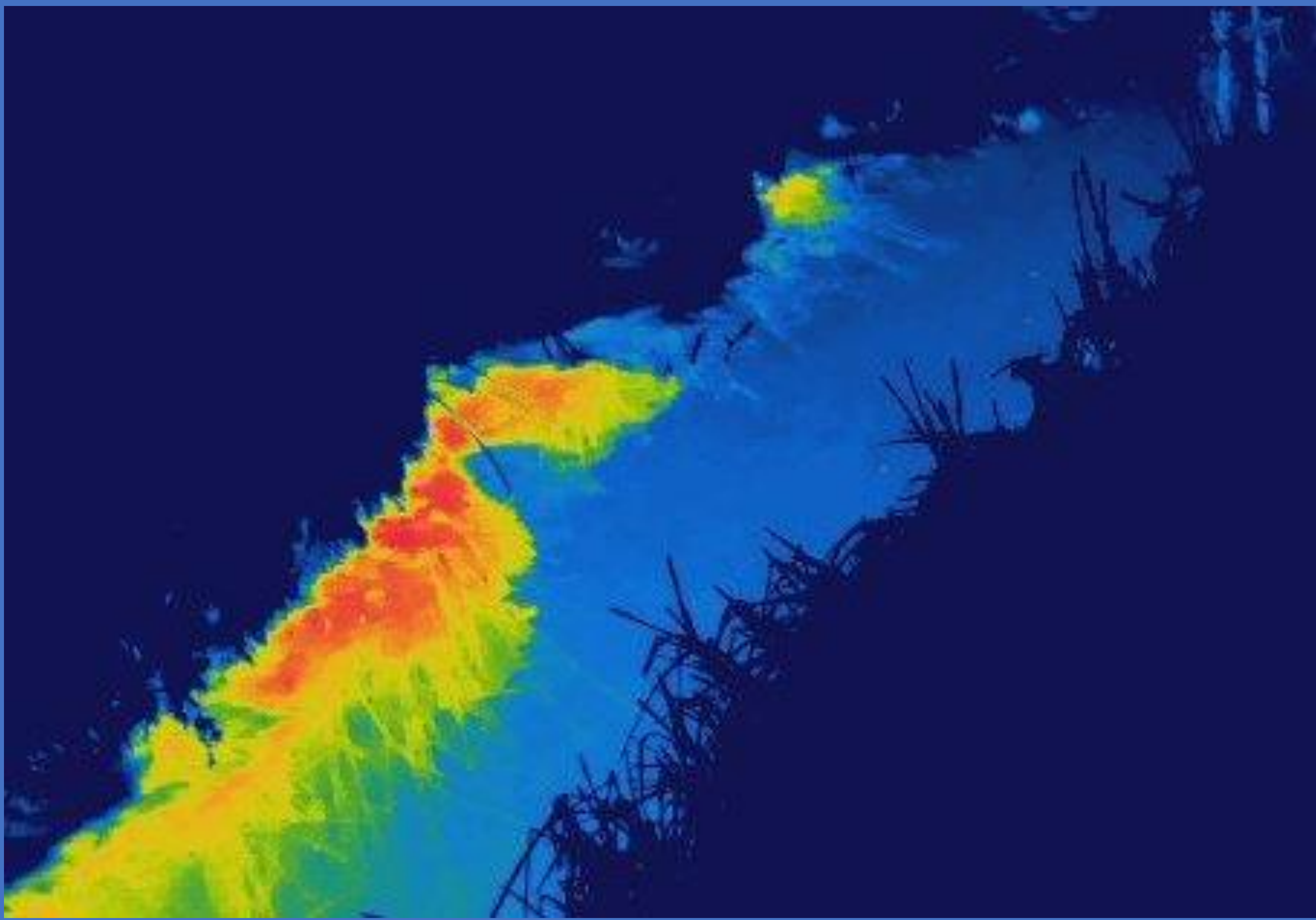


Quantifying thermal imaging data for piping detection

Analyzing the flow rate and the size of the temperature spots in thermal images of seepage at dykes

I.C. van Klaarbergen

Master Thesis

July 2022

BZ Ingenieurs & Managers

UNIVERSITY OF TWENTE.

Cover picture by USGS: Martin A. Briggs (2013)
via *https://www.usgs.gov/media/images/thermal-image-groundwater-discharging-stream*

Quantifying thermal imaging data for piping detection

Analyzing the flow rate and the size of the temperature spots in thermal images of seepage at dykes

Master Thesis
University of Twente
Department of Water Engineering and Management

Author:
I.C. (Iris) van Klaarbergen

Guided by:
dr. J.J. (Jord) Warmink — Head of the committee UT
dr. ing. R. (Rolands) Kromanis — Daily supervisor UT
ing. W.S. (Wouter) Zomer MSc. — Supervisor BZIM
ir. S. (Sander) Steenblik — Daily supervisor BZIM

Preface

This thesis is the finishing piece of my master Civil Engineering and Management at the University of Twente. During the six years of studying in Enschede, my interest has grown towards the Water Engineering and Management side of Civil Engineering. I was very happy that the subject of this master thesis gave me the opportunity to write a thesis about a subject that interests me a lot. During my internship at BZ Ingenieurs & Managers I got the opportunity to learn about many projects that were related to Water Engineering and Management, giving me a better view on the possibilities of this profession.

First of all, I would like to thank all colleagues at BZ Ingenieurs & Managers for being so enthusiastic about their and my work. Especially I would like to thank my daily supervisor Sander Steenblik for the weekly meetings, the feedback sessions and helping me move forward in tough parts of the research. Thank you for always being available for my questions. Furthermore, I would like to thank my supervisor Wouter Zomer, for the enthusiasm about the project, the feedback on my work and expanding my knowledge on the subject.

I would also like to thank my supervisors at the University of Twente, being Jord Warmink and Rolands Kromanis. Jord, thank you for the clarifying feedback sessions and giving me insight on how to improve this research. Rolands, thank you for all the help with analyzing the thermal imaging data in MATLAB, improving my report and improving my English. I got a lot out of our sessions and I am very grateful that I could come to you with all my questions.

Lastly, I would like to thank my family and friends. Thank you for all the support during my study time in Enschede, especially during this thesis, and the faith you have in me. I appreciate all of you very much and am grateful to have you in my life.

I hope you will enjoy reading this thesis report.

Iris van Klaarbergen

Enschede, July 2022

Summary

The failure mechanism piping is a significant problem for the safety of river and sea dykes located in delta areas. This is the case in large parts of the Netherlands. Piping is a process that can reach an advanced stage before any sign is remarkably visible. Therefore, it is important to monitor the dykes to have insight in the current conditions with respect to piping. The measurement of temperature is proven to be an effective tool for the detection of seepage spots which are possible piping locations in dykes. With the use of a camera with infra-red sensors, the surface temperature of structures can be measured, resulting in a thermal image. With thermal imaging, potential problematic seepage locations can be mapped very fast and efficiently since seepage often has a different temperature than its surrounding. With human inspection they are sometimes missed due to a coverage of grass or water plants.

The application of thermal imaging for management and inspection of flood defenses is still relatively limited. The technique is used to detect locations of seepage spots in the field but more than the location is not derived yet. The quantification of thermal imaging data for piping detection is a subject that needs to be investigated. In this study, the sizes of the seepage spots in the thermal images and the relation between these spots and the flow rate through the sand boils is quantified. The aim is to determine to what extent thermal imaging can be used to make a flow rate prediction of a seepage spot. The case study which is used for this research is the full-scale piping experiment in a controlled field environment at the Living Lab Hedwige-ProsperPolder (LLHPP) which was performed in September of 2021.

To determine to what extent thermal imaging can be used to make a flow rate prediction of a seepage spot, three main steps are taken. Firstly, a data analysis method is developed to quantify the sizes of the areas of the found seepage spots in the thermal images. Secondly, a correlation analysis is performed to find out if there is a relation between the areas of the seepage spots in the thermal images and the flow rate of the seepage through the sand boils. Lastly, a regression analysis is performed to develop a practical tool to enable flow rate prediction based on the size of the seepage spot in the thermal images.

For a large part of the thermal imaging data it was possible to quantify the seepage spots. The best circumstances to quantify the seepage spots in the thermal images are found when the influence of the sun on the surface temperature of the ditch is not significant. So, during night or when the weather is not too sunny.

The found correlation showed promising results. The correlating part of the data had little influence of sunlight, there were no people in the camera view and one sand boil was mainly active. This part showed a strong correlation between the flow rate and the size of the seepage spot in the thermal image which can be used for a regression analysis.

The regression analysis resulted in a practical tool which enables flow rate prediction based on the size of a seepage spot in a thermal image. The regression analysis showed that the flow rate increases slower at areas which are bigger, and faster at areas which are smaller.

The results of this study can be implemented in practice as an additional tool to complement the inspection of dykes for piping. It makes human observations more accurate, supports the interpretation of what is seen and can find critical seepage locations which human observations can miss upon.

In conclusion, the results that are found in this study show that thermal imaging can be used to make a prediction of the flow rate of a seepage spot under certain circumstances. It is a very promising method to be used as a complementary tool to enhance the quality of the dyke assessment.

Contents

List of abbreviations

LLHPP . . . Living Lab Hedwige-ProsperPolder
ROI . . . Region of Interest
SMH . . . Structural Health Monitoring
TI . . . Thermal Imaging
RGB . . . Red Green Blue

List of Figures

List of Tables

1 | Introduction

The failure mechanism piping is a significant problem for the safety of river and sea dykes located in delta areas (Bersan et al., 2018). This is the case in large parts of the Netherlands. Dykes protecting the land against tides and storm surges were built since the 13th century (Inman, 2010).The country has reclaimed lots of land from the sea as well. The reclamation and consolidation of the land and sea level rise results in one-third of the area of the Netherlands being below the mean sea level. Flood defence structures such as dykes and dunes protect the country from flooding. Without them 65% of the Netherlands would be flooded during high tides (Hoeksema, 2007). Since the 1990s, due to global warming, the rise of the mean sea level has increased, thus threatening coastal areas of being flooded. If no future adaptation measures are made to the coastal defence structures, the annual damage costs due to flooding are estimated to be 20.7 billion euros (Vousdoukas et al., 2020).

Dykes are subject to multiple environmental and traffic-induced loads and extreme events, such as flooding. Therefore they age and deteriorate. Weakened links are vulnerable and may fail. For example multiple dyke breaches happened during the flood disaster in 1953 (van der Ham, 2006). There are multiple failure mechanisms that can cause these dyke breaches. The most dominant mechanisms are overtopping, surface erosion, macro-instability (inward and outward), micro-instability, piping and (in)stability of the dyke cover (t Hart, 2018). All of these failure mechanisms are taken into account when a dyke assessment is done and can be monitored in different ways. This research focuses specifically on the failure mechanism piping.

1.1 Problem context

Piping in dykes is a process that can reach an advanced stage before any sign is remarkably visible (Sjödahl et al., 2009). Therefore it is important to perform a condition assessment of dykes, however it is challenging. Visual inspections are difficult considering the extended length of dykes. They are also subjective to an inspector's interpretation. Dyke structural monitoring can offer objective and accurate insights into conditions of these extended flood defence structures (Mes and Reichart, 2015). Monitoring aims assessing and forecasting current conditions of a dyke by detecting an onset of anomalous behaviour which could be related to the loss of a structures strength. The application of dyke monitoring can improve dyke management and save costs by having more insight in the current conditions of the dyke (Vermeer et al., 2012).

Currently many new sensor techniques for dyke monitoring are becoming available, such as deformation measurement with satellite data, ground radar, 3D-scanning and temperature measurement. When these techniques are applied in a smart way, it can give reliable information on the current conditions of dyke sections (Mes and Reichart, 2015). However, these kind of monitoring techniques are not always the most efficient and are not always useful under every condition. Still, the monitoring techniques can be very useful in some situations, giving additional information to an inspector's interpretation (Mes and Reichart, 2015). One of these new sensor techniques for dyke monitoring, temperature measurement, is able to detect behaviour which can indicate piping. It is an example of a non-destructive monitoring method. Temperature can be measured in different ways. It is most commonly measured with the use of fibre optic cables or with the use of an infra-red camera. The measurement of temperature is the monitoring method where this research is focused on.

The measurement of temperature is used to detect indications of piping in other studies already. Bersan et al. (2018) and Bersan et al. (2019) made use of temperature measurement with fibre optic cables underneath the dyke, which is between the aquifer (sand) and the impermeable layer (clay). During this experiment they found that a sensor located circa 1 meter behind the toe of the dyke could detect temperature changes caused by seepage, even while the process is still in an early stage (Bersan et al., 2019). This seepage can potentially become a sand boil which can result in piping. The same approach of temperature measurement with fibre optic cables was used in the study of de Vries et al. (2010). They found that a measurement near the top of the sand layer underneath the dyke is effective for detecting signs of failure by piping. So, both studies prove that a change in temperature indicates the presence of seepage.

Another technique to measure temperature is thermal imaging. With the use of a camera with infra-red sensors, the surface temperature of structures can be measured, resulting in a thermal image (Peeters et al., 2013). Zhou et al. (2022) conducted research about leakage detection in dykes with thermal imaging. It is shown that thermal imaging is a useful tool to detect leakage spots in dykes. Zhou et al. (2022) argues that the advantages of thermal imaging are the fast visualisation, strong mobility and wide coverage. Still, they recommend to conduct more research on the influences on the data by outside aspects, meaning that the temperature of the atmosphere or other weather aspects could have an influence on the measurements. This is in accordance with the study of Dong and Catbas (2021), who investigated how these adverse influence factors can influence the accuracy of the computer-visioned data. Chen et al. (2018) used thermal imaging on dykes to show overtopping failure, slope failure, progressive erosion failure or a combination of two or more of these failure mechanisms. In their experimental data, they noticed piping occurring. In the thermal images, a low temperature at the toe of the dyke downstream was visible in the data, as can be seen in Figure 1.1(b). The experiments show that low temperatures at the toe of the dyke indicate vulnerability to piping (Chen et al., 2018). So, under certain conditions leakage locations can be detected with the use of thermal imaging.

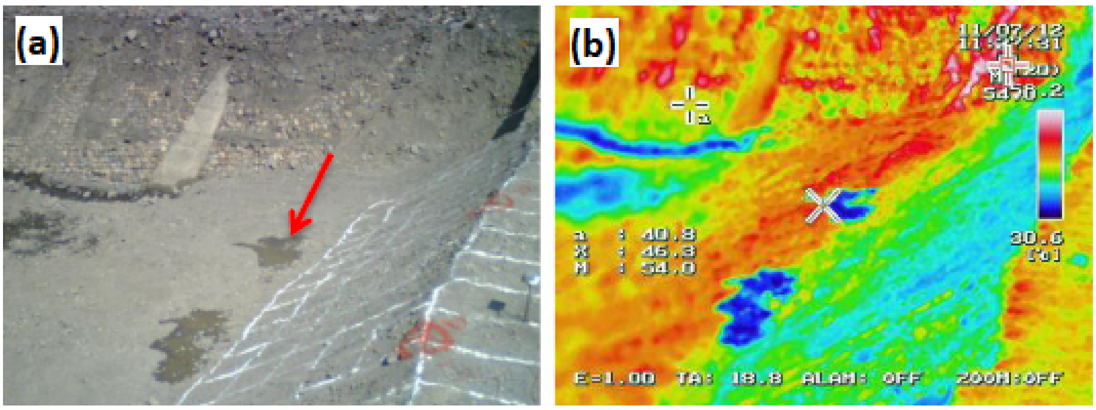


Figure 1.1: Base seepage downstream with (a) being a situation RGB image and (b) being a thermal image (Chen et al., 2018)

The study of Wiggers et al. (2019) addresses the technology of infra-red thermal imaging as well, where they found that this technique is very useful to detect the locations of seepage spots which are possibly sand boils. With thermal imaging, potential problematic seepage locations can be mapped very fast and efficiently, while with human inspection they sometimes would be missed due to a coverage of grass or water plants. An example of seepage visualised using thermal imaging can be seen in Figure 1.2. Wiggers et al. (2019) conclude that thermal imaging can be a good additional information source for further investigation of local weak spots (seepage locations) and as a tool for inspections and implementation control. So, the measurement of temperature has shown great potential of monitoring piping in past researches.

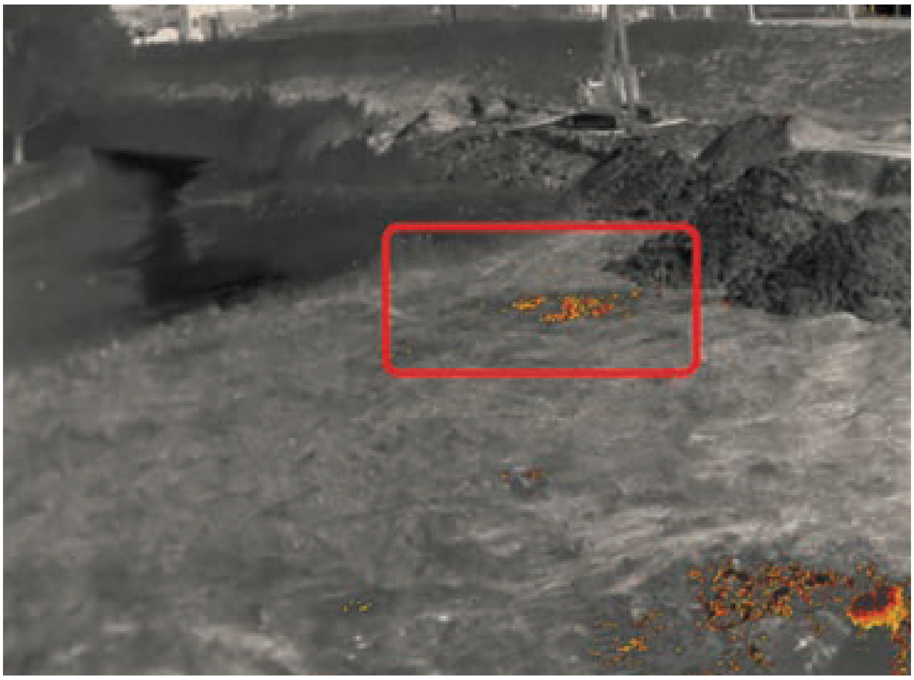


Figure 1.2: Seepage seen in a thermal image, the rectangle with rounded corners indicates one of the seepage spots where the colors indicate seepage spot by lighting up a certain temperature (Wiggers et al., 2019)

1.2 Problem statement

The application of thermal imaging for management and inspection of flood defences is still relatively limited: the technique is used to detect locations of seepage in the field but more than the location is not derived yet. It is known that a seepage spot or sand boil does not immediately have to indicate failure of the dyke due to piping. This is dependent on the amount of sand and water flowing out of this spot. The flow rate can give an indication of the size of the pipe (van Beek et al., 2013). So, when the information in thermal images can be connected to the outgoing flow rate of a seepage spot, this could indicate the severeness and development of the seepage spot over time in a very helpful way. This translation of thermal images to the development of current conditions for piping cannot be made explicit enough yet and is a subject that is needed to be investigated. As a result, this research adds on to the quantification of thermal imaging data for piping detection.

The research of van de Voort et al. (2021) made a first set-up of the possibilities of making this connection. They recommend to make an accurate estimation of the bandwidth of the flow rate of a sand boil captured with thermal imaging. As found, the flow rate of a seepage spot can give an indication of the current conditions of the dyke for piping. Therefore, a connection between the thermal image and the flow rate of the seepage spot could give an indication of the severeness of the situation going on within the dyke with respect to piping. This connection is investigated in this research.

1.3 Research aim and objectives

The aim of this study is to determine to what extent thermal imaging can be used to make a flow rate prediction of a seepage spot. Three objectives (O) are set to be able to reach the aim. The objectives are:

O1: To develop a data analysis method of the thermal images to:

- Extract useful thermal imaging data from the data set, being thermal images in which a clear temperature difference is visible due to the occurrence of seepage;
- Estimate the sizes of the areas of the seepage spots from these data.

O2: To quantify a possible correlation between the flow rate of the sand boil and the size of the seepage spot in the thermal images.

O3: To develop a practical tool to enable flow rate prediction based on the seepage spot size from the thermal images.

1.4 Thesis outline

Chapter 2 gives the theoretical framework of the study. Chapter 3 describes the Case study and Chapter 4 the Methodology. The results are given in Chapter 5 and a discussion of the results is given in Chapter 6. At last, the conclusion and recommendations are given in Chapter 7. Chapter 4 and 5 will follow the order of the research objectives.

2 | Theoretical Framework

In this chapter, some more background information is provided on the failure mechanism of piping, monitoring with thermal images and detection techniques for thermal images. At the end of the chapter, a summary of the takeaway points is given.

2.1 Piping

Piping is an internal erosion mechanism found in or underneath water-retaining structures such as dykes. In this process, hollow spaces, known as pipes, are formed as a result of water flowing through or underneath the structure. The mechanism consists of the following processes: seepage, backward erosion initiation, backward erosion progression, widening of the pipe, failure of the levee and breach (Semmens and Zhou, 2019) as can be seen in Figure 2.1.

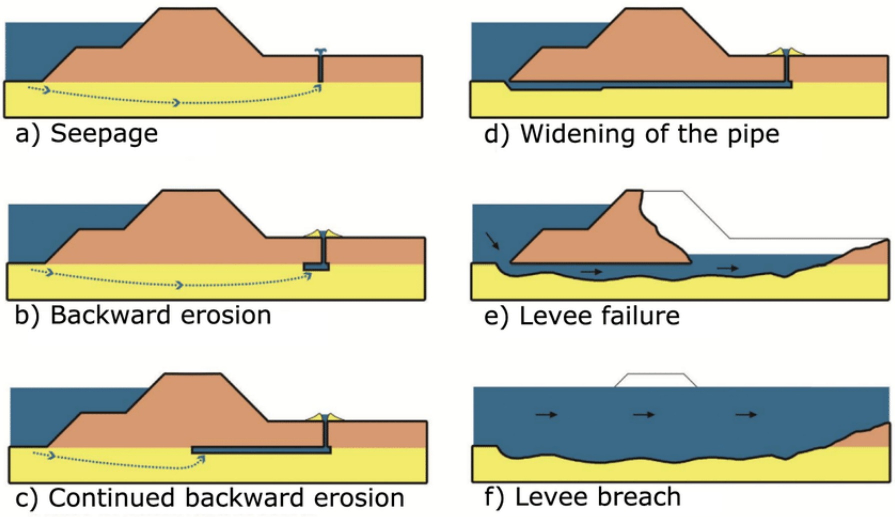


Figure 2.1: The phases of the piping process (Semmens and Zhou, 2019).

