HYDROLOGICAL VALIDATION OF GROUNDWATER LEVELS IN GROUNDWATER MONITORING NETWORKS

Jacob Belaiyneh (S2310732)

04/04/2024

Internal Supervisor: Anouk Bomers

Second Assessor: Alejandro Tirachini

External Supervisor: Emiel Groot

Preface

This report is the final product of my bachelor thesis research ‘Hydrological validation of groundwater levels in groundwater monitoring networks ‘. This thesis is written as part of my graduation from the bachelor’s programme of Civil Engineering at the University of Twente and was commissioned by BZ Ingenieurs & Managers (BZIM).

Conducting this research made me realize how interesting but also how broad the water section of civil engineering is. I am very thankful for the opportunity to conduct this research for BZIM to further develop myself. I really appreciate the kindness, the genuine interest, and willingness to help of the employees of BZIM. In particular, I would like to thank my external supervisor Emiel Groot and I would like to thank Anouk Bomers from the University of Twente for the help and guidance of my thesis project and from who I have learned a great deal in research practices and concise writing. Finally I would like to thank all the interviewees for making time in their schedules and helping me get the answers needed to conduct this research.

As of last, I hope that you enjoy reading my thesis and feel free to contact me if you have questions.

Jacob Belaiyneh

Enschede, 4 April 2024

Summary

In the Netherlands, managing groundwater effectively is essential, since groundwater is crucial for the Netherlands. Because it affects everything from drinking water supply to nature conservation and urban development. Monitoring groundwater levels is a key part of this effort, but it is becoming more complicated because of new laws and the large amounts of data being collected more frequently than ever before. Previously, observing and checking the accuracy of groundwater data was mostly done by hand, a detailed but increasingly slow method given the growing amount of data and the legal need to process this information quickly. Consequently, driven by the requirements of new legislation and advances in technology this research focuses on exploring ways to automate this process of checking groundwater data. Therefore during this research the new BRO legislation, ways of automating the validation process, hydrological validation and the current protocols are discussed. This research aims to gain insight into what steps can be automated in these validations protocols.

The primary goal of this thesis is to investigate the feasibility of automating steps within the anomaly detection phase of the groundwater data validation process. Traditionally and still in the majority of the current protocols this step is performed manually. Because of the recent shift of measuring methods to telemetry the amount of data entries significantly increased. Making it more tedious to validate manually.

To navigate towards this goal, several research questions need to be answered, including:

- What requirements does the BRO have for the validated data?
- What are the differences and contributions of hydrological validation in comparison to automatic validation?
- How do the different validation protocols compare to each other?

The research methodology covers a comparative analysis of hydrological versus automated validation techniques. Furthermore literature studies are also performed to find manners of executing automated and hydrological validation. Interviews and expert consultations are held to illustrate the requirements of the BRO and the validation protocols of the different organisations.

The results show the capability of the automated tool to significantly improve the validation process's efficiency. Incorporating statistical tests, consistency checks, timeseries models and exploring the application of machine learning algorithms. The automated tools not only reduced the likelihood of human error but also accelerate the anomaly detection process, allowing for swift decision-making. It is important to note that automated tools are not perfect, and very dependent on among other things the quality of training data.

The main conclusions drawn in this research marks the potential of automation in groundwater data validation. Although automation could improve data validation processes, human oversight remains important. The complex nature of some anomalies within groundwater data needs expert judgment, underscoring the importance of a hybrid approach that leverages the strengths of both automated tools and human expertise.

Table of Contents

1. Introduction

1.1. Context

The Netherlands is a country where 19% of the territory is covered by water, and 26% of its landmass is below sea level [1]. It is very important to manage the water in the land. This is done by different levels of water governing bodies; namely Rijkswaterstaat, the provinces, the waterboards and the municipalities [2]. The different water governing entities have different tasks. The municipalities are responsible for managing urban groundwater [3]. Groundwater refers to water that gradually infiltrates into the subsoil formations through waterbodies such as ditches, lakes, and rivers. Filling the cracks and spaces within soil, sand, and rock where it is stored [4]. Moreover, groundwater plays a crucial role in sustaining healthy ecosystems, and a decrease in groundwater levels can lead to soil subsidence [5]. In recent years there has been a significant shift in the approach to measuring groundwater compared to the past [6]. Traditionally, groundwater level measurements were conducted using monitoring wells, which are tubes with a transparent filter allowing water to enter. Once the water entered the filter, parameters such as water level and quality could be derived. These measurements were previously performed manually or with dataloggers. Manual measurements required a person to be present for the experiment or to read values from the monitoring well. Datalogger data, although recorded, had to be physically retrieved from the well. The retrieved data from these manual measurement methods were riddled with administrative and communicative errors. Since making mistakes is human, and the possibilities for employing detailed, systematic procedures and precise definitions were limited [7]. Therefore, data validation which is a process of ensuring that data used in various applications and systems is accurate, consistent, and meets specific quality standards [8] is necessary. These manual methods have seen a decline in recent years, and the adoption of groundwater monitoring networks, especially in urban areas, has surged. Groundwater monitoring networks utilize telemetry for data transmission, enabling remote parameter measurement and results sent to a central computer. This transition provides water governing entities with real-time information for decision-making and early issue detection. However, it also leads to a substantial increase in recorded measurements. Which goes hand in hand with a significant increase in workload for the expert that checks the measurements on anomalies.