The first stage of the process is seepage. Seepage is caused by a water level difference across the dyke. An opening is required on the inner side of the dyke to let seepage lead into a sand boil and thus backward erosion. In some situations, an opening is present by for example a ditch, while in other cases, a local crack in the cover layer is needed (van Beek et al., 2013). The process where a crack in the covering layer is formed is a result of an uplift process. In the case of uplift, the water pressure under the covering layer becomes greater than the weight of the covering layer, causing it to push up and eventually crack open (Technical Advisory Committee, 1999).

The second stage is the initiation of backward erosion. This can happen once an opening in the cover layer in combination with seepage is present. The initiation of backward erosion requires the sand entering a fluid state near the opening in the cover layer. The fluid state of sand can be reached when the pore pressures in the sand become higher than the effective stresses in the sand. The seepage velocity increases the pore pressures in the sand. If these pore pressures are sufficient for the fluid state to be reached, it still does not immediately mean that pipe formation will happen (van Beek et al., 2013). In other words, the initiation of backward erosion does not necessarily lead to the progression of backward erosion. Due to the crack being filled with sand in the fluid state, the resistance in that crack will increase. Now there are two options: (1) the flow speed at the crack in the cover layer decelerates to such a degree that the erosion process stops due to the increased resistance, meaning that no pipe formation will happen, or (2) the flow speed does not decelerate enough, so the sand transport trough the crack will continue, meaning that a pipe will start to form. In the first option, only clean water is flowing out of the crack in the surface layer. In case of the second option, a fluid sand mixture will flow out of the crack in the surface layer and will be deposited around it, creating a sand boil (Technical Advisory Committee, 1999). The two possibilities are schematically displayed in Figure 2.2.

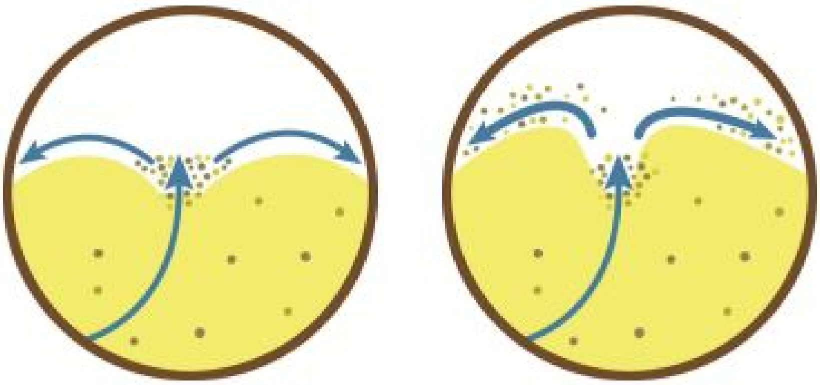


Figure 2.2: Left: a sand boil without sand deposition, the fluid flow is in equilibrium with the weight of the sand mass. Right: a sand boil with sand deposition, the flow velocity is high enough to deposit sand (van Beek, 2015).

The formation of a pipe is the third stage: continued backward erosion. In this stage, sand will keep flowing out of the sand boil and a pipe will grow starting at the sand boil and ending at the sea or river side of the dyke.

The fourth stage of the process is widening of the pipe. This process can start as soon as the pipe is connected from upstream to downstream of the dyke. Inside the pipe, the flow will increase, since the flow resistance of the sand will decrease. With this increase, more sand will be eroded and the pipe will be widened (van Beek et al., 2013).

When the process of widening progresses, the last stage of piping will happen, which is dyke breach. At this point, cracks and deformations can be seen in the dyke. There are two scenarios which can happen: sliding and settlement. In the case of sliding, a whole part of the dyke will be gone by erosion. In the case of settlement, the whole dyke will collapse, in which case the pipe will be closed (van Beek et al., 2013).

2.1.1 Early indications

Multiple studies are carried out to identify early indications of piping. Some find early indications which are not helpful, while others find helpful early indications. For example, Bersan et al. (2019) state that backward erosion only shows detectable deformations of a dyke when the structure is close to collapse. Therefore, they state monitoring of displacement is not helpful for finding early indications of piping. In contrast, pore water pressure can be a good early indication, but the effects seen in the pore water pressure are highly localised, meaning that a very high number of sensors in the dyke are needed (Bersan et al., 2019). Peeters et al. (2013) agree with these statements: piping is hard to identify using pore pressure sensors or inclinometers (used to measure deformation).

Talukdar and Dey (2019) search for early warnings of piping such that people can be warned at least 60 minutes before collapse of the dyke. In their study, they describe the initiation phase as being all stages happening before the continued backward erosion stage. They state that the initiation phase is the best phase to find warning occurrences: erosion through concentrated leaks is an early indication of possible piping, as well as cracking of the cover layer (Talukdar and Dey, 2019).

Erosion through concentrated leaks, known as sand boils, are an important indication of the possible start of backward erosion (Talukdar and Dey, 2019). The process of sand boils is seen in the first phases of piping, as described above. Two examples of a sand boil can be seen in Figure 2.3. However, when a sand boil is found, it does not immediately indicate that the process will result in failure. This is dependent on the amount of sand coming out of the sand boil (van Beek et al., 2013). Therefore, an even better early indication would be the sand flow rate of the sand boil, but the seepage flow rate indicates the size and thus state of the forming pipe as well. These results are similar to those reported by Sjödahl et al. (2009), who state that first indications are higher seepage rates, a visually observable concentrated leak, meaning a crack or a sand boil at the downstream side of the dyke, or high turbidity of the seepage water. After comparing all these studies, it can be concluded that sand boils are the most mentioned and best visible early indication of piping.

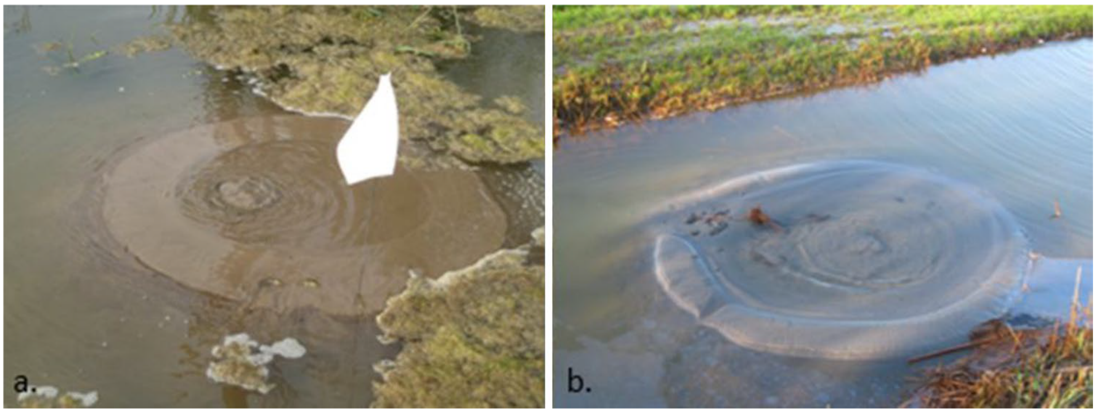


Figure 2.3: Two examples of a typical sand boil in the ditch on the outside of the river dyke. Locations: a. Mississippi River b. Waal River Netherlands. The size can range from millimeters to meters in diameter (Talukdar and Dey, 2019).

2.1.2 Temperature seepage water

Seepage through the aquifer underneath the dyke, can cause change in temperature of the seepage water (Rösingh et al. (2017); van de Voort et al. (2021)). Water infiltrating from the river or sea has a different temperature compared to the groundwater underneath the dyke. During the seepage process, seen in Figure 2.4, this water seeping through the soil can take up some of the temperature of the groundwater. The surface temperature of the water in the ditch at the toe of the dyke, or the surface temperature of the (grass) cover of the dyke toe, responds to the atmosphere temperature much faster than the groundwater temperature. As a result, when water is seeping through the dyke and coming up at the dyke toe, it often has a different temperature than the surrounding. This can be captured with an thermal image. However, a combination of circumstances can result in seeing no temperature difference.

If seepage can be seen by temperature differences at the dyke toe depends on (van de Voort et al., 2021):

- Seasonal variation of the atmospheric temperature, river temperature and groundwater temperature;
- Day- and night variation of the atmospheric temperature, river temperature and groundwater temperature;
- Geometrical characteristics of the sand boil (diameter, location in the ditch) and dynamic characteristics of the sand boil (flow rate, temperature seepage water);
- Geometrical characteristics of the ditch (depth, bathymetry) and dynamic characteristics of the ditch (flow rate of the ditch, temperature of the ditch).

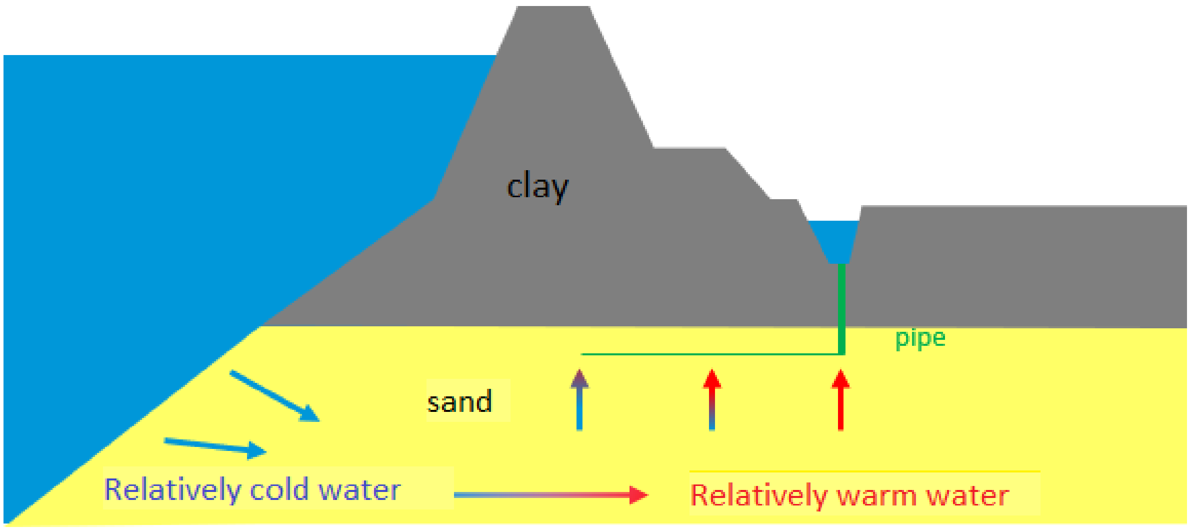


Figure 2.4: Temperature change of seepage water (van de Voort et al., 2021), an example of relatively cold infiltrating water which is changing to a higher temperature by the seepage process.

So season variation of the temperature of air, river water and groundwater, rain and day and night variations have an influence on the surface temperature of the water of the ditch. This can have a big influence on what can be seen in a thermal image, because it can cause the surface temperature to be one small range of temperatures. It can result in the water coming from the sand boil being the same temperature as the surface temperature. This can have a big influence, since the infrared cameras can only detect surface temperature differences (Donchyts and Koelewijn (2016); van de Voort et al. (2021)). Therefore, one should be aware of this while performing a data analysis of thermal images.
When the seepage spot is detectable, it is shown that in the summer, the seepage water will have a lower temperature than the water in the ditch, while in the winter, the seepage water will have a higher temperature than the water in the ditch. In fall and spring, temperature differences are harder to detect (van de Voort et al., 2021).

2.2 Monitoring with thermal imaging

Condition monitoring in the engineering perspective has the main purpose "to identify potential catastrophic failure with the intention to accurately schedule urgent maintenance activity so as to prevent operational interruptions" (Ma, 2007, p. 1). It is used in very many branches to understand processes that are going on or to track changes over time. The current condition of a structure can be assessed and with that the future condition can be possibly predicted (Ma, 2007). An example is structural health monitoring (SHM). SHM is the process of tracking the operational status of a structure, i.e. tracking if the structure is still able to do what it needs to do, assessing the condition, and detecting the damage (Dong and Catbas, 2021). In principle, this is a process that is used for the other branches as well. When zooming in on dykes, monitoring is often used to be able to say something about the risk of breaching by tracking change of certain parameters over time. The principle can then be used to detect failure mechanisms which are starting to develop.

When looking at processes that are relevant for piping, monitoring systems can be very helpful to identify zones where stronger seepage is seen. As mentioned before, these stronger seepage zones often represent weak spots where failure is more likely to occur (Bersan et al., 2018)). When monitoring systems are used as early warning tools, their effectiveness is based on how much in advance the monitoring systems can detect the start of failure compared to the time that is needed to take action to prevent the failure (Bersan et al., 2018).

Thermal imaging is an example of a computer-vision monitoring approach, which is an approach that aims to come up with computer systems that can perform and surpass detection of what the human visual system can detect (Huang, 1996). The computer vision technology for monitoring has a lot of advantages, such as the non-contact aspect, long distance possibility, fast result, the need of low labour and low costs. Since the object does not have to be touched, thermal imaging is a non-destructive surface detection method. A disadvantage is it being indirect measurements. Most often the measurement is not directly the parameter you are searching for such as the pore water pressure, flow rate, or other. A number of adverse influence factors can influence the accuracy of the computer-visioned data, since the measurement is from a distance (Donchyts and Koelewijn, 2016). One has to be aware of this fact when performing a data analysis (Dong and Catbas, 2021).

2.3 Summary

To summarize, the main takeaway points drawn from the literature and what they will be used for in this research are listed below.

- The process of piping starts with seepage. Seepage is required for the initiation of backward erosion. However, seepage on itself it is not a dangerous process. Once a seepage spot has evaluated in a sand boil, a high risk of piping occurs. This should be kept in mind when evaluating the results of this research.
- Clear early indications of piping are significantly higher seepage rates, the occurrence of sand boils and the flow rate of the sand boils. A relation between seepage rates/flow rates of sand boils and thermal images will be searched for in this research.
- The temperature of infiltrating seepage water changes on their way to the other side of the dyke and results in a temperature difference between the seepage water and the surroundings of the seepage spot. The visibility of temperature differences in thermal imaging can be highly influenced by rain, season variation of the temperature of air, river water, groundwater and day and night variations. In this research, a selection is made for thermal images in which temperature differences are clearly seen.
- Monitoring with thermal imaging has the advantage of being a non-destructive surface detection method. A disadvantage of this method is the possibility of the measurements being influenced by external factors. This will be taken into account when giving recommendations in this report.

3 | Case study

In September of 2021, a large-scale piping experiment in a controlled field environment was carried out at the Living Lab Hedwige-ProsperPolder (LLHPP) to investigate the influence of heterogeneous substrates in the emergence of piping. During this piping experiment, one of the used monitoring methods was infrared thermal imaging. The main goal of the measurement of these thermal images was to investigate the possibilities in supporting and interpreting the results of the piping experiments (Steenblik, 2021). In total a big set of data was gathered during the 10 day experiment. These data and the experimental setup of this experiments at the LLHPP are explained in detail in this chapter and are used as the case study of this research.

3.1 Experimental setup

The large-scale piping experiment at the LLHPP is divided in two experiments: one on the North side and one on the South side of the artificial dyke. In the first week, starting at September 6th, the first experiment was performed at the South side of the dyke. Starting at September 13th, a second experiment was performed at the North side of the dyke, known as the rerun-experiment. The dyke aquifer consists of tidal sand, with on top of that clay as the cohesive soil (Steenblik, 2021). The water flowing through the dyke is regulated with infiltration tubes. By increasing the water level in these infiltration tubes, the hydraulic head in the closed tidal sand package is increased until continued pipe growth is reached. These water levels are increased with steps of ±0.25 meter (or smaller at critical moments). The hydraulic head is regulated by holding that water level at every new step until a stationary situation is reached in which the water pressures and water flows became constant (Steenblik, 2021).

A picture of the artificial dyke at the experiment can be seen in Figure 3.1. For each side of the dyke, a ditch was created as the exit point for the pipes. These ditches are designed with the dimensions of 10 meter in length, 1 meter in width and 1.6 meter in depth, wherein a very small water level of 0.1 meter is maintained (Steenblik, 2021). This situation is shown in the schematic drawing in Figure 3.2. A sketch of the experimental set-up and additional pictures of the real experimental set-up can be seen in Appendix A in Figure A.1 and Figure A.2.

Figure 3.1: Experimental set-up at the LLHPP. The infiltration tubes can be seen on the top of the artificial dyke and the ditch of the North side of the dyke can be seen on the left side.

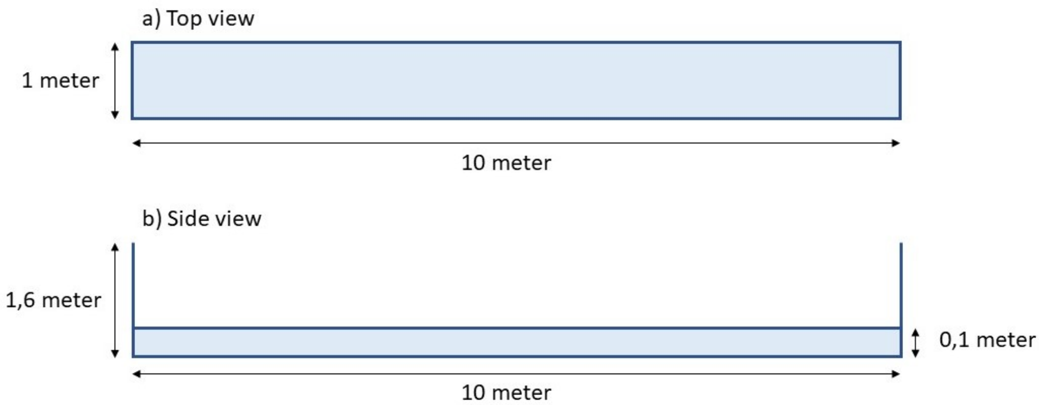


Figure 3.2: Schematic drawing of the dimensions of the ditches at the LLHPP.

3.2 Data

A large amount of data was gathered at the experiment. The first part of the experiment, at the South side of the dyke, started at September 6 8:00h and went on until September 7 11:00h, when the pipe reached the last row of water pore pressure gauges at 6,5 meters with respect to the sand boil in the ditch. The second part of the experiment, at the North side of the dyke, started at September 13 7:00h and went on until September 16 17:00h, when the pipe reached the ditch: continued pipe growth. These start and end times are based on the start and end moment of increasing the water height in the infiltration tubes.

Before the start of the experiments, a low water level was already in the infiltration tubes to make sure the dyke aquifer was already saturated with water. For the experiment at the South side, the infiltration tubes were filled with a constant small water level from September 3 12:30h until the start of the experiment. For the experiment at the North side from September 13 00:30h until the start of the experiment.

The measurement of thermal images started earlier and ended later than the increasing and decreasing of the water height as well. For the experiment at the South side the measurement started at September 6 7:00h and ended at September 9 12:00h. For the experiment at the North side the measurement started at September 13 00:00h and ended at September 16 21:15h. The RGB images are captured at the same moments as the thermal images. An overview of the data sets that are used in this research are listed in Table 3.1 and are explained in more detail in the following subsections.

Table 3.1: Data sets that are used in this research with measurement frequency and sizes of the data sets.

Available data	**Measurement frequency**	**Size data set South side experiment**	**Size data set North side experiment**
Thermal images	Every 10 seconds	25.820	44.106
RGB images	Every 10 seconds	25.820	44.106
Log book	When something noticeable happened	34 notation moments	97 notation moments
Incoming flow rate at the infiltration tubes	Every 15 minutes	654	410
Outgoing flow rate at the ditch	Random	None	65

3.2.1 Thermal imaging data

The thermal images were generated with an infra-red sensor containing a thermal resolution of 307.200 points per image. The output is a RAW 16 bit thermal .xpng image, which is a matrix of 480x640 pixels. Each pixel contains a measurement of the temperature. The thermal images are measured every ten seconds. The measurement system is built on an off-road vehicle on an 8 meter high telescopic mast containing a stability system and an Uninterruptible Power Supply (UPS) (Steenblik, 2021). The off-road vehicle is ±2 meter high, so the camera is positioned ±10 meter high. The set-up of this measurement device can be seen in figure 3.3.

Figure 3.3: Set-up of the thermal imaging measurement device (Steenblik, 2021).

3.2.2 RGB images

Next to the infra-red sensor, a camera measuring the visual spectrum of the light was installed on the telescopic mast shown in Figure 3.3. This camera generates RGB images with a resolution of 1.920x1.080 pixels (Steenblik, 2021). The camera captures an image every time a thermal image is generated, so every ten seconds. This is done so that the processes which are happening can always be compared to RGB images on that exact time. In appendix A Figure A.3 an example of a thermal image together with a RGB image can be seen where in both images a sand boil can be spotted.

3.2.3 Log book

During the experiments, the observers kept track of visual observations in a logbook. For this log book, no fixed period of making a notation is used, the time in between notation ranges from 10 minutes to several hours. During the night, no observations are done. Codes were used that correspond to the state of a sand boil in combination with the location of the sand boil in meters measured from left to right of the width of the ditch. Furthermore, the size of the sand boil is noted, if it boils sand or only water, the time it went active and the time it went inactive (Steenblik, 2021).

3.2.4 Flow rate measurements

The outgoing rate was intended to be measured by an automatic measurement device. However, this measurement device did not work. After this was noticed, at the 14th of September 14:24h, hand measurements were started. These measurements are taken at random times and are logged in the logbook. They are converted to the unit liter per minute.

There are two types of flow rate data: the flow rate coming into the dyke by infiltration tubes per 15 minutes, and the hand-measured flow rate coming out at the ditch. These two data sets are compared in Figure 3.4. In the figure it can be seen that during the night of 15 to 16 September, the flow rate was lowered to be sure that the point of continued pipe growth would occur during the day when observers were present.

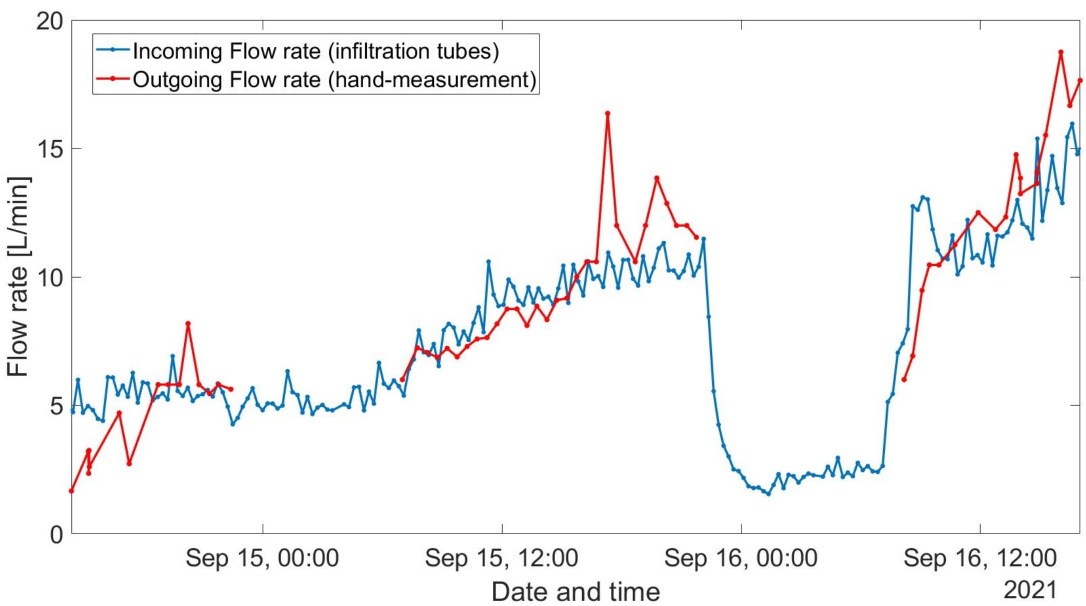


Figure 3.4: Comparison of the flow rate at the infiltration tubes and the hand-measured flow rate out of the ditch.

In the figure, it can be seen that the hand measured outgoing flow rate follows nearly the same path as the incoming flow rate in the infiltration tubes. Furthermore, a delay can be seen between the incoming and outgoing flow rate, which was best visible during the increase of the incoming flow rate in the morning of September 16. The delay is explainable because the flow rate which goes into the infiltration tubes will take a while to arrive at the ditch at the toe of the dyke, where the outgoing flow rate is measured. In the data set where Figure 3.4 is based on, it can be seen that the delay in the incoming and outgoing flow rate is approximately one hour. Moreover, the amount of measurements of the outgoing flow rate is less than the amount of measurements of the incoming flow rate. That a flow rate and thus seepage is taking place in the dyke does not necessary mean that piping is occurring. This is only the case when the seepage carries sand and thus backward erosion is happening. Based on the flow rate as a sole indicator, one can not conclude that sand is moved. But, an increasing seepage flow rate through the dyke can indicate that a pipe is forming since water can flow trough the dyke more easily.

4 | Methodology

In this chapter, the methodology of the three objectives is explained. In general, the steps that are taken are shown in Figure 4.1.
The first objective (O1) is reached by starting with detecting seepage spots in the thermal images. After that, the locations of these seepage spots are identified. Next, the size of the areas of these seepage spots is quantified. The second objective (O2) is reached by firstly preparing the data for the correlation analysis. After that, the correlation analysis between the areas of the seepage spots and the flow rates is performed, including a test for significance. The third objective (O3) is reached by performing a regression analysis for the best correlating data sets. This regression analysis will result in the final product, being a tool which estimates the flow rate of a seepage spot in a thermal image. In the next subsections, the methodology will be elaborated in more detail.

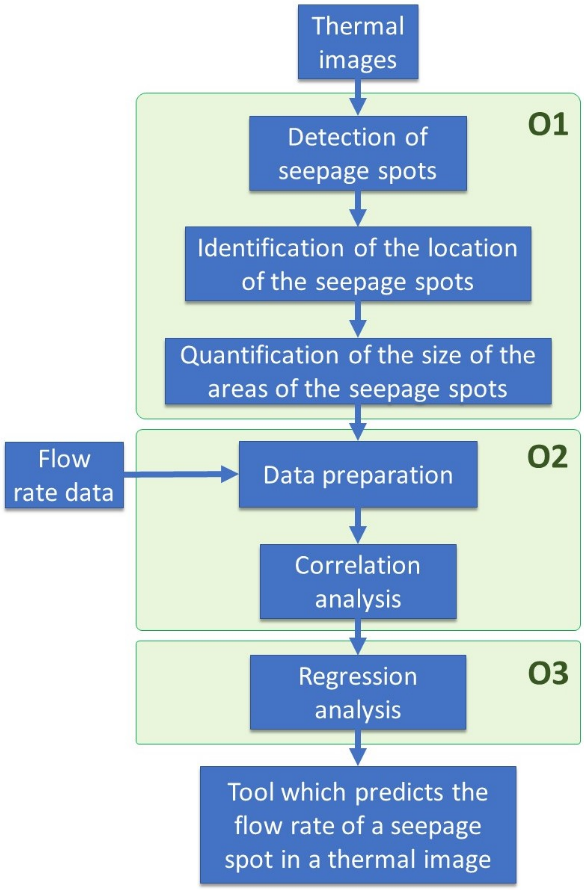


Figure 4.1: Flow chart of the methodology which summarizes the steps taken in the study, where the green squares make a division per objective (O).

4.1 Data analysis of thermal images (O1)

The first objective to develop a data analysis method of the thermal images to:

- Extract useful thermal imaging data from the data set, being thermal images in which a clear temperature difference is visible due to the occurrence of seepage;
- Estimate the sizes of the areas of the seepage spots from these data.

Three main components form the steps which are taken to analyse the thermal images: the detection of seepage spots in the thermal images, the identification of the location of these seepage spots and the quantification of the size of the areas of these seepage spots. This way of analyzing monitoring data is used a lot in Structural Health Monitoring (SMH) (Dhiraj et al. (2021); Downey et al. (2017); Larsen et al. (2007)) and is an effective way to convert RAW data into a quantified result. To extract useful thermal imaging data from the data set, the first step (detection) is used where the presence of seepage anywhere in a thermal image is detected. To extract the sizes of the areas of the found seepage spots in these data, the second step (identification of the location) and the third step (quantification) are used, where the location of the seepage spot is determined and the size of the area is quantified. Inside these three components, smaller sub-steps have to be taken to get to the next component. In the following subsections, the steps are explained in more detail.

4.1.1 Detection of seepage in the thermal images

The first component of the data analysis is to detect the presence of seepage anywhere in the thermal images. It is assumed that a temperature difference in the ditch that is not caused by an obstacle implies that seepage is present. As explained before, seepage is expected to have another temperature than the temperature of the water in the ditch. So, images in which a temperature difference in the ditch is present are extracted in this step. The disturbances in the thermal images are detected and kept out of the regions of interest. In the case of disturbance by for example people, the images are removed from the useful image list. Three criteria have been set up to select the images that are suitable for this research. To determine to what extent thermal imaging can be used to make an estimation of the flow rate of a seepage spot, it is important that flow rate data is available for the thermal imaging data set. The thermal imaging data is considered to be useful if it meets the following criteria:

1. Flow rate data is available for the thermal imaging data;
2. There is no disturbance by people or other obstacles in the thermal image, meaning that there are no people or other obstacles in front of the selected regions of interest, so that the seepage spot could possibly not be seen.
3. A temperature difference is visible in the thermal image: there is a temperature difference between the water in the ditch and the water flowing out of the sand boil, and there is little effect from the weather conditions, meaning that there are no such weather events that the temperature difference at the water surface of the ditch cannot be seen in the thermal image anymore;

The difference between a seepage spot and a sand boil can not be derived from a thermal image. Therefore, a temperature difference in a thermal image is assumed to be a seepage spot in the first place, since water flowing out of a sand boil is seepage as well. With the information in the log book or by visual inspection in the field, a check can be done to find out if the seepage spot is caused by a sand boil and thus piping.

The detection of seepage in the thermal images is done by firstly identifying the region of interest in the images. Subsequently the data set is reduced to the thermal images with the presence of seepage in three steps. Firstly, the amount of data is reduced to the part wherefore a flow rate is available. This is because the end goal of this research is to detect a possible relation between the thermal images and the flow rate. Secondly, the data set is reduced by excluding images with people in front of the region of interest. Thirdly, the data set is reduced by detecting the visibility of temperature differences (being seepage spots).

Regions of interest

The first step in approaching the objective is to select the regions of interest. These regions of interest (ROIs) are selected based on the expected location where seepage and sand boils are occurring. This location depends on the situation. When a ditch is present, this is the location where sand boils are expected and thus the region of interest, since this is the location of smallest resistance. In the case of no ditch, the process of uplift can cause a crack in the cover layer of the dyke toe, where the sand boil can then occur (van Beek et al. (2013); Technical Advisory Committee (1999)). In that case, the dyke toe area is the region of interest. Furthermore, obstacles and other hard structures where piping can surely not occur in, or which are blocking the view, have to be left out of the ROIs.

Availability of the flow rate

The first reduction of the thermal imaging data set is made based on the availability of the flow rate. The part where no flow rate is available is excluded from the data analysis.

People

The second reduction of the thermal imaging data set is made by excluding the images in which people are standing in front of the ROIs. To detect these images, histograms are created using the temperatures inside the regions of interest in the thermal images. These histograms show that in time, there are differences between images with and without people in front of the ROIs. The histograms of images with people in front of the ROIs show higher occurring temperatures, which are found to be local intensities (Li et al., 2012). It is concluded that when a temperature of 20 degrees or higher is present in the ROIs, the image contains people standing in front of the ROIs. A MATLAB script is created to extract the images with temperatures above 20 °C present inside of the ROIs. This results in a list with image numbers of thermal images without people.

Visibility of temperature differences

The third and last reduction of the data set is based on the visibility of temperature differences in the Regions of interest. To be able to extract an area of the seepage spot from the thermal images it is important that the data set for the research consists of images in which a (clear) temperature difference is visible. Tipping points where the seepage water is changing from cold to warm seepage relative to the ditch temperature and weather events are causes of no clear temperature difference being visible (van de Voort et al., 2021) and are therefore extracted. The temperature differences are detected statistically by histograms together with fitted curves. It is chosen to fit a Kernel type of curve to the histograms of the present temperatures, since the Kernel type follows the shape of the histogram the best. An example can be seen in Figure 4.2, where the frequency of the pixels of a certain temperature is shown with the blue bars and the fitted Kernel distribution is shown with a red line.

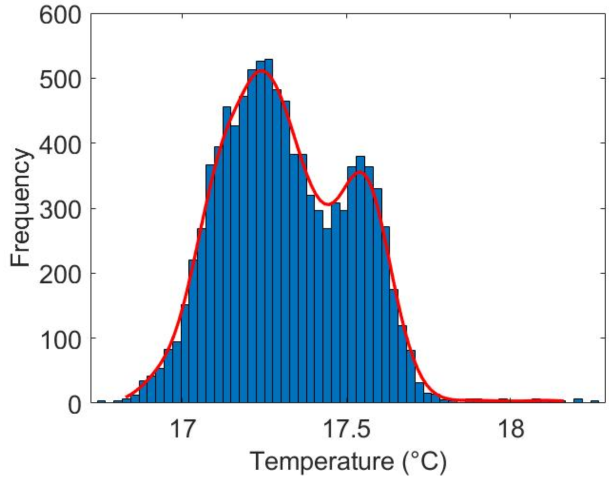


Figure 4.2: An example of an histogram with a fitted Kernel distribution curve.

Histograms with fitted curves show the intensity of the temperatures that are found in the ROIs. They are tools that show the different temperatures which are present in the ditch. A high intensity peak and small temperature range will indicate that there is no significant temperature difference between the water in the ditch and the seepage water being visible in the thermal image. A large temperature range and two or more lower intensity peaks indicate that a temperature difference can be seen in the thermal image. This method is based on Jiang et al. (2016) and Tsanakas and Botsaris (2012). Both studies also used histograms with fitted curves to indicate the peaks and to detect hot spots in their thermal images. Figure 4.3 shows the expected shapes of the peaks, where the green line indicates no clear seepage spot visible and red and blue line indicate that seepage is visible.

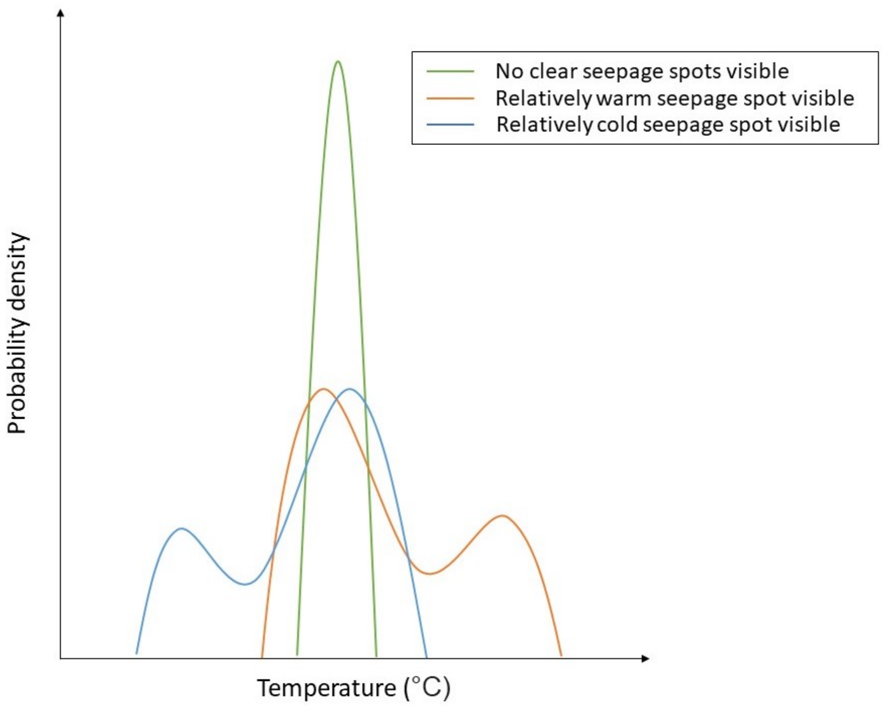


Figure 4.3: Expected temperature distributions in thermal images with and without seepage spots, based on the methods of Jiang et al. (2016) and Tsanakas and Botsaris (2012)

For every thermal image a fitted curve is created. These curves are combined in a 3D plot which shows all intensity curves over time. By sampling about 30 images from this 3D plot with different peak heights and opening these thermal images it is assessed that a clear temperature difference can be seen. A peak above a probability density of 3 is considered to be a thermal image with a too small temperature range in the ditch. This phenomenon is specifically valid for this data set and should be determined again in case of other data sets. A MATLAB script is developed to extract the images with a probability density curve which is higher than 3. After this step, the result is an image list with only useful images.

4.1.2 Identification of the location of the seepage spots

The second component of the data analysis of the thermal images is to identify the location of the seepage spots. This is done by analysing videos of the remaining thermal imaging data. When a spot with a temperature difference is seen in the video, it is presumably a location of a seepage spot. These locations seen in the videos are validated with the notes in the logbook telling the location based on the width of the ditch and with the visual images captured with a RGB camera. After that, the precise location of the core of the seepage spot is determined by zooming into the thermal images and comparing areas to see what part is always the same. This results in pixel coordinates.

4.1.3 Quantification of the size of the seepage spots

The third component of the data analysis is to quantify the size of the area of the sand boils in the thermal images. To be able to quantify this size, the core temperature is extracted from the pixel location of the seepage spot. After that, the temperature ranges of the seepage spots are determined. Then, the areas are extracted by counting the pixels with that temperature ranges. Lastly, the amount of pixels is converted to square meters. These steps are explained below.

Core temperature of the seepage spot

Once the pixel location of the core of the seepage spot is found, the temperature is extracted from that location by selecting that pixel and the 8 pixels surrounding that location and consequently calculating the average of these 9 pixels. It is chosen to make use of an average of 9 pixels to reduce the chance on outliers and small measurement errors. This temperature is needed to extract the area of the seepage spot in the thermal image. The result of this step is a matrix with the core temperature of the sand boil over time (per image).

Temperature ranges

Now that the core temperature of the seepage spot is found, the mixing with the water in the ditch has to be taken into account to determine the areas of the seepage spots. To find the best relation as possible between the area and the flow rate, it is decided to make use of three different temperature ranges: 0.2 °C, 0.4 °C and 0.6 °C. In this temperature ranges, the difference of mixing in case of cold seepage and warm seepage is taken into account. When water is seeping into the ditch at a colder temperature than the ditch, it will mix to a warmer temperature. When the seepage water is warmer than the temperature of the water in the ditch, it will mix into a colder temperature. Still, the area can consist of temperatures a little colder and warmer than the core temperature since the core temperature of the sand boil is only based on a few specific pixels. Taking all this into account, the temperature ranges are assumed to be as in Table 4.1.

Table 4.1: Temperature ranges around the core temperature of the sand boil

Range	In case of cold seepage	In case of warm seepage
0.2 °C	-0.05 °C & + 0.15 °C	-0.15 °C & + 0.05 °C
0.4 °C	-0.1 °C & + 0.3 °C	-0.3 °C & + 0.1 °C
0.6 °C	-0.2 °C & + 0.4 °C	-0.4 °C & + 0.2 °C

Areas

To extract the areas, a MATLAB script is developed to count the amount of pixels inside the ROIs that match with the temperature range explained above. To determine if cold or warm seepage is occurring, the temperature of the seepage spot is compared with the ambient temperature. In this case the temperature of the wooden beam laying over the ditch is taken as the ambient temperature. Once the situation is known (cold or warm seepage), the amount of pixels can be counted according to Table 4.1. This results in a data set giving an amount of pixels over time (per image). Since the flow rate corresponds to all seepage spots together, it is chosen to take the areas of all seepage spots in the ditch together.

Convert amount of pixels to square meters

To convert the amount of pixels to square meters, an approximation of the pixel size is made. This is done by counting the pixels of an object in the thermal image of which the real dimensions are known. This way, the real size of a pixel is calculated. The ditch at the experiment has dimensions of 1 by 10 meters. The viewing angle of the infrared camera is not exactly perpendicular to the ditch, so the distortion of the image is taken into account as well. The area of the water in the ditch that can not be seen in thermal image is determined by comparing the length of the wooden beam to the visible width of the water. They should both be 1 meter, but due to the viewing angle of the camera, part of the water in the ditch can not be seen. This way, the approximation of the area of the water which can be seen in the image is calculated. The total amount of pixels of water that can be seen in the image are counted. The approximation of the area of the water is divided with the amount of pixels, resulting in the size of a pixel. After the real size of a pixel is found, the area is calculated by multiplying the amount of pixels per image with the size of a pixel.

4.2 Quantification of a possible correlation (O2)

The second objective of this research is to quantify a possible correlation between the flow rate of the sand boil and the size of the seepage spot in the thermal images. Two main components form the steps to reach this objective: the data preparation and the correlation analysis. They are explained in more detail in the following subsections.

4.2.1 Data preparation

It is chosen to use the incoming flow rate data (infiltration tubes) of the experiment for the correlation analysis. The hand measured outgoing flow rate follows nearly the same path as the incoming flow rate in the infiltration tubes (see Figure 3.4). Moreover, the amount of measurements of the outgoing flow rate are very little, and therefore not very suitable to be used in a data analysis. Because of these two reasons, it is chosen to perform the correlation analysis on the following two data sets: incoming flow rate and thermal imaging seepage spot areas. As a first step of this objective, the outliers and fluctuations in the data are removed. After that, the data sets (flow rate and areas) are compared to find a possibly correlating part. Next, the selected parts of the data sets have to be changed into data sets of the same length (with the same time stamps per data). These steps are explained in more detail in the following subsections.

Smooth the outliers and fluctuations

Since both data sets show many small fluctuations over time, which are determined to be small errors, it is chosen to smoothen the data for outliers and fluctuations. For the flow rate, the fluctuations that are happening within an hour are smoothened, since it is determined that the flow rate can not change that fast. For the selected areas of the seepage spots, fluctuations are witnessed that occur within minutes and also within hours. The fluctuations within minutes are determined to be small errors in the selection method of these areas or small errors in the temperature measurement of the IR camera. The fluctuation within hours are determined to be caused by weather events such as a cloud flying over the ditch creating a moment of shadow or a moment where the sun reflects on the water. To remove the fluctuations in the data, it is decided to smoothen the data over an hour with the use of the MATLAB function $smooth$.

Compare the variables

The flow rate is plotted in one Figure with the selected areas of the seepage spots. This plot gives a first impression of a possible correlation between those two variables.

Fit the timestamps

To be able to perform a correlation analysis, the two data sets should be of the same length. Since the flow rate data is measured every 15 minutes and the thermal images are measured every 10 seconds, the area data originating from the thermal images is reduced to a value every 15 minutes.

4.2.2 Correlation analysis

To find out if there is a relation between the flow rate and the areas of the seepage spots in the thermal images, a correlation analysis is performed (Gogtay and Thatte, 2017). Firstly a scatter plot is made, and the correlation coefficients are calculated. After that, the significance of this correlation coefficient is calculated, to determine if the found relation is significant. These steps are explained in more detail in the following sections.

Scatter plot and Correlation coefficient

A scatter plot is made of the prepared data. To calculate the linear correlation coefficient the Pearson correlation coefficient was computed using MATLAB, which is based on equation 4.1 and 4.2 (Davis, 2002). To calculate the non-linear correlation coefficient the Spearman correlation coefficient was computed, which is based on equation 4.3.

$$r(A,B) = \frac{cov(A,B)}{\sigma_A \sigma_B} \tag{4.1}$$

$$cov(A,B) = \frac{1}{n-1}\sum_{i=1}^{n}\left(A_i - \mu_A\right)\left(B_i - \mu_B\right) \tag{4.2}$$

Where:
r = Pearson's Correlation coefficient;
A = Data set 1;
B = Data set 2;
n = Number of observations;
μ = Mean of A or B;
σ = Standard deviation of A or B.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{4.3}$$

Where:
ρ = Spearman's rank correlation coefficient;
d_i = Difference between the ranks of each observation;
n = Number of observations;

The Spearman and Pearson correlation coefficients range from -1 to +1. A coefficient of -1 indicates a perfectly related negative correlation, a coefficient of +1 a perfectly related positive correlation and a coefficient of 0 indicates no correlation Gogtay and Thatte (2017).

Significance

A significance test is performed. The null hypothesis that there is no correlation present between the flow rate and the areas (Equation 4.4) is tested against the alternative that there is linear correlation present (Equation 4.5). The significance is tested for a level of significance of 5%, $\alpha = 0.05$.

$$H_0 : r = 0 \tag{4.4}$$
$$H_1 : r \neq 0 \tag{4.5}$$

A p-value is calculated to test the significance of the correlation. The p-value is, assuming that the H_0 is correct, the probability that the data could be obtained by a random chance (Davis, 2002). This p-value is calculated using a normal distribution curve and is calculated using MATLAB.

If the resulting p-value is less than the significance level, the null hypothesis can be rejected. If the resulting p-value is higher than the significance value, the null hypothesis can not be rejected which means that no significant correlation was found (Davis, 2002).

4.3 Regression analysis (O3)

The third objective of this research is to develop a practical tool to enable flow rate prediction based on the seepage spot size from the thermal images.

The purpose of the regression analysis is to develop an equation which can be used to predict the flow rate based on a seepage spot found in an thermal image. The data set with the highest correlation coefficient is chosen as the input for the regression analysis, since this shows the best relation. So a regression analysis is performed between the areas of the seepage spots in the thermal images and the flow rate. A formula is developed which provides an indication of the flow rate connected to the found area in the thermal image.