In addition, due to a recent change in legislation, known as “de wet Basisregistratie Ondergrond (BRO),” this new law requires governing bodies to provide validated data about the Dutch subsoil to the BRO [9]. The validation aims to ensure the realism and trustworthiness of the telemetrically measured data. This submission of groundwater data to the BRO is also time-bound and must be done within 20 days. The fast amount of data that needs to be validated in this brief time frame asks for a more modern way of validating, namely automated validation.

Previously, this was not the case, and water governing bodies used the data to observe groundwater trends, quality, and make decisions within their managed areas. However, accessing this data was challenging for different organizations as it was only partially public and lacked standardization. The new BRO law addresses this issue by establishing a uniform method to describe the Dutch subsoil. This is done with a centralized data collection point ensuring standardization, reliability, and public accessibility. As mentioned, the data must be reliable, requiring validation. BZIM conducts this validation in two steps: an automatic validation step ensuring water levels are within the monitoring well's height and checking for unusual increases or decreases. This is followed by a plausibility check, where an expert assesses the data for realism and consistency. However, this step becomes challenging given the brief time frame and large data sample size. BZIM is thus exploring ways to automatically validate all data reliably.

1.2. Problem Description

The problem description will give some more background information about the research subject. First, the problem statement is given (1.2.1). Secondly, the involved parties of this research project are discussed.

1.2.1. Problem Statement

The surge in implementing groundwater monitoring networks, particularly in urban areas, poses a challenge in ensuring the reliability of the collected data. This is especially the case because of the shift to telemetric systems, the establishment of new networks and, consequently, the rise in the recorded measurements. Additionally, these recorded measurements need to be critically examined and validated to comply with the BRO standards. The examination is conducted in two steps: firstly, an automatic check serves as the preliminary phase, followed by a manual check as the subsequent step, which can prove to be challenging. This challenge is primarily due to the vast amount of data and the limited time available for data validation, thereby necessitating a scalable solution.

1.2.2. Involved Parties

The assignment that I will be doing is commissioned by BZIM. This project is relevant, because there has been a change implemented in legislation. This new legislation is called the BRO (Basisregistatie Ondergrond). The BRO is a law that commissions governing bodies to deliver reliable data on the subsoil. The BRO also hosts the database where organisations deliver their data, and since it orders all governing bodies to deliver data, it involves a lot of parties. In this research, the focus will be on the governing bodies which have a stake in the groundwater. So this new law will impact the provinces, municipalities and external companies that manage groundwater for the governing bodies.

1.3. Research Aim and Questions

The objective of this research is to investigate which validation steps can be done automatically and be part of the daily automatic validation process. This validation process checks groundwater levels on anomalies.

The main research question is formed by the research aim:

- **What validation steps can be automated in the new validation process of groundwater level measurements?**

To answer the main question, the question needs to be dissected into the following sub-questions:

- What requirements does the BRO have for the validated data?
- What are the differences and contributions of hydrological validation in comparison to automatic validation?
- How do the different validation protocols compare to each other?

1.4. Reading Guide

The remainder of this report consists as follows. First, in chapter 2 the theoretical background will be given. Which exists out of information of the BRO and the different kind of anomalies. In chapter 3 the methodology is explained. Here the method of answering the sub-questions is discussed. The requirements of the BRO are discussed in Chapter 4. In Chapter 5 different automatic validation methods and how this compares to hydrological validation method are explored. Different validation methods from different water entities will be discussed and how these methods compare to each other in chapter 6. Afterwards, a discussion is given in chapter 7, the conclusion in chapter 8 and at last the recommendations are given in chapter 9. Afterwards, the references and appendices are given.

2. Theoretical Background

This chapter is divided into two subjects. First, the BRO is explained (Section 2.1) and what kind of measurements it wants (Section 2.2). Afterwards, the several types of anomalies are discussed (Section2.3).

2.1. Basis Registratie Ondergrond (BRO)

The idea of including subsoil data in a basic registry has existed since the early 2000s [23]. Throughout this time, TNO was responsible for managing DINO-loket, where borehole-, well-, and groundwater level data were collected. Submissions to DINO was always voluntary, never a legal obligation. From the government a movement to include geo-information, especially subsoil data, within a legal framework started. This eventually led to the idea of also placing subsoil data under a basic registration, thus under a law. This basic registration encompasses an entire system, from people to vehicles, and now also the subsoil data. This law ensures that the administrative bodies (the data holders) take responsibility for the data complying with the BRO pillars. Where the following pillars for subsoil data are used by the BRO [10]:

- High quality (The data entries are reliable)
- Completeness (The data entries are complete)
- Timeliness (The data entries are up to date)

The data holdership now included in the BRO assigns a data holder, the party that monitors subsoil data (for example, a municipality, water board, province, Rijkswaterstaat), and makes them responsible for the data. If the data quality is assessed to be poor by other BRO users, it is up to the data holder to assess and improve it. In addition to the pillars of the BRO, there are also obligations associated with the basic registration, namely [20]:

- **Supply obligation** (for administrative bodies): The data holders must submit subsoil data collected to the BRO.
- **Usage obligation**: Administrative bodies (national government, provinces, municipalities, water boards) must use BRO data if the assignment or project entails the subsoil.
- **Reporting obligation**: An administrative body that has doubts about whether an authentic piece of data in the BRO corresponds to the physical reality, must report this to the TNO. They must also state the reason for this.
- **Research obligation**: According to the Wet BRO, data holders are obliged to investigate feedback on their BRO submissions.