When a linear relation is found, a simple linear regression line is formed based on the formula given in Equation 4.6. When a non-linear relation is found, the best possible curve is chosen to fit the data. An indication of the best shape is based on the scatter plot. In the case of predicting the flow rate based on the seepage spot in a thermal image, a polynomial curve is expected to fit the data the best. The general equation for a polynomial curve is given in equation 4.7. When the flow rate has become large, it is expected that at a certain point the area of the seepage spot in the thermal image does not grow much bigger anymore. While in case of a small flow rate which increases, the area of the seepage spot in the thermal image is expected to grow very fast. Therefore, the shape of the curve is expected to increase rapidly at smaller areas and flow rates, and slower at larger areas and flow rates. Since over-fitting the data is a risk with a small data set as in this research, an equation with an low order should be used.

$$y = ax + b + \epsilon \tag{4.6}$$

Where:
y = the predicted flow rate [L/min];
x = the area of the seepage spot in the thermal image [m^2];
b = the y-intercept;
a = the slope or regression coefficient;
ϵ = the error term.

$$y = c_0 + c_1 x + c_2 x^2 ... c_n x^n + \epsilon \tag{4.7}$$

Where:
y = the predicted flow rate [L/min];
x = the area of the seepage spot in the thermal image [m^2];
n = the degree of the polynomial;
c_0 = the y-intercept;
c_n = the coefficients depending on the best fit;
ϵ = the error term.

5 | Results

5.1 Data analysis of thermal images (O1)

In this section, the results of the data analysis of the thermal images are shown. Useful thermal imaging data is extracted from the data set and the sizes of the areas of the seepage spots are extracted from the remaining thermal images. The results are divided in three main steps:

- Detection of seepage in the thermal images;
- Identification of the location of the seepage spots;
- Quantification of the size of the seepage spots.

5.1.1 Detection of seepage in the thermal images

Firstly, the useful images were detected by setting the region of interest, checking the availability of the flow rate, detecting people and checking the visibility of the temperature differences in the images.

Regions of interest

The ditch is the region of interest for this experiment. This is where the sand boils are expected to arise (van Beek et al., 2013). In front of this ditch, which is the rectangular area enclosed by wooden planks which can be seen in Figure 5.1, there are a few other obstacles that are not of interest. These are measurement devices for other aspects of the experiment and wooden gangways for inspection purposes. The obstacles can be seen in Figure 5.1 where they are numbered. The wooden gangways are numbered as 1, the larger measurement devices as 2 and the small measurement devices as 3. All obstacles are excluded from the ROIs since they will show temperature differences in the thermal images which are not caused by seepage.

Figure 5.1: Obstacles inside of the ditch at the experimental set-up where the numbering corresponds to 1) wooden gangways 2) larger measurement devices and 3) small measurement devices.

It was found that the wooden gangways are moved two separate times. Therefore, three settings are used where the Regions of Interest are adjusted:

- Setting 1 (S1): 14 September 13:48h up to and including 15 September 8:20h, with regions of interest as indicated with the blue polygon in Figure 5.2(top);
- Setting 2 (S2): 15 September 8:20h up to and including 15 September 13:47h, with regions of interest as indicated with the blue polygon in Figure 5.2(middle);
- Setting 3 (S3): 15 September 14:02h up to and including 16 September 17:02h, with regions of interest as indicated with the blue polygon in Figure 5.2(bottom).

A time gap can be seen between the settings since it took some time to move the wooden gangways to the new location.

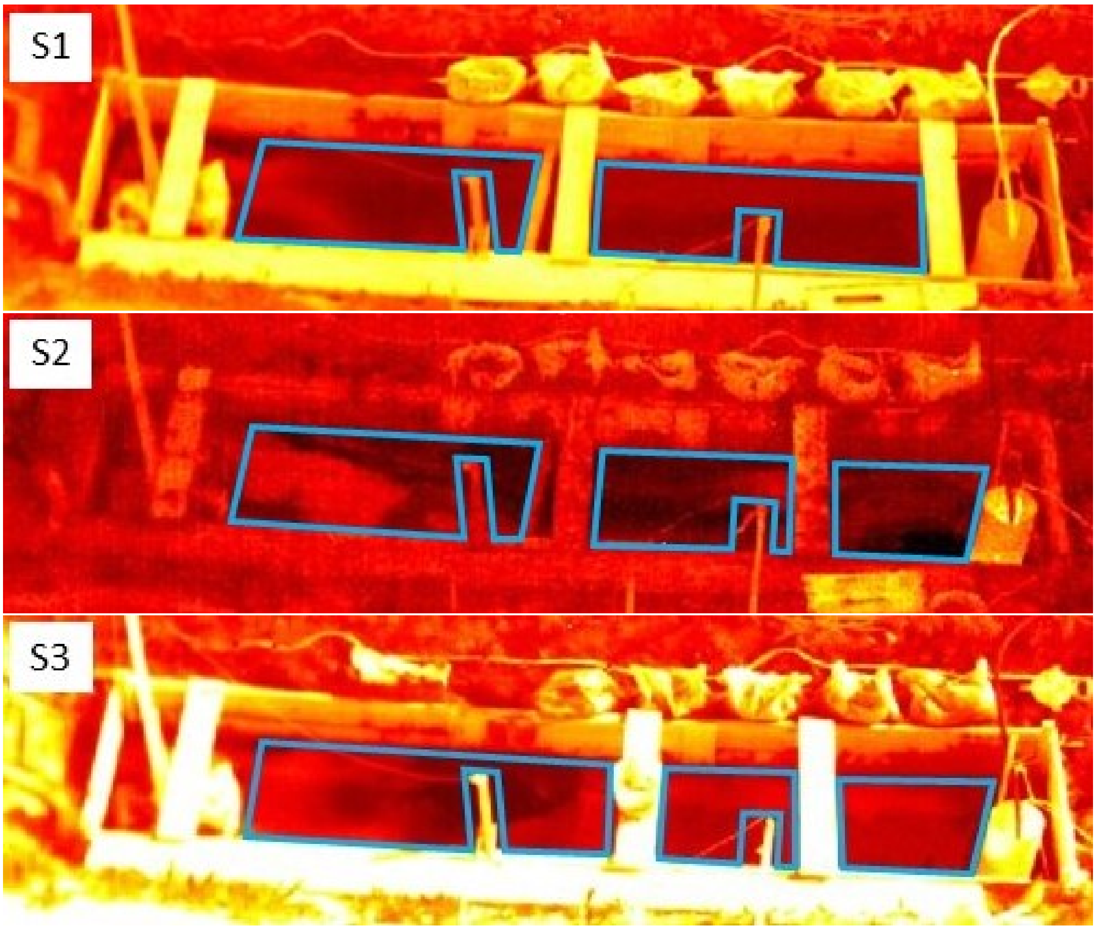


Figure 5.2: The regions of interest for Setting 1 (top), Setting 2 (middle) and Setting 3 (bottom)

Availability of the flow rate

As mentioned in Section 3.2, the flow rate data is only available starting from the 14th of September 14:24h, until the end of the experiment. Therefore the used data for the analysis in this research is reduced to the time span from 14 September 14:24h until 16 September 18:02h. The result is a remaining data set of 17228 images.

People

The selection of thermal images with temperatures of 20 °C or higher inside of the ROIs resulted in a data set of images with people of 5988 images. The result is a remaining data set without people of 11240 images. An example of a thermal image with people in front of the regions of interest can be seen in Appendix B in Figure B.1.

Visibility of temperature differences

For every remaining thermal image after the first two data reductions (availability flow rate and images with people in it) a fitted curve is generated and plotted in a merged figure. A division is made in three figures showing the curves over time per Setting. Figure 5.3 shows the fitted curves of the temperatures inside the ROIs of Setting 1 of the data, which is the setting before the wooden gangways shifted and the ROIs had to be adjusted. It can be seen that in this setting of the data, the first section has a low, wide curve, while the last section of this data has a high, small peak. In the second setting (Figure 5.4) and the third setting (Figure 5.5) of the data a clear difference between high and low peaks through the time can be seen as well. As stated in the methodology, it is expected that high, narrow peaks do not show seepage spots in the thermal image while low, wide curves do. For this data set it is seen that this expectation is met. Some examples can be found in Appendix C in Figure C.1 and the corresponding Figure C.2a to Figure C.2f.

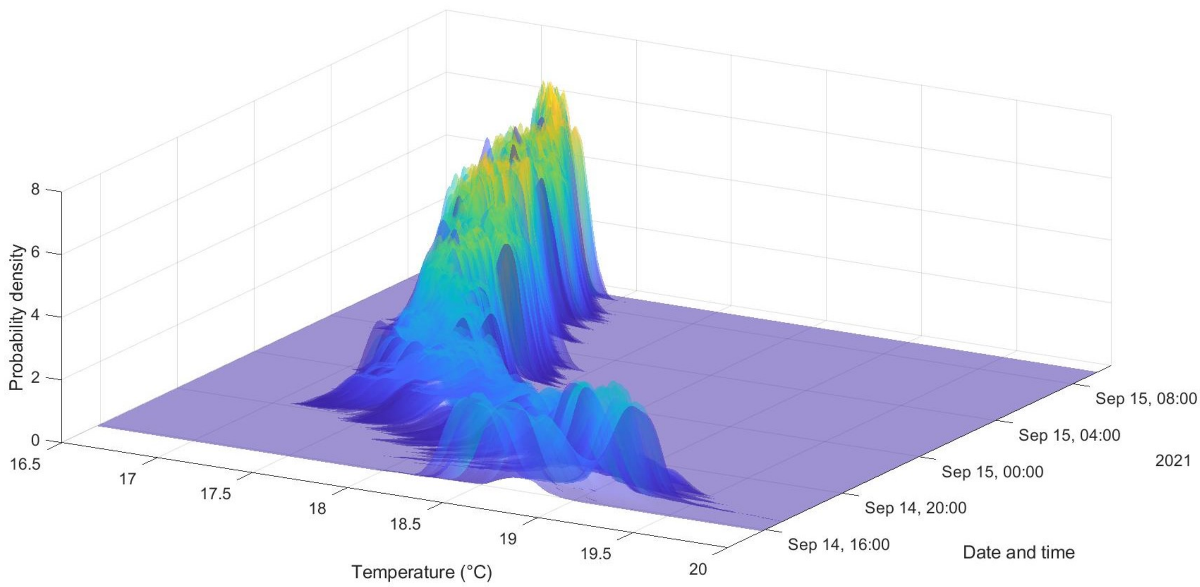


Figure 5.3: Plot of all fitted curves in time for Setting 1 with the purpose to show temperature differences in thermal images

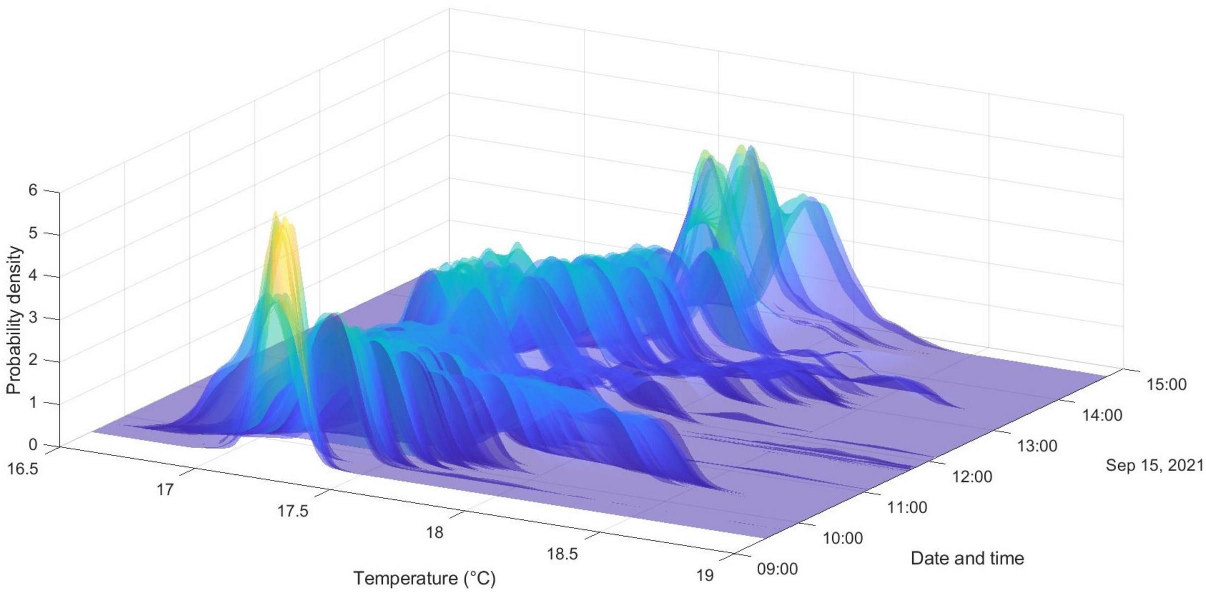


Figure 5.4: Plot of all fitted curves in time for Setting 2 with the purpose to show temperature differences in thermal images

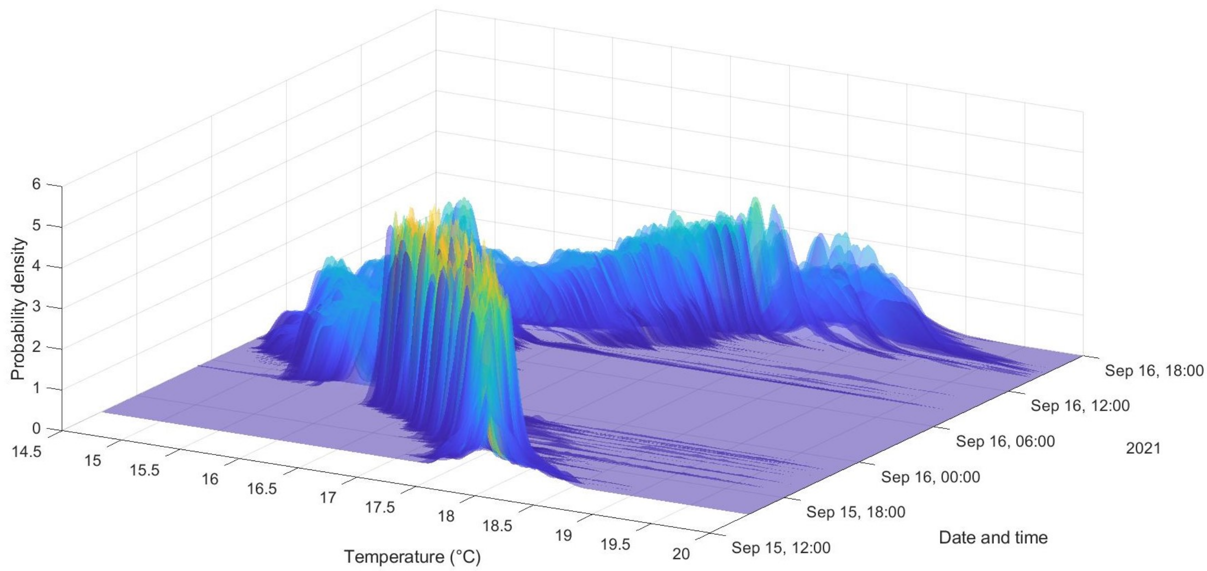


Figure 5.5: Plot of all fitted curves in time for Setting 3 with the purpose to show temperature differences in thermal images

The images with a peak above a probability density value of 3 are extracted from the useful data. This turned out to be 3756 images. After this step, the data set is successfully reduced to images that are useful for this research. This resulted in a remaining thermal images data set of 7484 images.

5.1.2 Identification of the location of the seepage spots

As the next step, the locations of the seepage spots are identified. It is found that there are three seepage spots in the remaining data set. Figure 5.6 shows the locations of these three seepage spots.

Figure 5.6: The locations of the three seepage spots in the remaining data set, indicated with a blue dot. The seepage spot in the middle is active during the whole data set, the seepage spots on the left and on the right are not active during the complete data set. The obstacles as shown in Figure 5.1 are kept out of the analysis.

The seepage spot in the middle is the most active seepage spot which is active during the whole data set. The seepage spots on the left and on the right side of the ditch are not active during the whole data set, but only partly. It is found that the seepage temperature of the three different seepage spots is relatively the same. Therefore, the middle sand boil is used for the core temperature extraction which will be used in the next step (Quantification of the size of the seepage spots). The pixel location of the core of the middle and most important seepage spot is (280,280).

5.1.3 Quantification of the size of the seepage spots

In this section, the results of the quantification of the size of the seepage spots are shown. The core temperature of the sand boil is extracted, resulting in a list of this temperature per image. There is made use of different temperature ranges for the extraction of the seepage spots. The areas of the seepage spots are extracted by counting the pixels and these areas are converted to the representing size in square meters.

Core temperature of the seepage spot

The core temperature of the seepage spot resulted in the temperature progression in time as shown in Figure 5.7.

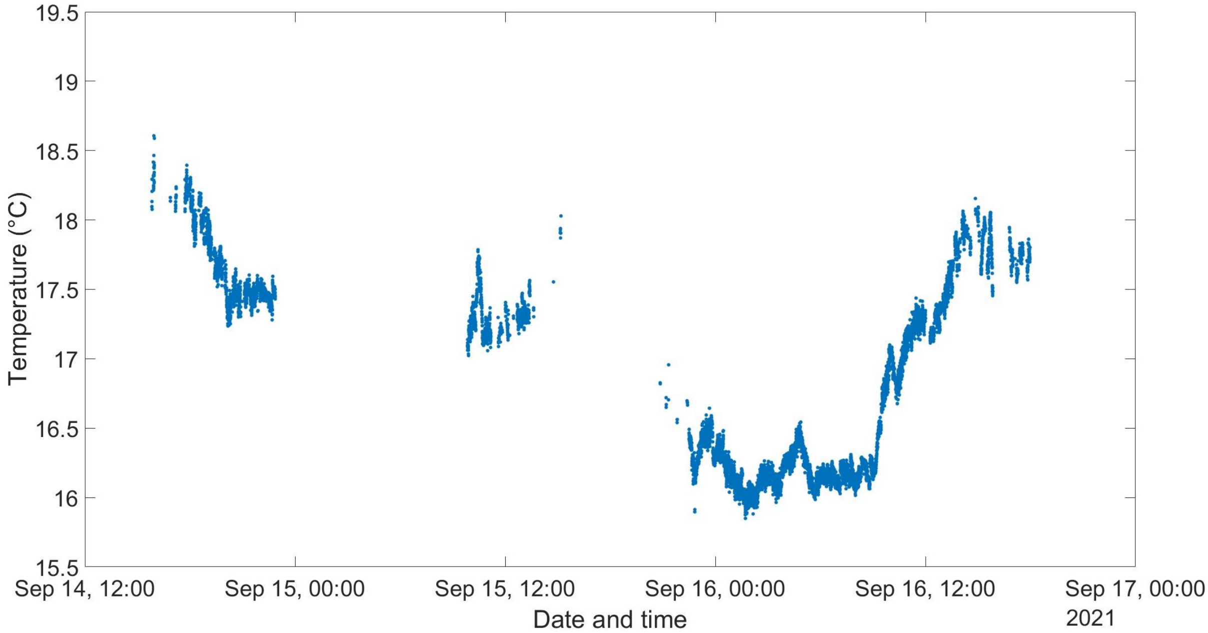


Figure 5.7: A plot of the core temperature of the main seepage spot over time.

Temperature ranges

The temperature ranges which are used result in different sizes of areas that are quantified. An example of the taken area for three different temperature ranges can be seen in Figure 5.8. In Appendix D other examples can be seen which shows the case of multiple active seepage spots at once in Figure D.1 and Figure D.2.

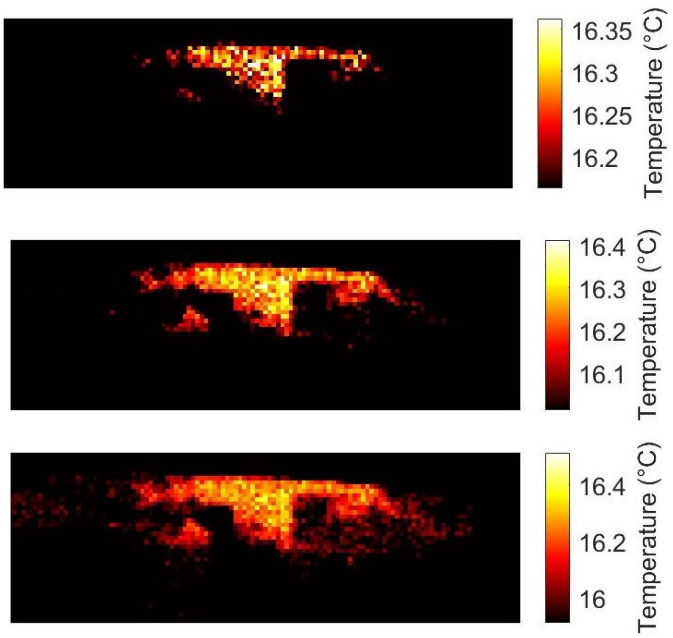


Figure 5.8: The quantified area for three different temperature ranges: 0.2 °C (top), 0.4 °C (middle) and 0.6 °C (bottom) where the yellow/orange/red colors indicate the pixels that are counted as the area. The image is a cropped part of the ditch.

Areas

The areas are extracted using the core sand boil temperature and the three different ranges as explained in the Methodology. The comparison of the ambient temperature with the core seepage spot temperature is shown in Figure 5.9. As can be seen, there are two periods where cold seepage occurs in the data set (beginning and end) and one period where warm seepage occurs (middle). The extracted areas are shown in Figure 5.10.