2.2. BRO Measurement Series

The BRO Act framework has a scope to determine which measurement data are mandatory. Within the groundwater level investigation (GLD) domain, the registration object contains the measurements of the variation in the groundwater level that is obtained through a groundwater monitoring well [21]. It is useful to include data from wells where measurements have been taken for a duration longer than a year. This choice was made so that a hydrological cycle of at least one year is incorporated into the data. If the data holder has measurements that only cover, for example, half a year, it is up to the data holder to decide if the measurement series is thought to be relevant enough to register. In the BRO database three distinct types of validated time series are incorporated namely:

- **Provisional series**: This series is expected to be quickly available, but not fully assessed yet. This data is labelled as a provisional series.
- **Definitive series**: This series is established after a thorough assessment has taken place of each data entry and is labelled as a definitive series.
- **Historical series:** Historical measurement series are voluntary, where TNO is legally obligated to migrate the data from the DINO desk to the BRO.

At last, it is possible to override the provisional series with the definitive series when it's available.

2.3. Anomalies

Before discussing the detection of measurement errors in groundwater levels, it is essential to define what constitutes a measurement error. Simply put, a measurement error is the discrepancy between the recorded value and the actual groundwater level. However, a measurement is expected to represent the groundwater level at a specific layer, location, and time. An error occurs if a measurement intended for one layer is interpreted as belonging to another, despite being accurate for its actual layer. Errors can arise from various sources, including manual data entry, equipment inaccuracies, and misinterpretation of data layers [25]. Anomalies in data, which include both random and systematic errors, highlight the importance of careful monitoring and validation to ensure data accuracy.

So for the data to be as reliable as possible, the data needs to be checked for anomalies. In this document seven anomalies will be discussed namely; drift, filter replacement, pattern change, jumps, temporary increases or decreases, and other unexpected measurements. In the category unexpected measurements four additional anomalies are added that are defined within the ArtDiver [26] anomaly checking software. This software is used by multiple water entities for their anomaly checks. These additional anomalies are dry, air, hang depth and frost. The level of importance to filter the anomaly is displayed in Appendix B1. The images and frequencies derive from a study done on the groundwater measurement series of the Dutch provinces of Noord-Brabant, Zuid-Holland, and Groningen [23].

- **Drift (Frequency Low):** Drift in groundwater measurement refers to a slow, consistent upward or downward trend in the data. This might not be apparent within the data series itself but becomes evident when analysing the differences between two measurement points over time. This drift could reflect actual changes in groundwater levels. For example those caused by increasing withdrawals, or it could be due to a gradual shift in the accuracy of the pressure sensors used for measurements.
- **Filter swapping (Frequency Low):** Filter swapping can occur in various forms, such as occasional mix-ups over time, or a consistent swap from a certain point onwards. Sometimes this is clearly visible, but other times it only becomes apparent when analysing the series of differences between datasets.
- **Pattern Change (Frequency: Low):** The pattern of the measurements can change during the monitoring period. For example, there may suddenly be more peaks visible, or the data might show a flattened curve that later becomes more variable.

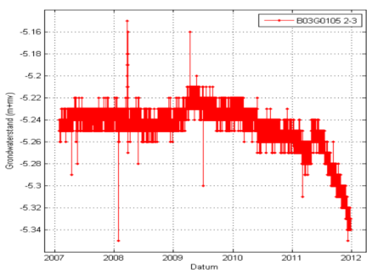

Figure 1: Drift [23]

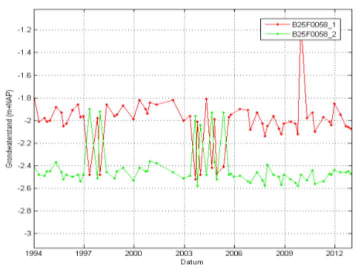

Figure 2: Filter Swapping [23]

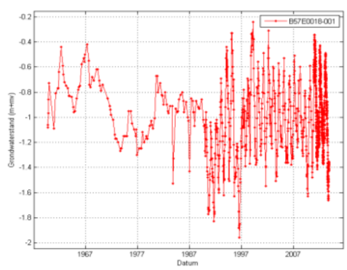

Figure 3: Pattern Change [23]

- **Spikes (Frequency: Moderately High):** Spikes are occurrences where the groundwater level shows a sudden, noticeable increase or decrease.
- **Temporary Decrease (Frequency: Low):** This category describes a scenario where the groundwater level suddenly drops but then, often after months or years, returns to its original level. It's typically an error similar to a spike, corrected eventually but not retroactively.
- **Temporary Increase (Frequency: Low):** Similar to a temporary decrease, this refers to when the groundwater level experiences a temporary rise before eventually returning to the previous level.

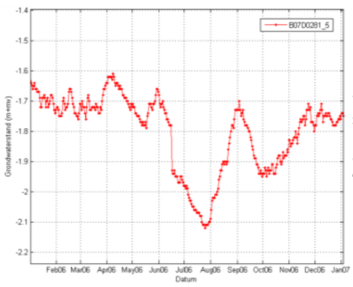

Figure 4: Spikes [23]

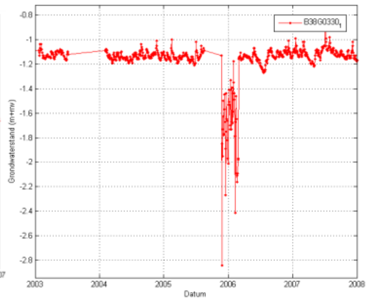

Figure 5: Temporary Decrease [23]

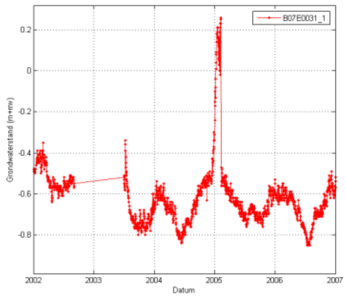

Figure 6: Temporary Increase [23]

- **Unexpected Measurements (Frequency: High):** This category includes all other anomalies that don't fit into the previously defined categories. These could be unexpected peaks, outliers, odd fluctuations, and any other abnormalities that stand out in the data series. Not all strange measurements are necessarily errors, but they are data points that require additional attention.
- **Dry/Air:** The logger only measures air pressure, often seen as a flat period at the bottom of a data series.
- **Hang Depth**: Errors related to the depth at which the logger is suspended, marked by periods where the data is at an incorrect level.
- **Frost:** Water in the well's pipe is frozen, making measurements not related to the actual groundwater level.

The same research [23] conducted on the groundwater measurement series of the Dutch provinces of Noord-Brabant, Zuid-Holland and Groningen concluded the anomalies that occurred the most in the series were strange measurements and spikes. Moreover, the same study also mentioned that all the anomalies had a high level of importance to be monitored in the measurement series.

3. Methodology

The objective of this project is to investigate which validation steps can be done automatically and be part of the daily automatic validation process. The methodology of this research describes the methods used to answer the research questions. A step-by-step method of answering the research question (RSQ) will be described. Where every sub question (SQ) will be addressed using specific research methods, with the associated coloured blocks indicating the chosen method for tackling each question:

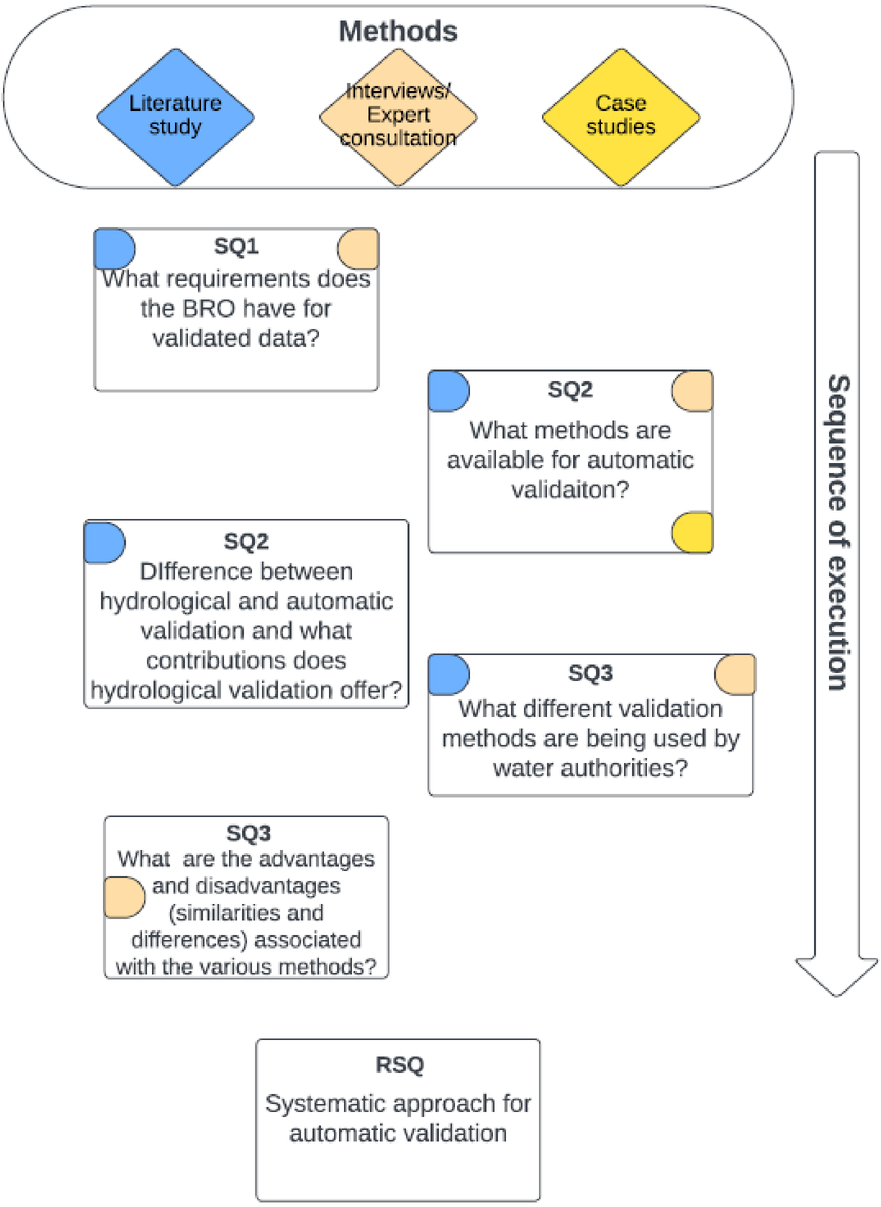

Figure 7: Schematic representation methodology

Research method	Color-codes
Literature study	
Interviews/expert consultations	
Case studies	

Table 1: Colour codes for the research methods

Sub question 1 (SQ1): What requirements does the BRO have for the validated data?

The initial phase involves documentation of BRO requirements, primarily through literature studies. Literature on the BRO requirements can be found on the website of the BRO and GitHub [22]. This encompasses an investigation of the guidelines that are given by the BRO. These guidelines range from type of delivery data to the timespan of delivery. Furthermore additional questions as; what requirements does the BRO have for the quality of the data? Does the BRO prefer a certain way of testing the data? Are answered by looking at the literature provided by the BRO itself and by interviewing Erik Simmelink from the Nederlandse Organisatie voor toegepaste-natuurwetenschappelijk onderzoeks (TNO). Erik Simmelink had a significant part in the creation of the BRO.

Sub question 2 (SQ2): What are the differences and contributions of hydrological validation in comparison to automatic validation?

To answer this question the first step is to identify what methods are available for automatic validation. Literature studies were conducted by searching through the internet for papers that give an overview of the possible different automated methods, and that compare different automated validation techniques with each other. A total of at least 30 papers over the last two decades were examined, with emphasis on automatic anomaly detection, what methods are used in modern water management and how accurate these methods are. Key search terms included "Groundwater anomaly detection", "Smart measuring methods groundwater", "models to forecast groundwater levels", "groundwater hydrographs automated detection of errors"," automated validation groundwater", " comparison variable specific and model-based anomaly detection" and "Machine learning algorithms for modelling groundwater level changes", among others.

Furthermore, for the hydrological validation process an expert consultation session with Johan Bouma and Emiel Groot two hydrologists from BZIM was held. This expert consultation session was executed, because hydrological validation on its own is not a term that is used outside of BZIM. So to get a better understanding of this term, this consultation was held. At last, the two methods got compared with each other by looking for the similarities and differences and how these may contribute or restrict one another.

Sub question 3(SQ3): How do the different validation protocols compare to each other?

To answer this question, it is important to start by identifying the different available methods. An overview of different validation protocols is supplied by the same BRO document as discussed in sub-question 1. Also an expert consultation session was held with Johan Bouma and Emiel Groot to use their knowledge in this field to find additional validation protocols. After identifying the different protocols, interview questions were prepared intending to delve into the techniques employed by each organization in data validation. Five semi-structured interviews were performed, meaning the interviews had a specific focus while also allowing for flexible discussions. These experts were chosen based on their knowledge and experience in areas directly related to this research. The interviewees were hydrologists from various organizations: Jeroen Castelijns (Brabant Water), Pim van Santen (Aa en Maas), Clen Poulie (Vitens), Daniel Heimans (Blik), and Frank van Vliet (Artesia). For these interviews, experts were provided with a list of the interview questions in advance, allowing them to prepare their responses. After interviewing the water authorities, the different answers from the interviews were mapped out and compared with each other to determine the advantages and disadvantages of these methods. And the overall similarities and differences were within the different organizations and their way of validating.

4. BRO Requirements

The requirements of the BRO can be split into two sections. What requirements does the BRO have for the validation process therefore for the data quality. And timewise how quickly does the BRO want the data to be delivered?

4.1. Data Quality

To ensure the reliability of data quality the measurement series must be fully evaluated. The content of the submitted measurement series is considered complete when each groundwater level entry has a quality control status given by a validator from the data holdership. The status of the quality consists of the following categories [22]:

- **Rejected**: This means there is a reason to consider this data as incorrect based on the applied assessment procedure, and the actual value cannot be determined.
- **Approved**: This means there is no reason to doubt the accuracy of this data based on the applied assessment procedure.
- **Not Yet Assessed**: This means there has been no assessment of the quality performed yet, and the assessment will take place later.
- **Undecided**: This means there are doubts about the accuracy of this data, but a definitive conclusion could not be reached based on the applied assessment procedure.
- **Unknown**: This applies only to historical data series. These are data series that have been measured and stored before the introduction of the BRO. These data series need to be reassessed and the title unknown implies there has been no assessment of the quality, or it is unknown whether an assessment has been conducted.