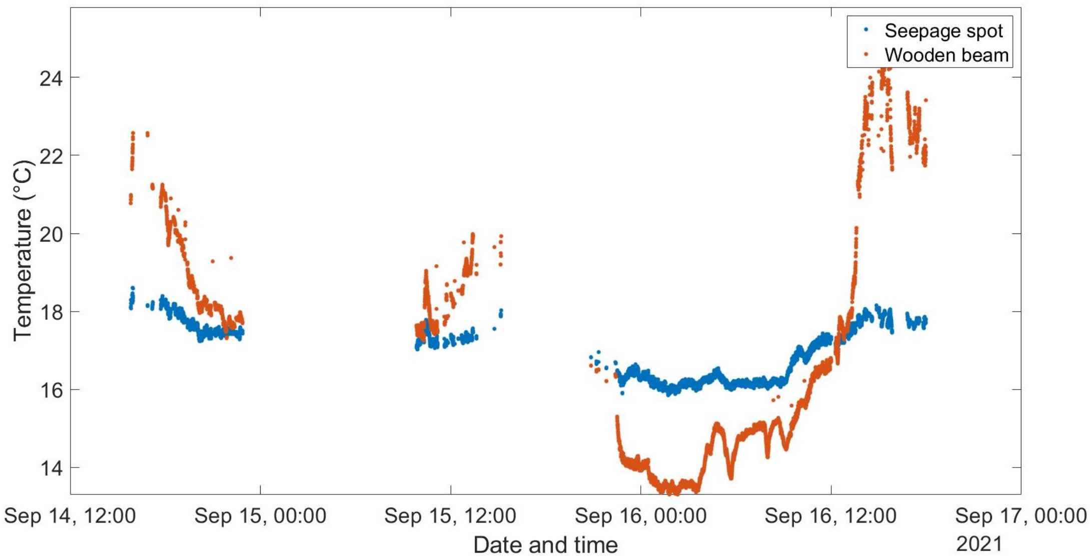


Figure 5.9: A plot of the core temperature of the main seepage spot and the ambient wooden beam temperature over time.

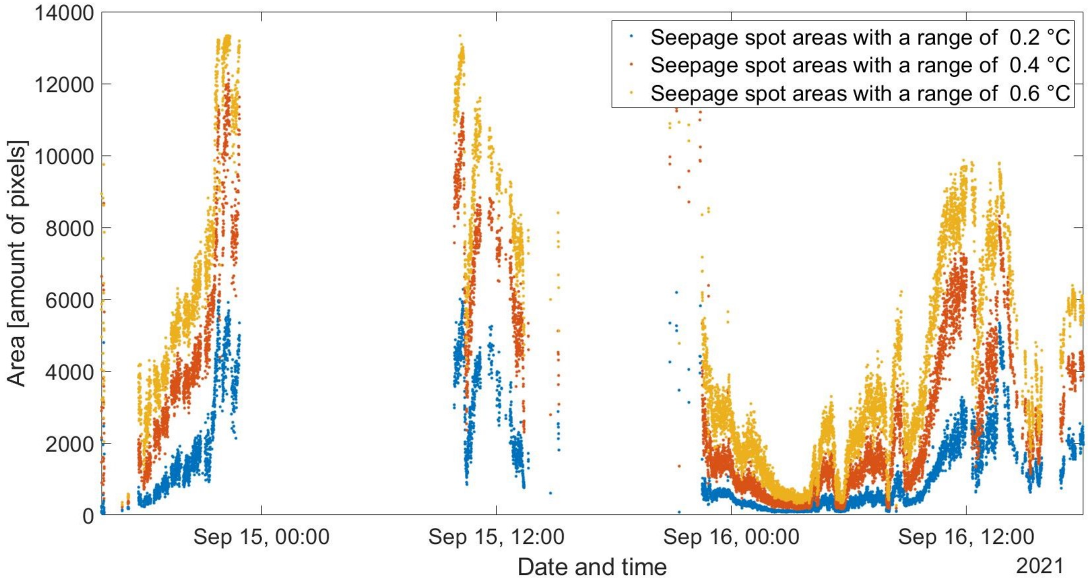


Figure 5.10: The extracted areas from the thermal images in amount of pixels over time for the three different temperature ranges.

Convert amount of pixels to square meters

Figure 5.11 shows one of the thermal images with the known dimensions of the ditch (1 by 10 meters) in it. The wooden beam on the right side of the thermal images (see Figure 5.11) is found to be 64 pixels in length. So 64 pixels is set to be equal to 1 meter. The cross-size of the water is found to be 41 pixels. With this information it is concluded that the part of the water that can be seen in the thermal image is approximately 0,64 meters. On the left side of the image, the wooden beam and cross-size of the water contains some more pixels. Because this has a small effect on the resulting width, it is chosen to stick to the approximation of the cross-size of the water in the ditch of 0.64 meters.

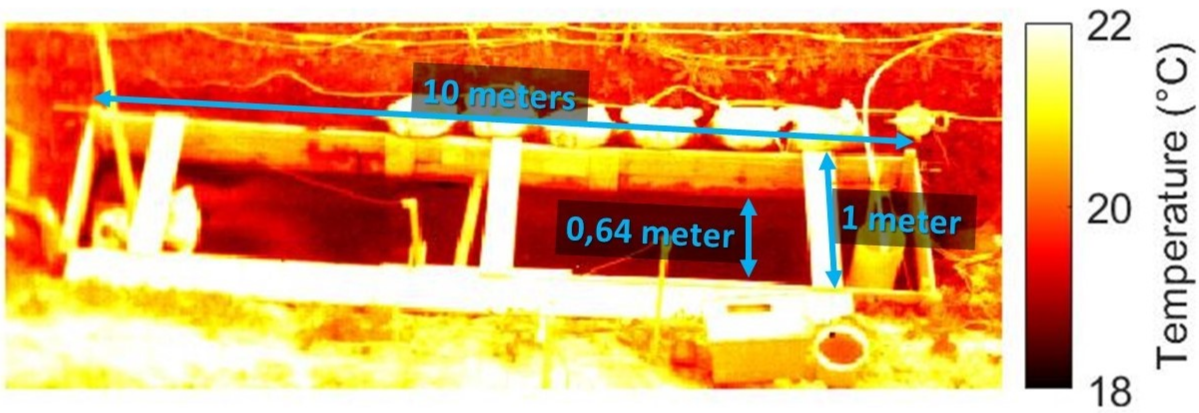


Figure 5.11: Known dimensions in the image which are used for the determination of the pixel size.

Furthermore, it is known that the ditch has a width of 10 meters. This width can be completely seen in the thermal image, so it is concluded that the water seen in the thermal images has an surface area of $10 * 0,64 = 6,4m^2$. When the amount of pixels is counted inside the region which is shown in Figure 5.12, the number of pixels is found to be 17425. Now the size of one pixel is calculated: $6,4m^2/17425 = 3,7 * 10^{-4}m^2 = 3,7cm^2$. This is an approximation of reality since the picture is taken with an unknown viewing angle. The error of this approximation is expected to not be a significant problem for this research. In conclusion, the size of one pixel in the thermal image is determined to be $3,7cm^2$.

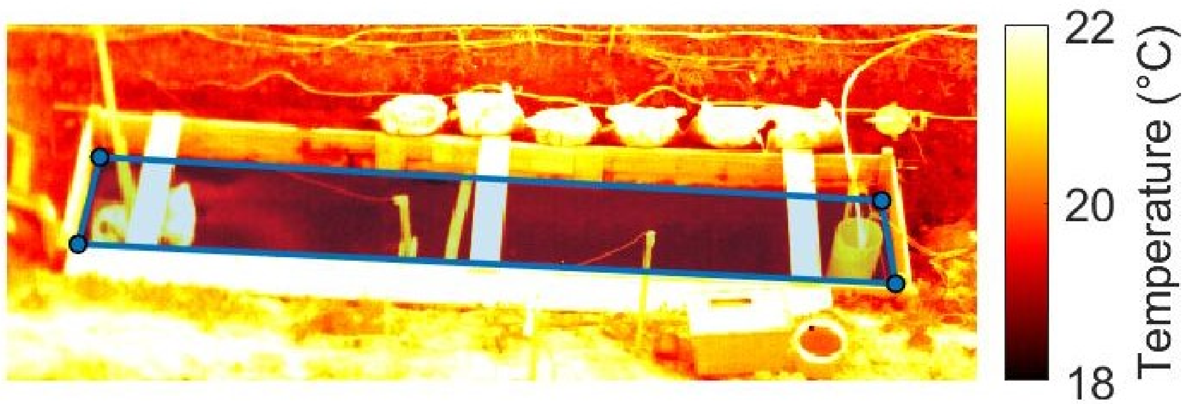


Figure 5.12: The area defined with the blue lines shows the pixels which are counted to calculate the size of the pixels.

5.2 Quantification of a possible correlation (O2)

In this section, the results of the quantification of a possible correlation are given. The flow rate data and seepage spot areas data is prepared for a correlation analysis. A correlation analysis is performed for a linear and non-linear correlation and these results are tested for significance.

5.2.1 Data preparation

The data is prepared by smoothing the outliers and fluctuations in the data sets, comparing the variables for a possible correlating part and by fitting the timestamps. The results are given in this section.

Smooth the outliers and fluctuations

In Figure 5.13 the incoming flow rate and the seepage spot areas are compared. It can be seen that there are many fluctuations in both data sets. The fluctuations in the flow rate are found within an hour. This is presumably caused by the step-by-step intake of the water in the infiltration tubes. The fluctuations in the areas are found within minutes of time, while the area of the sand boil can not change that fast. The cause of these fluctuations is expected to be a combination of weather events (sun/shadow/rain), fluctuations in the seepage temperature and the validation of the camera. Therefore it is chosen to smoothen the data.

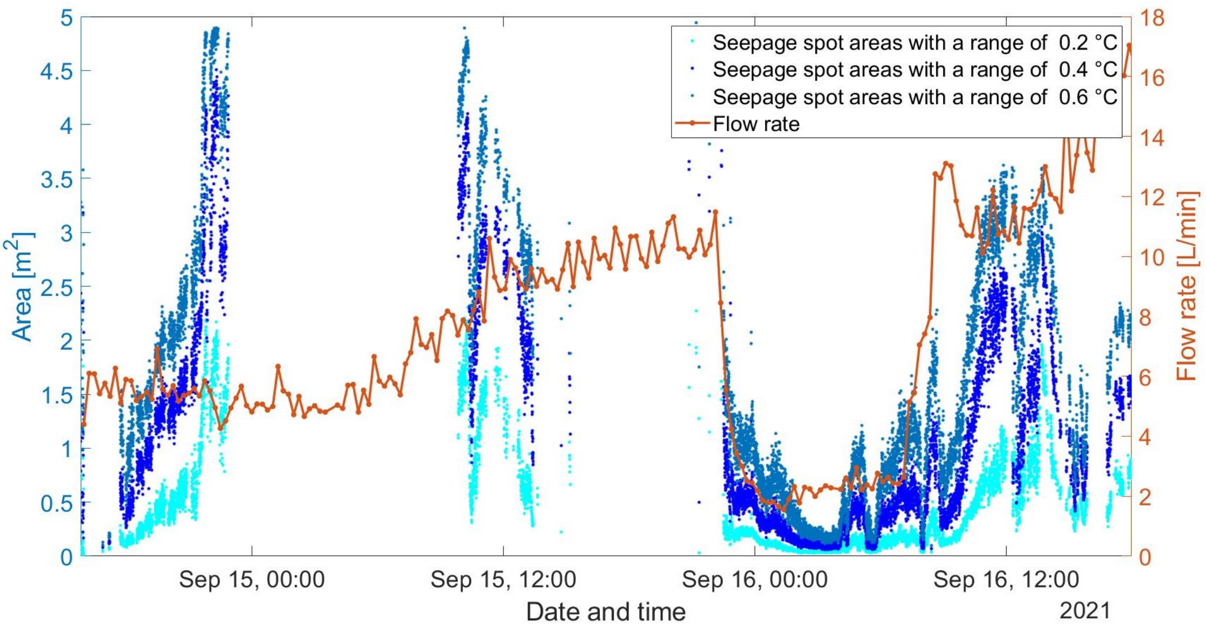


Figure 5.13: Flow rate compared with the areas of the seepage spots for different temperature ranges.

The data is smoothed with the goal to remove fluctuations within an hour. The moving average used for this is therefore based on two hours of data around the data point. This process results in the lines shown in Figure 5.14. In Appendix E the curves can be seen together with the data points.

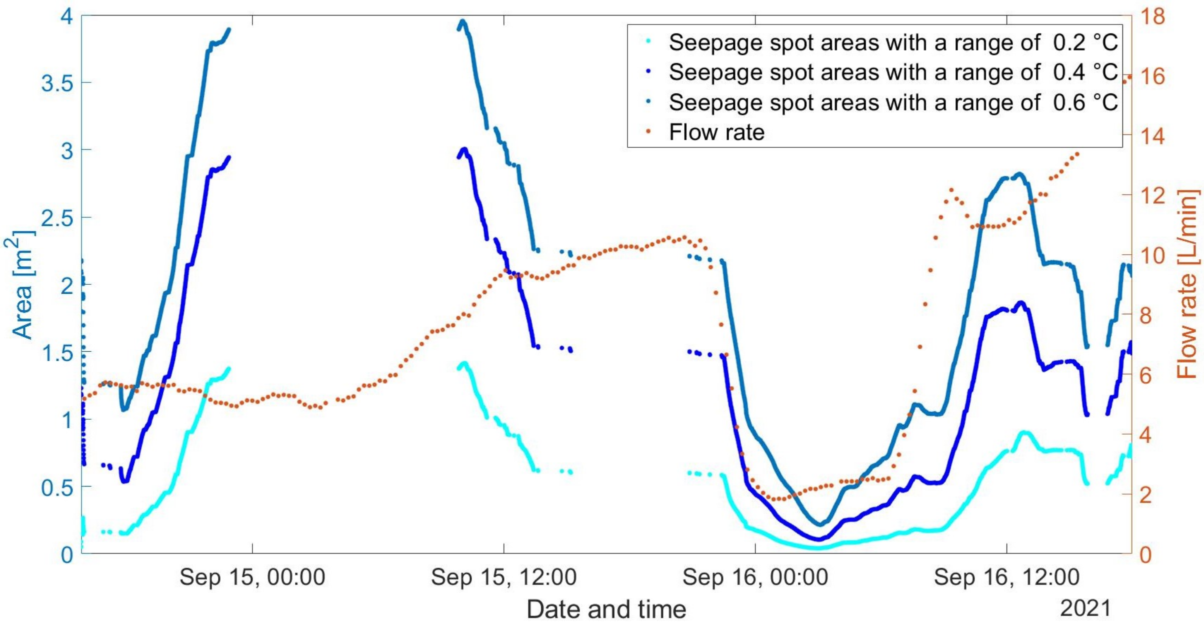


Figure 5.14: Flow rate compared with the smoothed areas of the seepage spots for different temperature ranges.

Compare the variables

In Figure 5.14 can be seen that the first part of the data set (Sept. 14 15:50h up to Sept. 15 22:00h) shows no clear correlation. The two lines/time-series that are left in this part of the figure after the reduction of the data set in objective 1, show a completely different shape than the flow rate at that time. Just before the first big gap in useful data (Sept. 14 22:50h) and just after this gap (Sept. 15 9:50h), it can be seen that the areas of the seepage spots become really big. This is possibly caused by a too small temperature difference in the ditch, as can be seen in Figure 5.9. Furthermore, in this part of the data, it was found that the pump which kept the water level in the ditch constant had a big influence on the size of the areas. Starting at September 16 12:00h up to 18:00h, it was found that the sun was having a big influence on the data. Half of the ditch became a different temperature due to shadow and the other half due to the sun. Since these parts of the data are therefore not very reliable and suitable for the correlation analysis, it is chosen to focus on the data part starting at 15 September 21:20h up to 16 September 13:00h. In the Figure a likewise path of the lines can be seen for this part, which is therefore promising for a correlation analysis. Since this part is mainly during the night, there was no big influence of the sun and no big influence of people standing in front of the ditch.

To be able to compare the part even better, a new Figure 5.15 is made where the part starting at 15 September 21:20h up to 16 September 13:00h is zoomed into and the y-axis are normalized. This was the period in which the water level in the infiltration tubes was decreased and increased so that continued pipe growth would not happen during the night. In this Figure, it can be seen even better that the lines follow a likewise path. Furthermore a delay can be seen between the flow rate and the areas. This is explainable because the flow rate is the amount of flow going into the infiltration tubes. Logically, it takes time before this flow arrives at the ditch at the toe of the dyke. In the data set where Figure 3.4 (the incoming flow rate versus the outgoing flow rate) is based on, it can be seen that the delay in the incoming and outgoing flow rate is approximately one hour. In Figure 3.4 the delay between the flow rate and the areas of the seepage spots seems to be up to 2 hours, but this can be caused by the temperature distribution in the ditch or other reasons as well. Therefore, it is chosen to base the delay on the difference in incoming and outgoing flow rate and take a shift of one hour of the flow rate into account when looking for the best correlation in the next section.

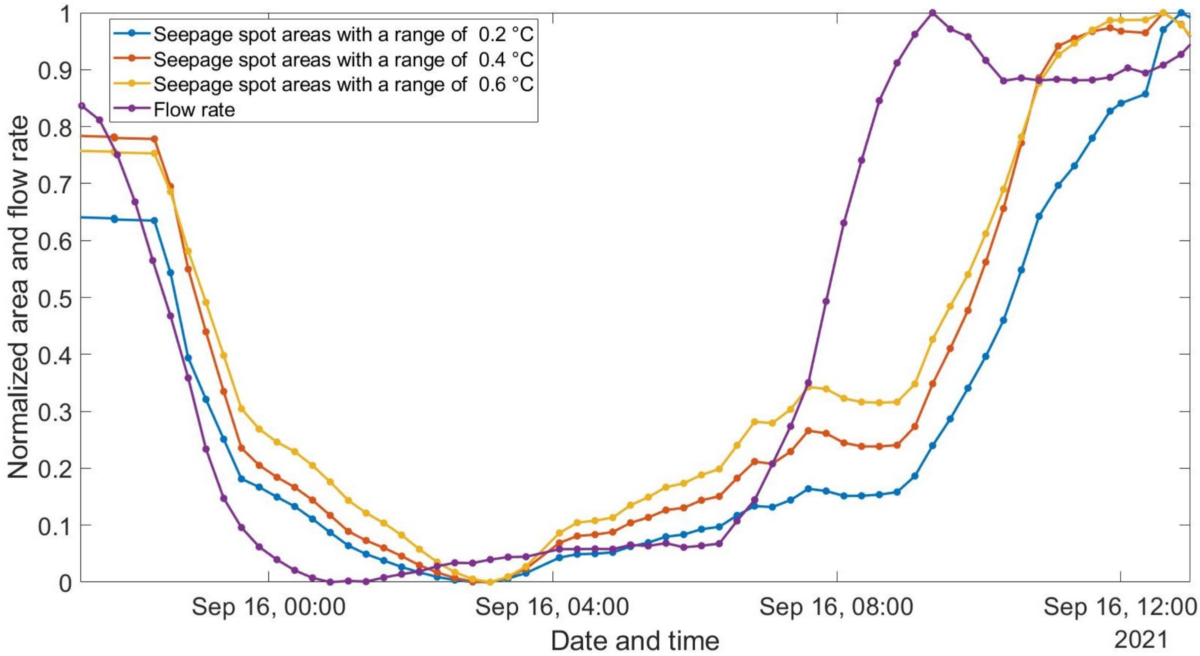


Figure 5.15: Normalized flow rate and areas for the time period of 15 September 21:20h up to 16 September 13:00h.

Fit the timestamps

The data part shown in Figure 5.15 are made of the same length and frequency by converting the areas into data points of once every 15 minutes. This resulted in 4 data sets of 63 data points: 1) Flow rate, 2) area with a range of 0.2 °C, 3) area with a range of 0.4 °C and 4) area with a range of 0.6 °C. The data points are added in Figure 5.15 as dots on the line.

5.2.2 Correlation Analysis

In this section, the results of the correlation analysis are shown. Scatter plots of the data are created and the correlation coefficients are given for the different scenarios. After that, the significance of the correlation coefficients is given.

Scatter plot and correlation coefficient

The correlation coefficients and scatter plots are calculated and created for 6 scenarios of three different temperature ranges. This is done for the original time series and a time series where the flow rate is shifted one hour in time to cover the delay of the incoming flow rate versus outgoing flow rate. The resulting correlation coefficients per scenario for the Pearson linear correlation test are shown in Table 5.1. The resulting correlation coefficients per scenario for the Spearman non-linear correlation test are shown in Table 5.2. In the case of the linear correlation coefficient, it can be seen that the correlation is stronger between the flow rate and a larger used temperature range, and even stronger when the time series of the flow rate are shifted. In the case of the non-linear correlation coefficient, it can be seen that the correlation is the strongest for the shifted time series and a temperature range of 0.2 °C. All scenarios and scatter plots can be seen in Appendix F. In Figure 5.16 the scenario with the highest correlation coefficient can be seen: the non-linear Spearman correlation for a shifted time series and a temperature range of 0.2 °C. Here can be seen that for small areas, the flow rate does not grow very fast. For areas between $\pm 0.2 m^2$ and $\pm 0.5 m^2$ the flow rate grows very fast. For areas higher than $\pm 0.5 m^2$, the flow rate does not grow fast and decreases a little in the end.

Table 5.1: Correlation coefficients Pearson (linear correlation)

Range	**Correlation Coefficient original time series**	**Correlation Coefficient shifted time series**
0.2 °C	0.776	0.878
0.4 °C	0.813	0.917
0.6 °C	0.833	0.923

Table 5.2: Correlation coefficients Spearman (non-linear correlation)

Range	**Correlation Coefficient original time series**	**Correlation Coefficient shifted time series**
0.2 °C	0.835	0.952
0.4 °C	0.853	0.945
0.6 °C	0.848	0.947

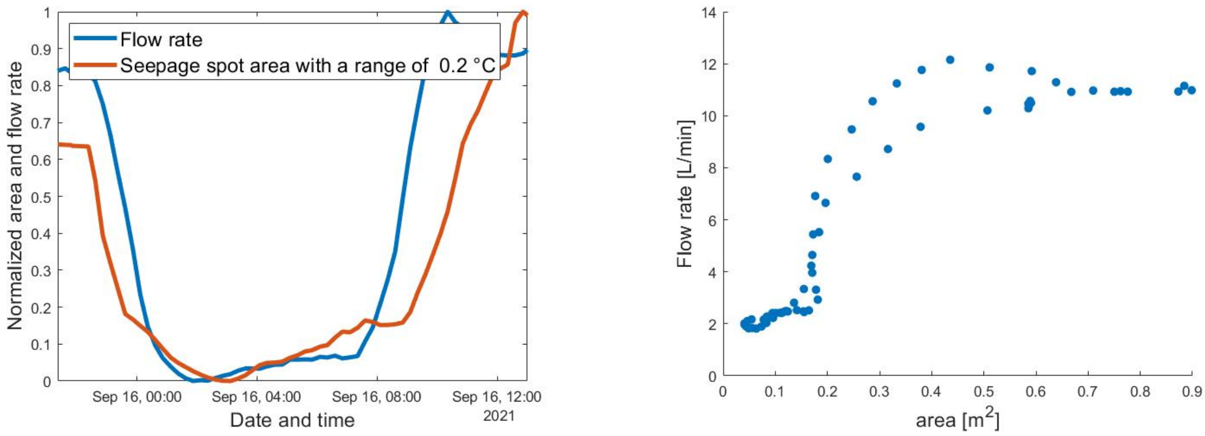


Figure 5.16: Scenario where the area with a range of 0.2 °C is correlated with the shifted flow rate. Left: a normalized plot of the flow rate and area in time. Right: a scatter plot of the flow rate versus the areas.

Significance

The p-values for the Pearson linear and Spearman non-linear correlation analysis are found to be < 0.0001 for every scenario. When this is compared with the significance level of 5 %, it can be said that the p-values are less than the significance level for all scenarios and therefore the null hypothesis can be rejected. This means that the relations found with the correlation coefficients are found with statistical evidence (Davis, 2002).

5.3 Regression analysis (O3)

The regression analysis is done for the data sets with the highest correlation coefficient. The highest correlation coefficient is found for the non-linear Spearman coefficient and is found between the shifted flow rate and the areas with a range of 0.2 °C. The curve that represents the data the best is a third degree polynomial calculated in MATLAB. The resulting regression line is shown in Figure 5.17. The 95% confidence interval is shown as well. Within this interval, one can be 95% sure that the flow rate is the correct value (according to this sample).

The shape of the curve is what is expected. The flow rate increases slower at areas which are bigger and faster at areas that are smaller. So a big flow rate makes the seepage spot spread more than a small flow rate. In Figure 5.1 this is seen, the regression line is steep at the beginning and more gentle at the end.

The equation corresponding to the regression line is given in equation 5.1. With this equation, an area found in a thermal image (x) can be converted into an approximation of the flow rate coming out of the sand boil (y). The equation is fitted on a data set for areas ranging between 0 and 0.9 m^2. Therefore, this equation can only be used for areas between that range.

$$y = 12.5x^3 - 45.0x^2 + 43.2x - 1.1 + \epsilon \tag{5.1}$$

Where:
y = the predicted flow rate [L/min];
x = the area of the seepage spot in the thermal image [m^2];
ϵ = the error term.

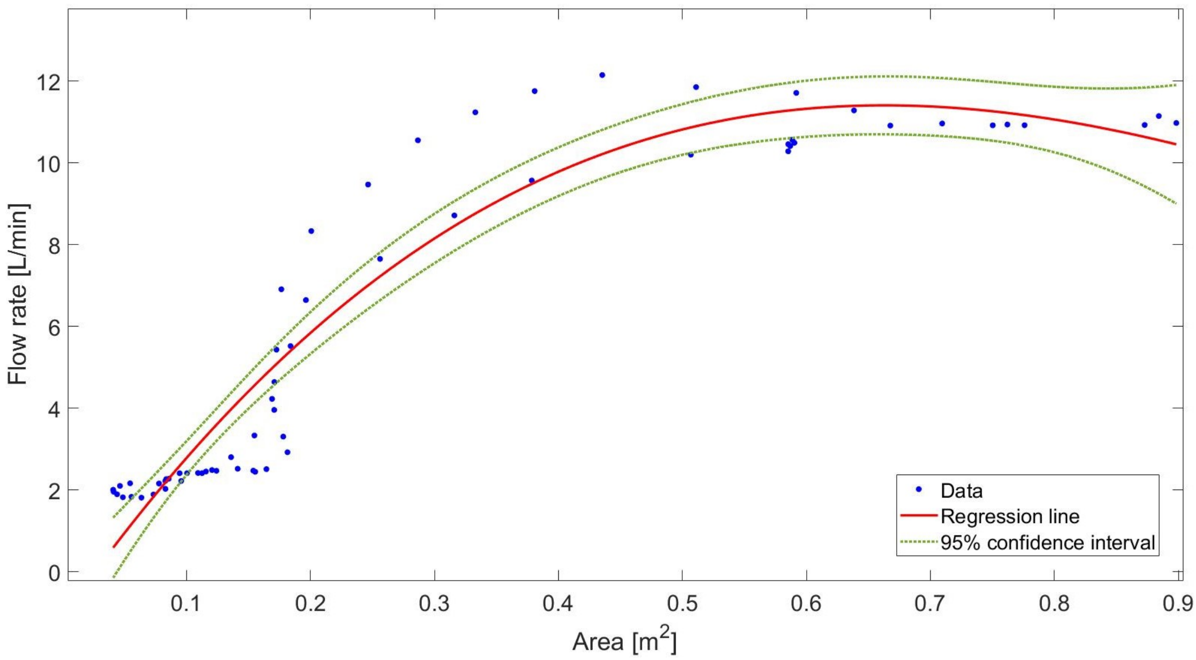


Figure 5.17: A third degree polynomial Regression line that represents the best correlating data set. The 95% confidence interval of the regression line is given as well.

6 | Discussion

The aim of this study is to determine to what extent thermal imaging can be used to make a flow rate prediction of a seepage spot. This has been determined by three analyses: a data analysis of the thermal images, a correlation analysis between the flow rate and the size of the seepage spots in the thermal images, and a regression analysis for the best correlation.

The results of this research are explained and interpreted in the first section of this chapter. After that, the limitations of the used data set are given. Lastly, the implementation in practice is discussed.

6.1 Explanation and interpretation of the results

The results of the research are discussed here and are divided in the subsections data analysis and correlation & regression analysis.

6.1.1 Data analysis

The first part of the data analysis (detection of seepage in the thermal images) is performed with the use of histograms with fitted curves. An image analysis based on histograms is a method that is often used. For example Li et al. (2012) uses histograms to detect pedestrians in thermal images. Local intensities and a larger range of temperatures in the histograms showed the presence of pedestrians. Another example is Salgadoe et al. (2019), who developed a method to extract an Orchard tree in a thermal image from its surrounding. Their research found that histogram gradient thresholding based on temperature intensity changes was the best method to perform this. A high peak in the histogram corresponded to the tree and made it able to extract it.

The results of this research show that it was not possible to automate the whole data analysis for this data set yet. Four actions require human effort:

1. The identification of the regions of interest;
2. The identification of obstacles;
3. The identification of the used temperature ranges;
4. The identification of the location of the seepage spot where the seepage temperature is extracted from.

It was not possible to automate these actions yet since the thermal behaviour of piping was found to be very unpredictable in the context of this research. Complicating factors proved to be the differences in cold and warm seepage, mixing processes with the water in the ditch, weather events, the location of the active seepage spot, and the seepage temperature. To handle these complicating factors in the context of this research, human validation is embedded in the method to ensure the right area is selected as the seepage area. Doing this by sampling in the time series should be sufficient. Kuchi et al. (2019) show that another possibility to detect seepage spots in thermal images is by the use of deep learning. Deep learning can be a solution to make the data analysis process fully automated. Unfortunately, this method was outside of the scope and available time of this research.

Application of the data analysis method on different data sets requires analysis to determine to what extend the method is valid, considering the region of interest, obstacles and characteristics of the ditch (water depth, dimensions) which can be different for an other data set. For now, the development of the data analysis method is based on a specific LLHPP experiment and the method is not validated on other experiments. The used temperature ranges in the case of sand boils in the grass should are expected to be different, since mixing with surrounding water is not likely to occur in that situation. Moreover, the water in the ditch of this case study is kept at a water level of 10 centimeters, while in other surroundings, this water level is expected to be larger. This can cause very different mixing of the water temperatures as well, and therefore the temperature ranges suitable for that situation can be significantly different. These should be determined for that specific situation.

In the data set of the LLHPP experiments, the difference in temperature of the seepage water and the ditch water is very small, especially during the shifting point from cold to warm seepage. This varies from a difference of ±0.1 °C to ±6 °C. The small temperature difference made it a challenge to quantify the size of the seepage spot in the thermal image. In other experiments, it has been found that the seepage temperature differs more from the ditch temperature of minimal 3 °C (Wiggers et al. (2019); Chen et al. (2018); Rösingh et al. (2017); Hafkamp (2021)). Therefore, in case of a critical situation with a high water level against the dyke, it is expected that temperature differences between the seepage water and the surrounding water are clearly visible.

During the data analysis of this research it was chosen to focus on the images with a clear temperature difference. In further research, the choice of the extraction of images with a too small temperature difference by a maximum peak height in the curves which were fitted on the histograms (>3) could be optimised. It can be that for example a peak height of the probability density of 2.5 in some cases would show a clear enough temperature difference as well. The value that is used now, worked the best for the majority of the data set of this experiment.
Another possibility would be to make a different data analysis part for the data with smaller temperature differences, with the use of smaller temperature ranges for the extraction of the size of the seepage spot.

6.1.2 Correlation & regression analysis

The correlation of the areas with the flow rate resulted in a well-correlated data part for the time period of 15 September 21:20h up to 16 September 13:00h. There was a better non-linear correlation than linear correlation and a stronger relation when the delay of the incoming flow rate was taken into account. This can be explained by what is seen in the scatter plots, the flow rate increases slower at areas which are bigger than $\pm 0.8 m^2$, and faster at areas that are within a range of $\pm 0.2 m^2$ and $\pm 0.8 m^2$. So apparently a big flow rate makes the temperature of the seepage spot spread more than a small flow rate, as one would expected.

Furthermore, it is seen that the data sets does not correlate well for the whole time series. There is insufficient evidence data to be sure about the reason for this. However multiple processes are expected to possibly cause this:

- The identification of the seepage spot sizes for these parts could not work so well for images with smaller temperature differences between seepage and ditch temperature. So in the step 'Detection of seepage in the thermal images', a lower probability density peak could have been more suitable. Or the extracted temperature ranges should be adapted for smaller temperature ranges.

- The correlation is performed using one flow rate for the whole experiment, while multiple sand boils were active. The flow rate is thus not connected to a specific sand boil.
- There are multiple sand boils active. The smaller sand boils have not been active for the whole time series, while the biggest sand boil has been. In the strongly correlating part, the biggest sand boil was mainly active.
- The pump which maintains the water level in the ditch, stretches the area of the seepage spots at the moment it is switched on. This pump switched on automatically at moments when the water level was too high.
- People are in front of the ditch for a big part of the data, making these data parts not suitable for the extraction of the size of seepage spots.
- During the day, the sun has a big influence on the data, causing temperature differences between parts of the ditch in the shadow and parts of the ditch being in the sunlight. This can cause the script to extract a different seepage spot size than there actually was.
- The low water level of 0.10 meter is very sensitive to sudden temperature changes by for example the sun as explained in the previous point.

The part of the data for which a strong correlation was found, had the following circumstances:

- It was mainly during the night, so there was little influence on the data by sunlight and people in front of the ditch.
- One sand boil was mainly active and had the biggest influence on the seepage spot and thus on the flow rate.
- The correlating part is exactly the part where the flow rate was decreased during night and increased the following day. This gives the indication that a cause of the correlation is possibly due to that process. Due to the decrease of the flow rate, the pressure on the aquifer decreases. The flow rate that does flow through the aquifer to the ditch will take the path with the lowest resistance. When (part of) a pipe is present already, which is the case here, the path with the lowest resistance will be the pipe. Therefore, it is expected that for this part of the data the flow rate is the most accurate flow rate that flows through the specific sand boil.

The regression line is fitted on the data set with areas between 0 and 0.9 m^2 and therefore not suitable to be used in situations with larger areas. A regression analysis on a data set with larger areas should be performed to validate this regression line and expand it to larger areas.

6.2 Limitations of the data set

In this section, the limitations of the data set which is used for this study are discussed. A division is made between thermal imaging data, flow rate data and the correlation analysis.

6.2.1 Thermal imaging data

The data set which is used in this research is a full-scale experiment in a controlled field environment. The focus of the laboratory experiment at the LLHPP was less on thermal imaging measurement than on other measurements. This has hampered the quality of the thermal images in some way. When the focus is on capturing data with thermal imaging, one can make sure that the images are captured perpendicular to the region of interest, no obstacles are in front of this region, no people walk in front of the images and the sunlight is not too bright on the region of interest.

In the experiment, a pump is used that extracts the water from the ditch with the purpose to keep the water level of 10 centimeters. This possibly deformed the areas of the seepage spots in the thermal images unrealistically. However, it can as well provide a realistic situation, since in a real life situation the current in the ditch can deform a seepage spot area as well (van de Voort et al., 2021).

All sand boils found in this study occur to be on the edges of the ditch. Other results are possible when a sand boil would occur in the middle of the ditch, as can be seen in Figure 6.1. The area of the seepage spot in the center can be bigger due to more mixing. Therefore, it would be good to have situations where the sand boil is in the middle of the ditch in the data set as well. However, one can not influence the place where the sand boil occurs in an experiment. Still, the sand boil being on the edge of the ditch can occur in practice as well (van de Voort et al., 2021).

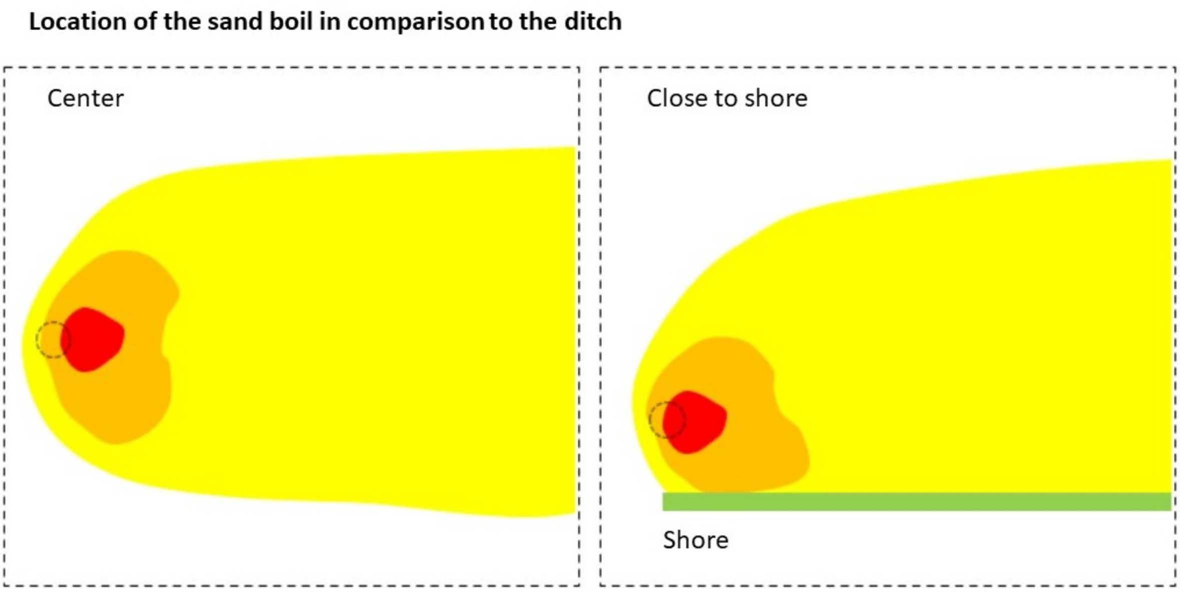


Figure 6.1: Sand boil location in comparison to the ditch shore, resulting in different areas in the thermal image. Adapted from van de Voort et al. (2021).

6.2.2 Flow rate data

For the experiment at the LLHPP, it was planned to measure the outgoing flow rate automatically by a measurement system. However, during the experiment, this measurement device did not work. At that moment, the decision was made to start the measurement by hand. This resulted in the flow rate measurements as shown in Figure 3.4 in Section 5.2. As can be seen, there were less frequent measurements due to this hand-measurement than the automated measurements as done at the incoming flow rate. Still, it is found that the incoming and outgoing flow rate follow approximately the same quantities and therefore the incoming flow rate could be used.

The flow rate was not measured for a specific sand boil but for the whole ditch/system. It is therefore needed to make an estimation on how big the contribution of a sand boil was to the flow rate, or to take the areas of the sand boils together and analyse the flow rates in that way. The second option is chosen in this research, since the first option would need a too big assumption to perform a correlation analysis. It is expected that accurate data on the flow rate of a specific sand boil will increase the reliability of the correlation found in this research.

6.2.3 Correlation analysis

The correlation analysis requires an equal number of data points for flow rate and areas. To meet this requirement the number of data points for the area are reduced, this way assumptions for interpolation are avoided. Consequently, the number of data points reduced to 63.

6.3 Implementation in practice

In this section, the possibilities of using the results of this research in practice are discussed.

In other studies, it has already been shown that thermal imaging can be used to detect the location of a seepage spot (Zhou et al. (2022); Chen et al. (2018); Wiggers et al. (2019)). The difference between a seepage spot and sand boil can not be seen in a thermal image, since the occurrence of sand transportation doesn't change the thermal image. It is important to check if the area which can be seen in the thermal image is a seepage spot or a sand boil, since a seepage spot does not have to be dangerous while a sand boil has a high risk of resulting in piping (van Beek, 2015). A sand boil indicates erosion occurring in the dyke. Therefore human interference is always needed in the assessment of the dyke for piping. Thus, thermal imaging is a complementary tool to enhance the quality of the dyke assessment rather than a stand alone tool. A big advantage is that the thermal images show the progression of the sand boil very clear during night, when the human eye can see less.

There are different platforms with different application possibilities for thermal imaging. In this study, an off road vehicle is used. Other possibilities are a hand-held device, drone, quad, airplane/helicopter or a satellite (Rösingh et al., 2017). When an off road vehicle is used, a mast is attached so that the camera is as perpendicular to the region of interest as possible. With this platform, stationary images of a specific location can be captured making an off road vehicle ideal for monitoring a specific sand boil. With a quad, one can reach more inaccessible locations which can not be reached by car. However, there is less space for a stable mast or for analysis equipment. A hand-held device has the advantage that one can carry it with them and make a direct measurement of a suspicious location. A drone can be used to capture the situation of a larger search area and has a large reachability independent of vegetation and accessibility. A disadvantage is that it is dependent on the wind, when there is too much wind, the drone can not capture stable images. It can specifically be used to detect locations of seepage. When the location is found, the situation can be checked and an off road vehicle can be used to monitor the specific sand boil. A helicopter or airplane can cover an even larger search area. An advantage is the speed with which large ranges of dykes can be measured. An airplane can fly in more extreme weather circumstances than a drone. However, the costs are much higher than the use of the other platforms and a higher resolution camera is needed due to the large distance. Lastly, a satellite can be used. However, this platform can capture images with a resolution of 100 by 100 meter so far, which is too large to detect seepage. An advantage is that a satellite captures data at a very constant recurrence (Rösingh et al., 2017). So, to estimate the flow rate of a sand boil in a thermal image as done in this research, the most suitable platforms are an off road vehicle, quad, handheld device or drone since they give accurate data with which one can extract a specific sand boil area.

The regression analysis that is performed in this study, resulted in an equation that can transform a found seepage spot area into the corresponding flow rate. This is an additional tool that can be used in practice when inspecting or guarding the dyke. When an inspector detects a sand boil, they can choose to monitor it with thermal imaging to be able to follow the progression or state of this sand boil. When using the data analysis method to extract the area and the results of the regression analysis to give an indication of the flow rate, growth or shrinkage of the pipe can be detected. This can indicate the severeness of the sand boil. So, the tool can be used in the care taking duty or assessment of the dykes as a support of the interpretation of what is seen. A disadvantage is that the critical flow rate of a sand boil is not known yet. Therefore a follow-up research on the critical flow rate can be a very helpful addition to be able to make an estimation of the severeness of a sand boil using thermal images. Still, with the information we have now, one can monitor the change of a sand boil and have an indication of the increase or decrease of the flow rate. But it is recommended to perform this study on more data, to make the estimation of the relation between the flow rate and the area of the sand boil more reliable. When a followup study is performed with the use of flow rates specifically originating from one sand boil, a more accurate regression analysis can be done. In summary, the results of this study can be implemented in practice as an additional tool to complement the inspection of dykes for piping. It makes human observations more accurate, supports the interpretation of what is seen and can find critical seepage locations which human observations can miss upon, especially during night. It is a very promising method to be used as a complementary tool to enhance the quality of the dyke assessment.

7 | Conclusion & Recommendations

In this chapter, the conclusion and the recommendations of the research are presented.

7.1 Conclusion

The aim of this study is to determine to what extent thermal imaging can be used to make a flow rate prediction of a seepage spot.

This goal is achieved by dividing it in three objectives. Starting with a data analysis of the thermal images, and continuing with searching for a correlating part of the data set. After finding a correlation, a regression analysis is used to make an estimation of the flow rate of a seepage spot in a thermal image.

The main conclusion is that the flow rate prediction tool developed in this research can be used to complement the visual inspection of dykes for piping. The method makes human observations more accurate and can find critical points which human observations can miss upon, especially during night.

The conclusion per objective is given in the following subsections.

7.1.1 Data analysis of thermal images

The first objective is to develop a data analysis method of the thermal images to:

- Extract useful thermal imaging data from the data set, being thermal images in which a clear temperature difference is visible due to the occurrence of seepage;
- Estimate the sizes of the areas of the seepage spots from these data.

A data analysis method is developed for the thermal imaging data set, divided into three main components: detection of seepage in the thermal images, identification of the location of the seepage spots and quantification of the size of the seepage spots. The first component (detection of seepage in the thermal images) extracts the useful thermal imaging data from the data set. The second and third component (identification of the location of the seepage spots and quantification of the size of the seepage spots) extract the sizes of the areas of the found sand boils from these data. Unfortunately, the analysis could not be made fully automated yet for this data set. Four moments of manual action are needed:

1. The identification of the regions of interest;
2. The identification of obstacles;
3. The identification of the used temperature ranges;
4. The identification of the location of the seepage spot.

So, a data analysis method on the thermal images is developed which extracts the areas of the seepage spots from the thermal images with a clear temperature difference due to seepage. Temperature differences in the thermal images are the clearest when there is no (big) influence of the sunlight on the region of interest.

7.1.2 Quantification of a possible correlation

The second objective is to quantify a possible correlation between the flow rate of the sand boil and the size of the seepage spot in the thermal images.

A correlation coefficient is calculated for 6 scenarios of three different temperature ranges and with the original time series, and a time series where the flow rate is shifted one hour in time to cover the delay of the incoming flow rate versus outgoing flow rate. Furthermore, a linear (Pearson) and non-linear (Spearman) correlation are compared. The scenario with the strongest correlation coefficient of 0.952 is the non-linear Spearman correlation for a shifted time series and a temperature range of 0.2 °C. This relation has been tested for significance. The test concluded that the relation is found with statistical evidence. So, a non-linear relation is identified between the flow rate of the sand boil and the area of the seepage spots in the thermal images. The correlation that is found is promising: this data had little influence of sunlight, there were no people in the camera view and one sand boil was mainly active, making it a reliable part of data.

7.1.3 Regression analysis

The third objective is to develop a practical tool to enable flow rate prediction based on the seepage spot size from the thermal images.