The content of a measurement submission is considered complete (definitive series) only when the full assessment has been performed. The BRO remains neutral on how this validation should be carried out. This means that ensuring the same level of data quality may be challenging because different parties use different validation methods. However, the BRO does not recommend any specific protocol. Nevertheless, the BRO recognizes the value of a uniform protocol to ensure consistent data quality. Although, it does not have the role of enforcing this, as it is a neutral and agnostic system designed to register acquired data for users [23]. Furthermore, the BRO does not concern itself with the measurement frequency of the series. It is up to the data holder to determine what is sufficient for their area [23].

At last as mentioned in the unknown category there are also measurement series that are recorded previous to the creation of the BRO. These are called the historical series, and these series are also kept on the BRO database. When these measurement series are fully assessed they fall into the following quality regimes: IMBRO and IMBRO-A. IMBRO-A allows for data submission without detailed metadata, and this is only allowed for historical data. If the metadata is available the IMBRO title is connected to the historical measurement series. Furthermore for new data entries, it is mandatory to provide information about the measuring instrument (IMBRO) [22].

4.2. Data Delivery

After fully assessing the measurement series a legal term is set for data delivery. This legal term to send the data to the BRO is 20 days. In practice this is more complex, since the 20-day term only starts when both the data acquisition process and the assessment has been performed. The 20 days serve as a legal incentive to encourage a certain rhythm of fast deliveries to promote the speed of assessment for the benefit of society [23].

5. Data Validation

The process to control data is normally referred to as data validation but can also be referred to as quality control. This is inspired by the protocol named "protocol voor datakwaliteitscontrole (QC)" [7] the validation protocol used by the Dutch provinces. Therefore the blocks displayed below follow a similar structure as the QC protocol. Hereby the blue coloured blocks indicate which step of the process is being discussed. In this chapter however only the monitor and plausibility phases will be discussed, because these parts can benefit the most from automatising. Moreover the measuring and storage phases are shortly brought up in the discussions. Furthermore the measuring phase is explained in Appendix A1.

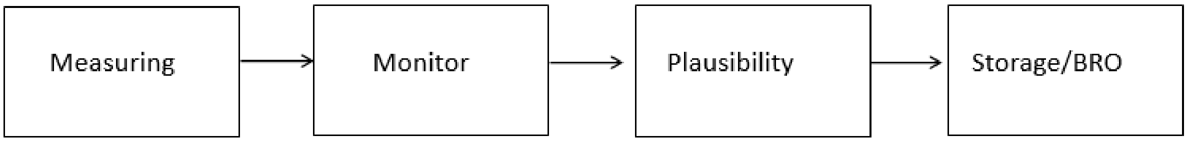

Figure 8: Quality Control Phases

5.1. Monitoring

The first step in this chapter is monitoring the data. This entails checking the collected data for anomalies. Therefore automatic data quality checks will be discussed and at last manual control checks will be specified.

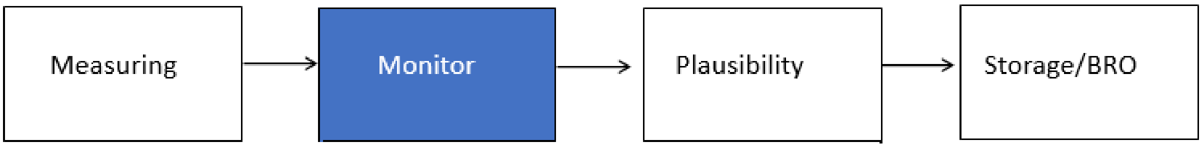

Figure 9: Monitoring phase.

5.1.1. Automatic Validation

Automated data validation processes refer to the use of automated techniques, tools, or systems to evaluate and ensure the quality and integrity of data in an efficient manner. These processes include the development and implementation of algorithms, rules, or procedures that automatically check and validate various aspects of the data, such as accuracy and compliance to defined standards [16]. Additionally given the increase in measurement frequency, manual data quality control can be highly laborious and inefficient [12]. Furthermore, manual checks are subjective, harder to standardize, and can introduce errors themselves. Therefore, developing a validation and quality assurance system that leverages automated methods and tests is recommended [11]. This document covers a wide range of possible tests. Initially, it discusses basic tests. Following that, it explores statistical tests that can be performed. Lastly, it delves into time series models and machine learning models.

5.1.1.1. Basic Tests

Before starting with the basic tests, it is important to get familiar with the terms consistency and plausibility. Where consistency means checking if it is possible given the data about the measurement setup, and plausibility is defined as testing for likelihood or probability [7]. There are a series of tests for assessing data consistency and plausibility in groundwater measurements. These six basic tests are explored [27]:

1. The top of the filter must be higher than its bottom (consistency).
2. The top of the filter cannot be higher than the top of the monitoring well (consistency).
3. Different filters at the same monitoring point must have identical coordinates (consistency).
4. Checks if measurements are dated in the future (administrative test).

5. Verifies if measurements are above the top of the monitoring well (plausibility).
6. Ensures measurements do not fall below the bottom of the filter (plausibility).
The first six relatively simple tests were applied to a real dataset, including the KRW networks of three provinces together with 1466 measurement series across the Netherlands [24]. The results of the tests are displayed in Appendix B.2 and conclude that these simple tests are effective in checking the datasets.

5.1.1.2. Statistical tests

Another method of filtering anomalies automatically is by using statistical tests. In this segment the following methods are going to be discussed namely a value range, flow curve and successive measurements. At last, a case study will reveal the trustworthiness of these tests.