A regression analysis is performed which developed an equation to transform a seepage spot area in the thermal image to an indication of a flow rate coming out of that possible sand boil. It is found that a third degree polynomial function fitted the data set the best. The resulting equation is $y = 12.5x^3 - 45.0x^2 + 43.2x - 1.1 + \epsilon$, with a range for x (seepage spot area) of $[0 - 0.9m^2]$. So, a practical tool is developed which translates the relation found in objective 2. With this tool, one can predict the flow rate of a seepage spot in a thermal image, if the seepage spot is in between the range of 0 and 0.9 m^2. At this moment, the tool is useful as a complementary tool to enhance the quality of the dyke assessment rather than a stand alone tool. It can support interpretation of the dyke inspector in the care taking duty and assessment of a dyke and can see more than the human eye (especially during night). So, the human dyke inspection and the thermal imaging can complement each other perfectly.

7.2 Recommendations

The first recommendation is to perform this research on an other data set to validate the results of the regression analysis. It is recommended to perform an experiment in which the flow rate of a specific sand boil is measured, since there are now three sand boils taken together in the analysis. In this way the relation between the flow rate and the area of the sand boil in the thermal image can be made more accurate and the reliability of the results of this study can be improved.

Furthermore, to be able to make the translation to practice even better, it is recommended to perform a research to the critical flow rate of a sand boil. When the critical flow rate is known, one can determine if the sand boil is at a critical point based on the method developed in this research, while now one can only monitor the increase or decrease of the flow rate.

Moreover, it is recommended to perform the research on other test sites as well. For example a ditch with a larger water depth is expected to behave differently. Another recommendation is to perform a research on the mixing of the temperature of water seeping into other water. With the results of such a test, the temperature ranges of a seepage spot can be predicted more accurately.

Another recommendation is to automate the four steps that have now been performed manually. These steps are 1) The identification of the regions of interest; 2) the identification of obstacles; 3) the identification of the used temperature ranges and 4) the identification of the location of the seepage spot where the seepage temperature is extracted from. A method that is recommended to investigate is deep learning. This method is able to recognise prevalent processes by learning this from example data sets.

The focus of the experiment at the LLHPP was less on the thermal imaging measurement than other measurements. Some recommendations can be given on gathering higher quality thermal imaging data:

- Be aware of walking in front of the region of interest.
- Be aware of other obstacles in front of the region of interest.
- The sunlight has a negative influence on the quality of the thermal images. The best images are gathered during night. In case of measuring for a longer time there can be chosen to create constant shadow with a shadow cloth.
- Measure the thermal image as perpendicular to the region of interest as possible.

Bibliography

van Beek V, Bezuijen A, Sellmeijer H (2013) Erosion in Geomechanics Applied to Dams and Levees, John Wiley and Sons, chap 3 Backward Erosion Piping, pp 193–269. DOI 10.1002/9781118577165.ch3

van Beek VM (2015) Backward erosion piping: initiation and progression. DOI 10.4233/uuid:4b3ff166-b487-4f55-a710-2a2e00307311

Bersan S, Koelewijn A, Simonini P (2018) Effectiveness of distributed temperature measurements for early detection of piping in river embankments. *Hydrology and Earth System Sciences* 22(2):1491–1508, DOI 10.5194/hess-22-1491-2018

Bersan S, Koelewijn AR, Putti M, Simonini P (2019) Large-scale testing of distributed temperature sensing for early detection of piping. *Journal of Geotechnical and Geoenvironmental Engineering* 145(9):04019052, DOI 10.1061/(ASCE)GT.1943-5606.0002058

Chen CY, Chen SC, Chen KH, Liu ZH (2018) Thermal monitoring and analysis of the large-scale field earth-dam breach process. *Environmental Monitoring and Assessment* 190(483):1–17, DOI 10.1007/s10661-018-6869-y

Davis JC (2002) Statistics and Data Analysis in Geology Third edition. John Wiley Sons

Dhiraj, Agarwal A, Agrawal A, Meruane V, Sangwan K (2021) Development of a machine learning based model for damage detection, localization and quantification to extend structure life. *Procedia CIRP* 98:199–204, DOI 10.1016/j.procir.2021.01.030, URL https://www.sciencedirect.com/science/article/pii/S2212827121000536, the 28th CIRP Conference on Life Cycle Engineering, March 10 – 12, 2021, Jaipur, India

Donchyts G, Koelewijn A (2016) Deltafact: Remote sensing voor waterveiligheid. URL https://www.stowa.nl/sites/default/files/assets/DELTAFACTS/Deltafacts%20NL%20PDF/Remote%20Sensing%20waterkwantiteits-kwaliteitsbeheer_DEF-converted.pdf

Dong CZ, Catbas FN (2021) A review of computer vision–based structural health monitoring at local and global levels. *Structural Health Monitoring* 20(2):692–743, DOI 10.1177/1475921720935585

Downey A, D'Alessandro A, Baquera M, García-Macías E, Rolfes D, Ubertini F, Laflamme S, Castro-Triguero R (2017) Damage detection, localization and quantification in conductive smart concrete structures using a resistor mesh model. *Engineering Structures* 148:924–935, DOI 10.1016/j.engstruct.2017.07.022, URL https://www.sciencedirect.com/science/article/pii/S014102961731115X

Gogtay NJ, Thatte UM (2017) Principles of correlation analysis. *Journal of The Association of Physicians of India* 65(3):78–81

Hafkamp R (2021) Weldetectie en welmonitoring met infraroodbeelden (afstuderen onderzoeksrapport). URL www.witteveenbos.com van der Ham W (2006) Watersnoodramp van 1953 was te voorkomen. *Tijdschrift voor Waterstaatsgeschiedenis* 12(2003):21–31

t Hart R (2018) Fenomenologische beschrijving faalmechanismen: Faalmechanismen wbi. URL `www.helpdeskwater.nl`

Hijma M, van Goor G (2021) Pipingproef hedwigepolder

Hoeksema RJ (2007) Three stages in the history of land reclamation in the netherlands. *Irrigation and Drainage* 56:S113–S126, DOI 10.1002/ird.340

Huang T (1996) Computer vision: Evolution and promise. pp 21–25, DOI 10.5170/CERN-1996-008.21

Inman M (2010) Working with water. *Nature reports Climate Change* 4:39–41, DOI 10.1038/climate.2010.28

Jiang L, Su J, Li X (2016) Hot spots detection of operating pv arrays through ir thermal image using method based on curve fitting of gray histogram. *MATEC web of conferences* 61(06017):1–4, DOI 10.1051/matecconf/20166106017

Kuchi A, Hoque MT, Abdelguerfi M, Flanagin MC (2019) Machine learning applications in detecting sand boils from images. *Array* 3:100012, DOI 10.1016/j.array.2019.100012

Larsen R, Ringgaard S, Overgaard K (2007) Localization and quantification of muscle damage by magnetic resonance imaging following step exercise in young women. *Scandinavian journal of medicine & science in sports* 17(1):76–83, DOI 10.1111/j.1600-0838.2006.00525.x

Li W, Zheng D, Zhao T, Yang M (2012) An effective approach to pedestrian detection in thermal imagery. In: 2012 8th International Conference on Natural Computation, IEEE, pp 325–329, DOI 10.1109/ICNC.2012.6234621

Ma L (2007) Condition monitoring in engineering asset management. In: Proceedings of APVC2007 12th Asia Pacific Vibration Conference, JSME, pp 1–16

Mes J, Reichart N (2015) Dike monitoring: Improving insight in actual strength of embankments. *Knowledge journal Water matters* 2, URL `https://www.h2o-watermatters.com/?ed=201511#`

Peeters P, Haelterman K, Visser KP (2013) About reinventing innovative technologies for levee monitoring. *Changing times: infrastructure development* 25:2121–2130

Rösingh O, Enschedé M, Groeneouwe I, Reichart N, Zomer W, Rinsema JG (2017) Infraroodmetingen bij dijken: Haalbaarheidsstudie voor de pov piping

Salgadoe ASA, Robson AJ, Lamb DW, Schneider D (2019) A non-reference temperature histogram method for determining tc from ground-based thermal imagery of orchard tree canopies. *Remote Sensing* 11(6):714, DOI 10.3390/rs11060714

Semmens SN, Zhou W (2019) Evaluation of environmental predictors for sand boil formation: Rhine–meuse delta, netherlands. *Environmental Earth Sciences* 78(457):1–11, DOI 10.1007/s12665-019-8464-0

Sjödahl P, Dahlin T, Johansson S (2009) Embankment dam seepage evaluation from resistivity monitoring data. *Near Surface Geophysics* 7:463–474, DOI 10.3997/2214-4609.20146268

Steenblik S (2021) Concept rapportage infrarood-monitoring pipingexperiment 2021

Talukdar P, Dey A (2019) Hydraulic failures of earthen dams and embankments. *Innovative Infrastructure Solutions* 4(42), DOI 10.1007/s41062-019-0229-9

Technical Advisory Committee (1999) Technical report on sand boils (piping). URL `http://resolver.tudelft.nl/uuid:f6d03006-7744-452e-8ff2-a4914f118184`

Tsanakas J, Botsaris P (2012) An infrared thermographic approach as a hot-spot detection tool for photovoltaic modules using image histogram and line profile analysis. *International Journal of Condition Monitoring* 2(1):22–30, DOI 10.1784/204764212800028842

Vermeer K, de Bruijne J, Heynert K, Nijhof A (2012) Flood control 2015: Five years of innovation in flood risk

van de Voort PH, van Broekhoven FJG, Naus FL (2021) Van infraroodbeeld naar actuele sterkte. URL `www.witteveenbos.com`

Vousdoukas MI, Mentaschi L, Hinkel J, Ward PJ, Mongelli I, Ciscar JC, Feyen L (2020) Economic motivation for raising coastal flood defenses in europe. *Nature communications* 11(2119):1–11, DOI 10.1038/s41467-020-15665-3

de Vries G, Koelewijn AR, Hopman V (2010) Ijkdijk full scale underseepage erosion (piping) test: Evaluation ofinnovative sensor technology. *Scour and erosion* pp 649–657, DOI 10.1061/41147(392)63

Wiggers A, Sanders M, Niemeijer H, Tonneijck M (2019) Pov pipingportaal: een publicatie van de pov piping. URL `https://issuu.com/pov-piping/docs/18-5853_rapportage_2_webpdf`

Zhou R, Su H, Wen Z (2022) Experimental study on leakage detection of grassed earth dam by passive infrared thermography. *NDT E International* 126:102583, DOI 10.1016/j.ndteint.2021.102583

Appendices

A | Case study

In Section 3.1 the experimental set-up of the case study is explained. In this appendix, Figure A.1 shows pictures of the experimental set-up at the Hedwige-ProsperPolder and Figure A.2 shows a sketch of the artificial dyke. Furthermore, Figure A.3 shows an example of a thermal image in combination with a RGB image. It can be seen with the indicated rectangles, that the sand boil can be seen in the thermal image as well as in the visual image.

Figure A.1: Real experimental setup at the LLHPP with at the left-hand side the front perspective (Steenblik, 2021) and on the right-hand side the side perspective (Hijma and van Goor, 2021).

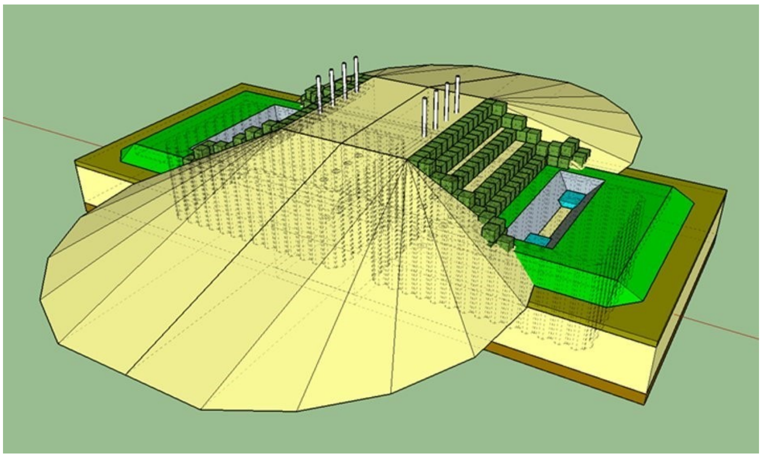


Figure A.2: Sketch of the experimental setup at the LLHPP (Hijma and van Goor, 2021)

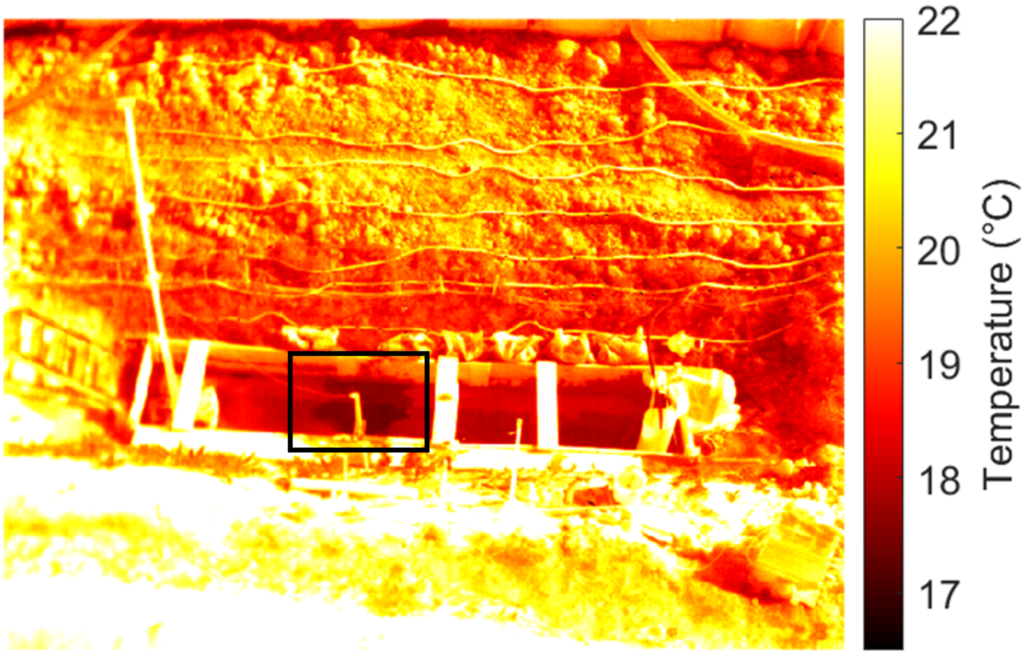


Figure A.3: A thermal (top) and RGB (bottom) image taken on 16 September 2021 at 15:54h. The black (top) and red (bottom) rectangle shows the location of a sand boil.

B | Data analysis of thermal images: People

In Section 5.1.1 the occurrence of people in front of ROIs is discussed. In this Appendix, an example of a thermal image with people in front of the ROIs can be seen in combination with the corresponding fitted curve in Figure B.1. It can be seen that the people are indeed above 20 °C and the temperatures inside the ROIs lower than 20 °C. In the fitted curve, the temperatures of these people are seen as well, as indicated with the red lined area in Figure B.1

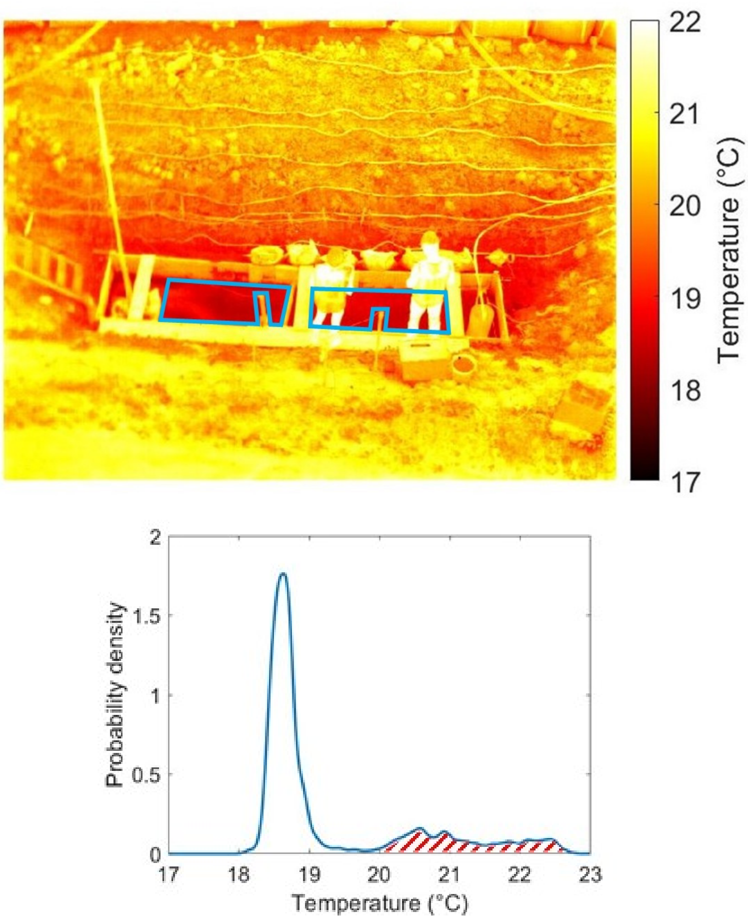


Figure B.1: A thermal image (top) and fitted curve (bottom) with people in front of the ROIs. The ROIs are indicated with blue polygons in the thermal image and the temperature of the people with red lines in the curve

C | Data analysis of thermal images: Visibility of temperature differences

In Section 5.1.1 the results for the visibility of temperature differences are discussed. In this Appendix, some examples of fitted curves in combination with the corresponding thermal image are given.

In Figure C.1, a low, wide curve can be seen for 14 September 20:00h. In the thermal image (Figure C.2a) it can indeed be seen that there are clear temperature differences inside of the ROIs. In Figure C.1, a curve can be seen that has a bit higher peak than the previous example for 15 September 0:01h, but it is still a low, wide curve. It can indeed be seen in the thermal image (Figure C.2b) that there are clear temperature differences which can be caused by the sand boil. For 15 September 6:00h (Figure C.1) a high, narrow peak can be seen. In the thermal image (Figure C.2c), it can indeed be seen that there is no clear temperature difference visible. For 15 September 12:00h a low, wide curve can be seen which shows clear temperature differences in the thermal image as well (Figure C.2d). 15 September 18:06h, shows a high peak, in which no clear temperature differences can be seen in the thermal image as well (Figure C.2e). 16 September 6:00h shows a low wide, peak at the end of the data set. In this thermal image, a clear temperature difference can be seen as well (Figure C.2f). More thermal images are checked as well.

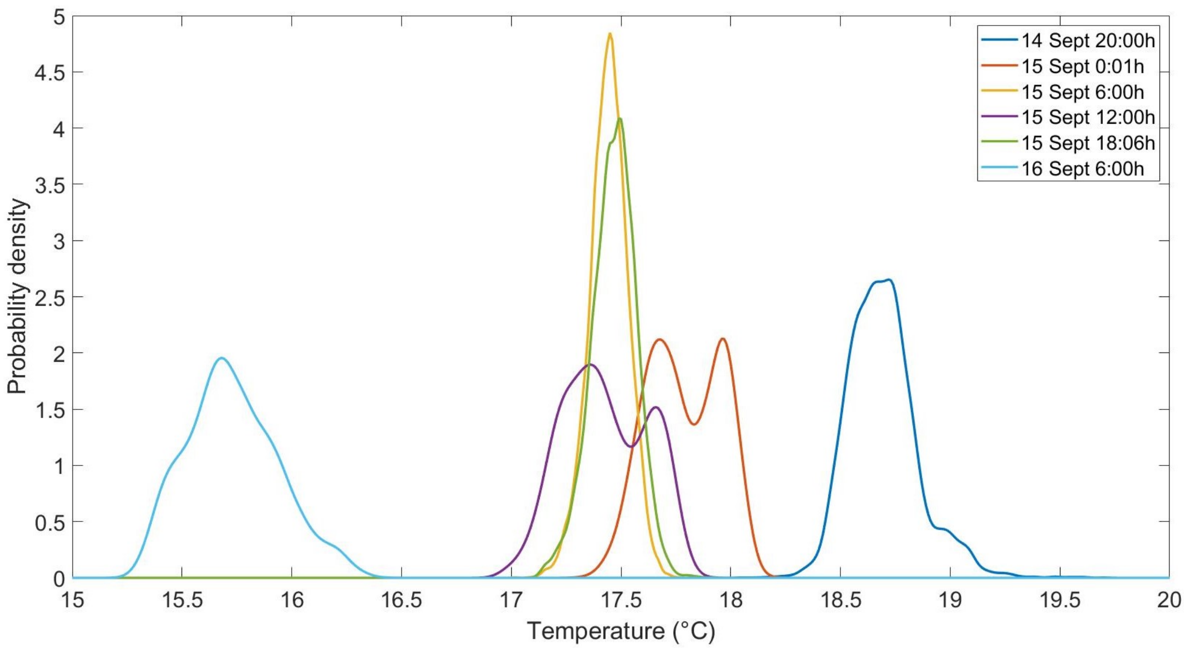
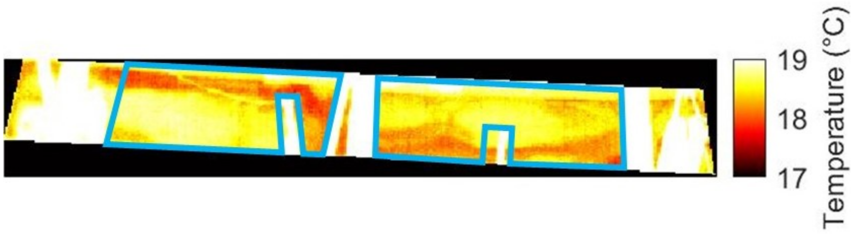
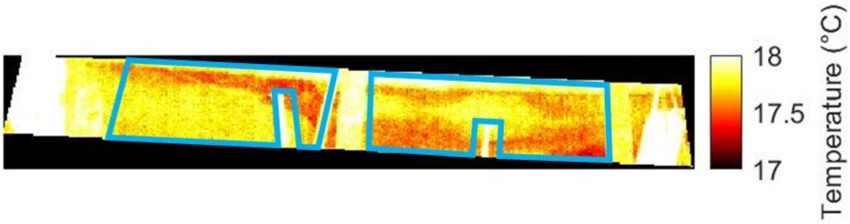


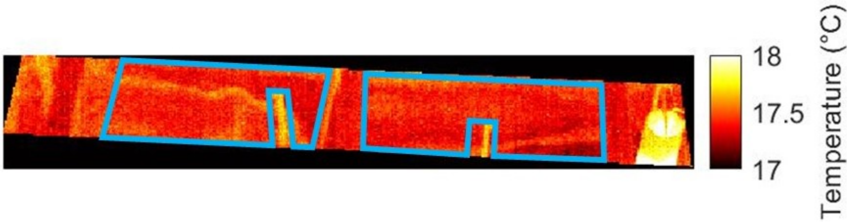
Figure C.1: Examples of fitted curves showing narrow ranges with high peaks and wide ranges with low peaks, the corresponding thermal images can be seen in Figure C.2a to Figure C.2f.



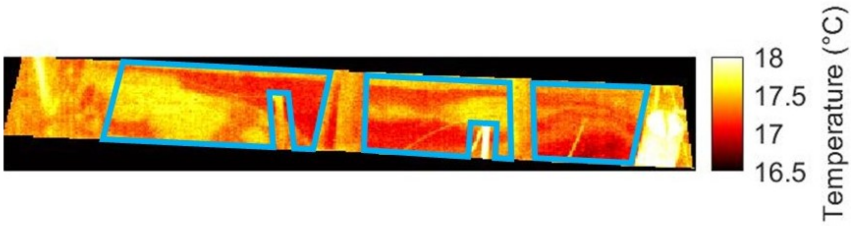
(a) Thermal image 14 September 20:00h



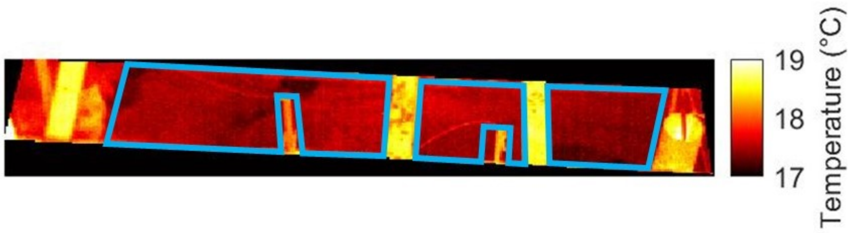
(b) Thermal image 15 September 00:01h



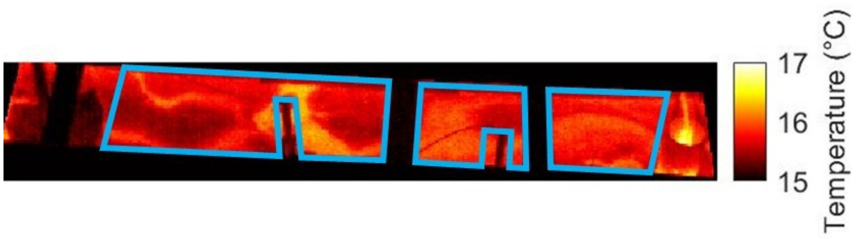
(c) Thermal image 15 September 6:00h



(d) Thermal image 15 September 12:00h



(e) Thermal image 15 September 18:06h



(f) Thermal image 16 September 6:00h

Figure C.2: Thermal images corresponding to the curves in Figure C.1, where the blue polygons indicate the Regions of interest.

D | Data analysis of thermal images: Temperature ranges

In Section 5.1.3, the results of the used temperature ranges are discussed. In this Appendix, some more examples are given in which more than one active sand boil can be seen (Figure D.1 and D.2).

Figure D.1: The quantified area for three different temperature ranges: 0.2 °C (top), 0.4 °C (middle) and 0.6 °C (bottom) where the yellow/orange/red colors indicate the pixels that are counted as the area.

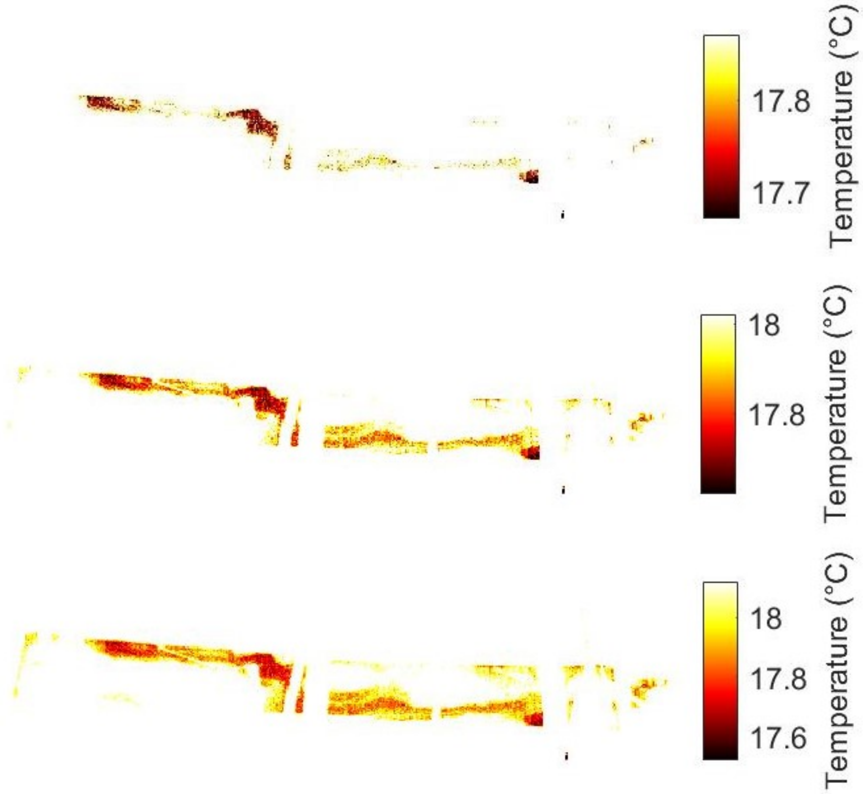


Figure D.2: The quantified area for three different temperature ranges: 0.2 °C (top), 0.4 °C (middle) and 0.6 °C (bottom) where the yellow/orange/red colors indicate the pixels that are counted as the area.

E | Smoothing of the correlation data

In Section 5.2.1 the data sets are smoothed. In this Appendix, the smoothed curves can be seen together with the original data points in Figure E.1 up to and including E.4.

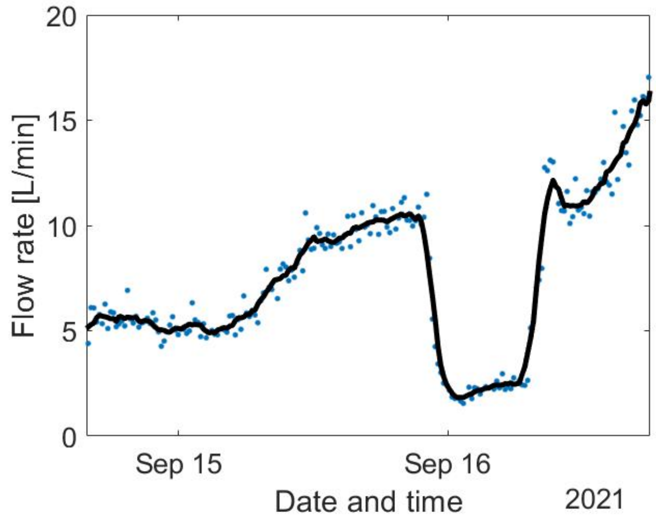


Figure E.1: Smoothed flow rate curve plotted with the original data points.

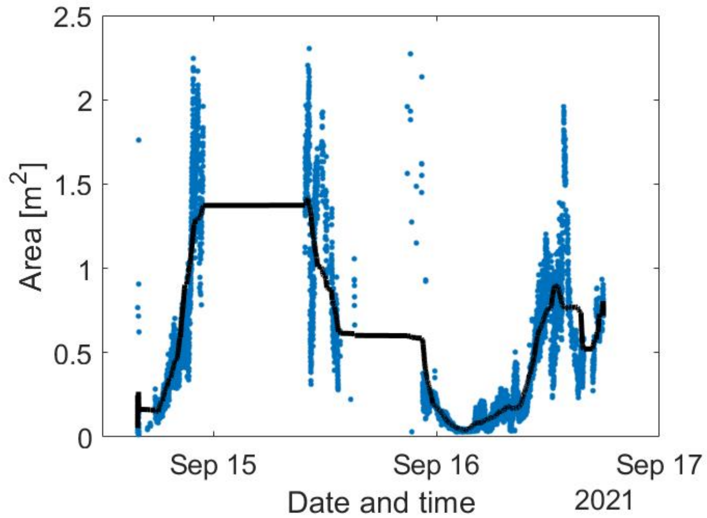


Figure E.2: Smoothed areas curve with a range of 0.2 °C plotted with the original data points.

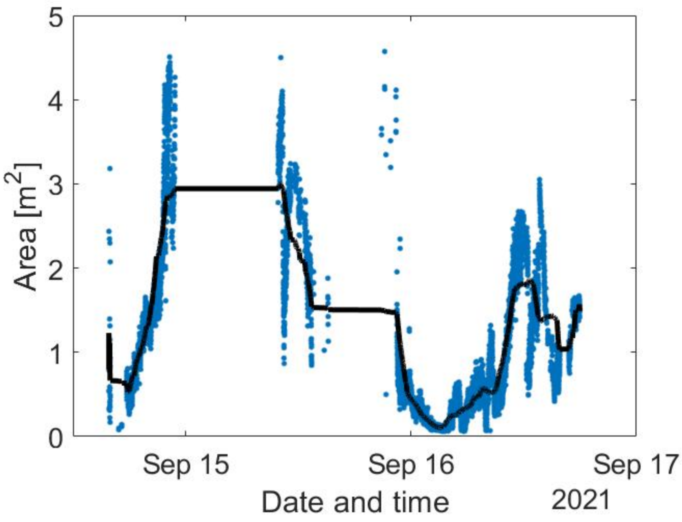


Figure E.3: Smoothed areas curve with a range of 0.4 °C plotted with the original data points.

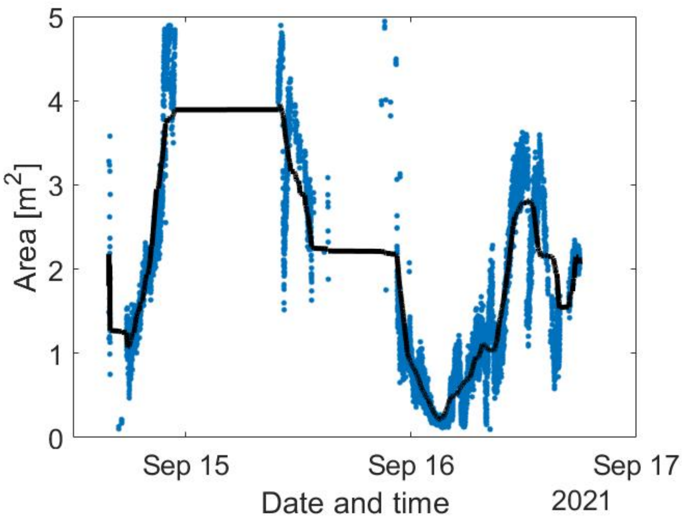


Figure E.4: Smoothed areas curve with a range of 0.6 °C plotted with the original data points.

F | Correlation analysis

In Section 5.2.2 the correlation coefficients of all six scenarios are given. In this Appendix, the scatter plots of the scenarios are given in Figure F.1 up to and including Figure F.6.

Original time series

Figure F.1, F.2 and F.3 show the scatter plots of the areas with different ranges against the original flow rate. This means that the delay of the incoming flow rate is not taken into account. In the scatter plots it can be seen that the data points are relatively far apart.

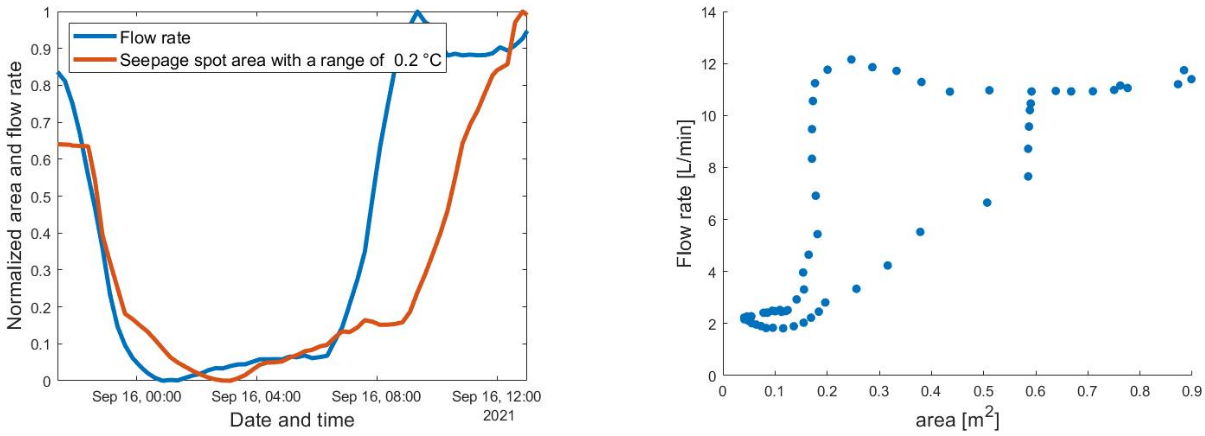


Figure F.1: Scenario where the area with a range of 0.2 °C is correlated with the original flow rate. Left: a normalized plot of the flow rate and area in time. Right: a scatter plot of the flow rate versus the areas.

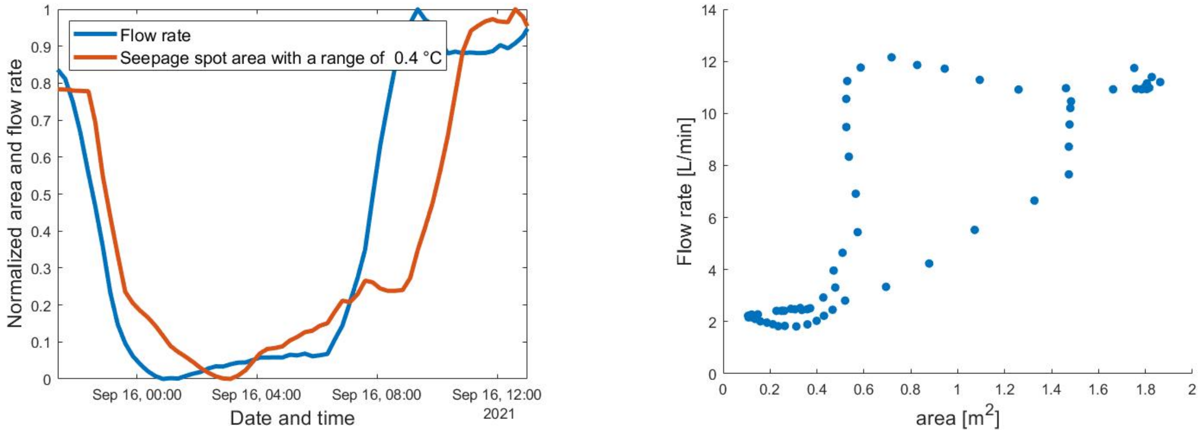


Figure F.2: Scenario where the area with a range of 0.4 °C is correlated with the original flow rate. Left: a normalized plot of the flow rate and area in time. Right: a scatter plot of the flow rate versus the areas.

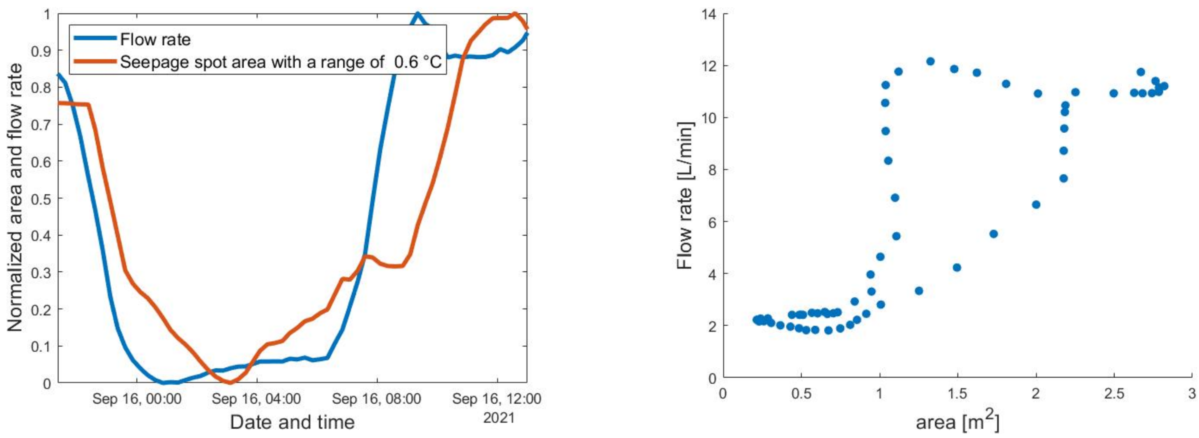


Figure F.3: Scenario where the area with a range of 0.6 °C is correlated with the original flow rate. Left: a normalized plot of the flow rate and area in time. Right: a scatter plot of the flow rate versus the areas.

Shifted time series

Figure F.4, F.5 and F.6 show the scatter plots of the areas with different ranges against the shifted flow rate. This means that the delay of the incoming flow rate is taken into account by shifting the flow rate forward in time with one hour. In the scatter plots it can be seen that the data points are more close to each other and seem like following a path in comparison to the scatter plots with the original time series. This indicates that a higher correlation is present for the shifted flow rate in which the delay is taken into account.

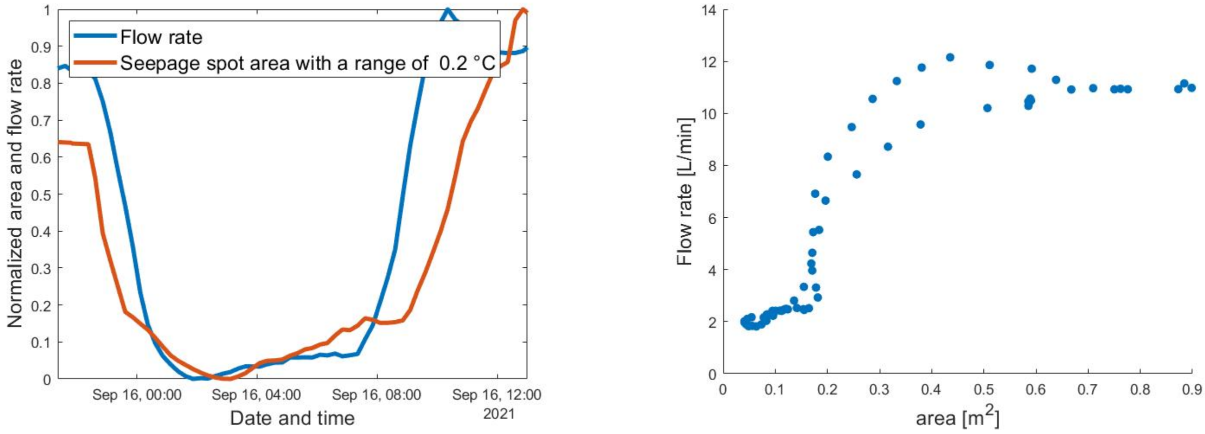


Figure F.4: Scenario where the area with a range of 0.2 °C is correlated with the shifted flow rate. Left: a normalized plot of the flow rate and area in time. Right: a scatter plot of the flow rate versus the areas.

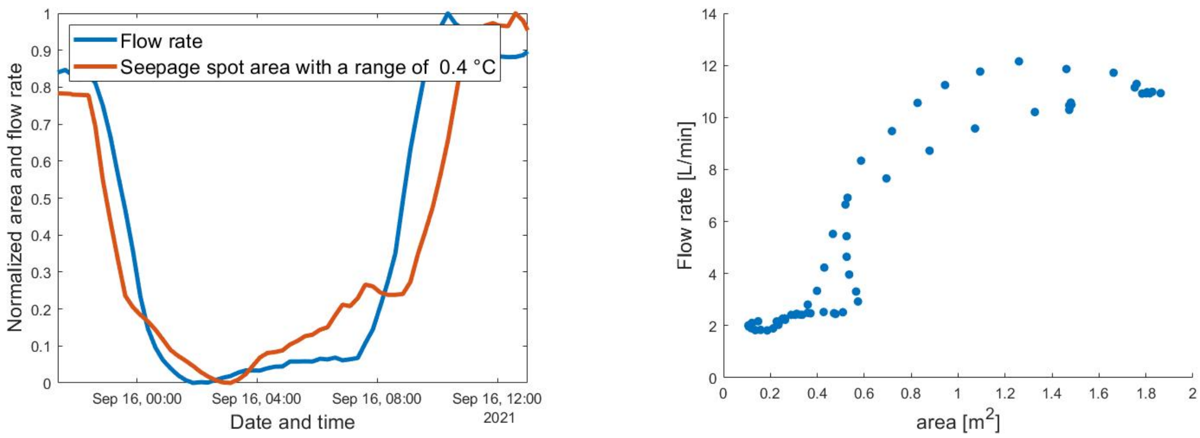


Figure F.5: Scenario where the area with a range of 0.4 °C is correlated with the shifted flow rate. Left: a normalized plot of the flow rate and area in time. Right: a scatter plot of the flow rate versus the areas.

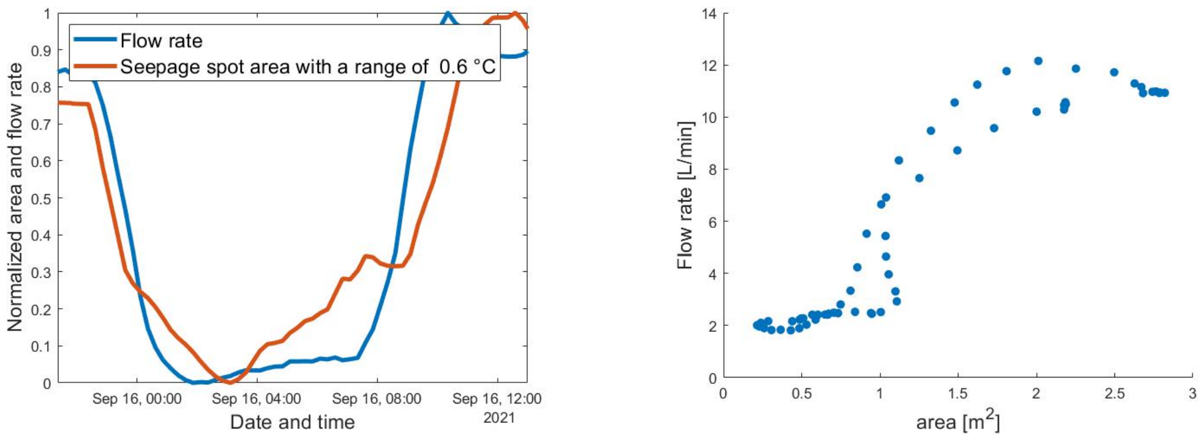


Figure F.6: Scenario where the area with a range of 0.6 °C is correlated with the shifted flow rate. Left: a normalized plot of the flow rate and area in time. Right: a scatter plot of the flow rate versus the areas.