Value Range

The time series of variables fluctuates around an average with a specific value range. For example, shallow groundwater levels typically range from 1.5 to 2 meters below ground level [24]. Observations outside this range are considered anomalies. An automated method can calculate the average and standard deviation over a selected measurement period. If an observation deviates from the average by more than a set value (e.g., +/- three times the standard deviation), it's flagged as suspicious. In [27], an observation is flagged as extreme or suspicious if it falls outside a predefined value range based on all measurements in the series. A common approach is to use plus or minus three times the standard deviation, which corresponds to an interval encompassing 99% of the values in a normal distribution, as displayed in figure 10.

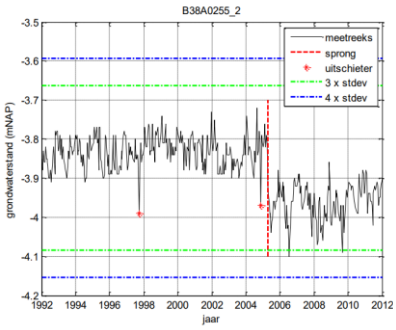

Figure 10: Value Range [24]

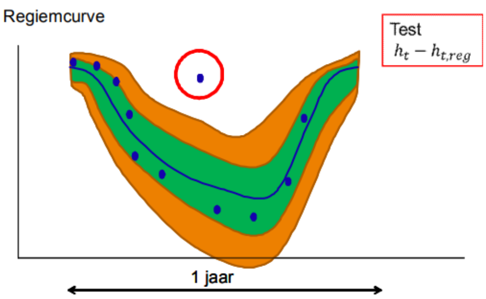

Figure 11: Flow Curve [27]

Flow Curve

Many groundwater time series exhibit seasonal patterns. A flow curve is made with the averages of observations on specific dates or time intervals across different years, which characterizes this variation [27]. It offers better distinction than the value range but requires a longer observation period (e.g., at least 5 years) as displayed in figure 11.

Successive Measurements

To account for correlations between successive measurements, differences between equally distant observations can be calculated [27]. The average and standard deviation of these differences is used to determine if a new observation falls within an acceptable range (e.g., +/- three times the standard deviation) as displayed in figure 12.

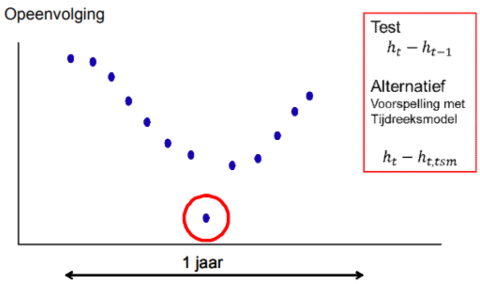

Figure 12: Successive Measurements [27]

Case study

In [24] a statistical test was made that uses the value range and successive measurement principles. For this test, the groundwater levels from the Kaderrichtlijn Water (KRW) monitoring networks in the provinces of Groningen, Noord-Brabant, and Zuid-Holland were visually checked for errors and anomalies in the time series. During the visual check, one or more anomalies were found in 163 groundwater level series. Thereafter the value range was determined meaning the upper and lower limits were established for all groundwater level series using the mean and standard deviation. The bounds of the upper and lower limits were varied by changing the standard deviation by 2, 3, 4, and 5 times the standard deviation. Subsequently this value range test was used to check the number of anomalies it could detect correctly, having the visually checked number of anomalies as a reference point. However, there are two risks associated with this approach:

1. False Positives: The test may indicate the presence of deviations in a series when there are none.
2. False Negatives: The test may indicate that there are no deviations in a series even though deviations were found during visual inspection.

From this test, the results are displayed in Appendix B3. The test concluded that the statistical tests appeared to be unsuitable for automatically detecting errors. With a bandwidth of 3 times the standard deviation, 27 out of 163 series are incorrectly approved, which accounts for 17% of the series with deviations. At the same time, 44% (72 out of 163 series) are still falsely indicated as containing deviations. Accounting for +/- 61% of the measurements being either false positives or false negatives. When comparing the results with the other standard deviations the percentage of false positives/negatives increases. This increase stretches from 59% with one standard deviation to 75% with five standard deviations. Noteworthy is that the number of false positives decreased per standard deviation step and contrary the number of false negatives increased. Concluding that the value range and successive measurement method are not trustworthy in this case study. However, this method does perform well when filtering for extreme spikes.

5.1.1.3. Time Series Models

The next method that can be used for filtering data automatically is time series models. This validation process aims to ensure that hydrological models or data accurately represent the real-world hydrological processes [15]. Here three distinct types of time series models are discussed namely.

- Time series models without external input
- Time series models with external input
- Time series models with multiple measurement series

But first it is important to know the different relevant objectives that can be used for time series models [17,18] mention:

- Prediction: Predicting a value at time step n + 1 (or beyond) based on a time series with n data points.
- Anomaly Detection: Identifying all intervals within a specific time series X that can be classified as "anomalous" based on a similar time series Y, where a different variable is measured for the same time steps.
- Clustering: Grouping individual time series in a database based on their degree of similarity to each other.

To create a time series model it is important to make use of different rules and preferably make use of external influences. The time series models that are going to be discussed in this document are created by Alterra [26]. This research utilizes and displays rules such as Spike and Dry detection in Appendix B4. For additional rules that are commonly used, please refer to [26]. Five time series model were made and compared with each other. These five models make use of different rules/checks, where the measurement series are checked per defined rule. The results from the different rules/checks are merged at the end as displayed in figure 13.

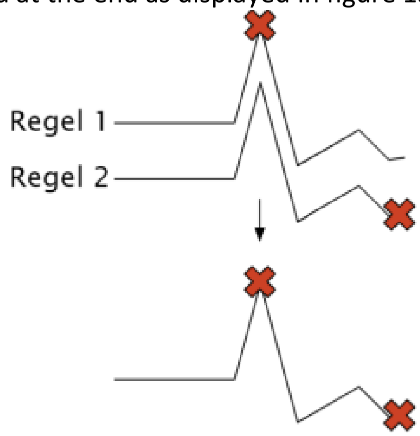

Figure 13: Merging different rules performed by the algorithms [26].

The five different time series model that were made are [26]:

- Time series models without external input

1. **Basic Method with Dry Detection**: Utilizes rules for spike detection, dry periods, and measurements above well tops (Greater/Smaller Than Threshold Rule).
2. **Basic Method with Liveliness:** Designed for cases without logger depths or raw pressure data, this method assesses signal liveliness, first using spike detection and then flagging flat signals (liveliness).

- Time series models with external input

3. **TRA Meteo Method:** Employs a time series model calibrated daily with validated data and weather variables (precipitation and evaporation) to predict measurement ranges.
4. **TRA Meteo+ Method:** Combines dry fall detection from the Basic Method with the TRA Meteo approach to improve dry measurement flagging.

- Time series models with multiple measurement series

5. **TRA Obswell+ Method**: Integrates dry fall detection with a time series model based on nearby measurement points, allowing for more accurate anomaly detection.

In figures 14 and 15 an overview of how well the various time series methods can replicate manual validation is given, considering manual validation was used as the truth. Figure 14 shows how many of the suspicious measurements are correctly detected by the automatic methods. Figure 15 shows how many measurements are incorrectly flagged as suspicious.

				Basic	TRA_meteo	TRA_meteo+	TRA_obswell+
Dataset	**Number of measurements (millions)**	**Number of signals (millions)**	**Percentage of signals (%)**	**Correctly signalled (%)**	**Correctly signalled (%)**	**Correctly signalled (%)**	**Correctly signalled (%)**
Aa en Maas (divers)	25.8	0.85	3.28	52.5	29.0	62.2	**67.9**
Aa en Maas (telemetrie)	7.6	0.02	0.23	1.9	**28.2**	-	-
Brabant Water	98.2	0.61	0.62	69.6	80.6	**83.4**	-
De Dommel	6.4	0.98	15.27	5.2	33.2	**36.0**	-
Rijn en Ijsel	7.2	0.07	0.98	20.7	**37.4**	-	-
Vitens	149.7	6.40	4.28	56.2	47.8	**67.7**	-

Figure 14: Percentage of correctly spotted anomalies by the different algorithms on the different datasets [26].

				Basic	TRA_meteo	TRA_meteo+	TRA_obswell+
Dataset	**Number of measurements (millions)**	**Number of signals (millions)**	**Percentage of signals (%)**	**Too many removed (%)**	**Too many removed (%)**	**Too many removed (%)**	**Too many removed (%)**
Aa en Maas (divers)	25.8	0.85	3.28	1.44	**1.37**	2.28	4.21
Aa en Maas (telemetrie)	7.6	0.02	0.23	**1.83**	2.34	-	-
Brabant Water	98.2	0.61	0.62	**0.38**	2.3	2.55	-
De Dommel	6.4	0.98	15.27	**1.35**	3.13	4.41	-
Rijn en Ijsel	7.2	0.07	0.98	3.34	**1.46**	-	-
Vitens	149.7	6.40	4.28	**2.13**	**2.13**	2.47	-

Figure 15: Percentage of incorrectly spotted anomalies by the different algorithms on the different datasets [26].

Given the variations in datasets and the quality of manual validation, comparing outcomes across datasets is not meaningful. However, comparisons within a dataset show that the TRA methods are better at correctly identifying errors than the BASIC method, although with an increase in false positives. Incorporating a model that accounts for weather conditions or nearby measurement points provides added value over relatively simple statistical rules but, also results in more measurements being flagged as suspicious. The error detection based on time series methods of nearby measurement points (TRA_obswell+) was only applied to one dataset (Aa en Maas (divers)). However it scored the highest in correctly signalling erroneous measurements, although it also had the highest number of false positives. (TRA_obswell+), showing it is quite a promising method.

5.1.1.4. Machine Learning Models

Methods that have been used in recent anomaly detection within groundwater measurement are machine learning models these are additional to hydrological models. Here traditional hydrological models have been employed to predict groundwater levels and detect anomalies, based on physical principles and historical data. These physical processes-based models often, predict groundwater levels by considering factors like precipitation, evaporation, soil moisture, and human extraction [44].These models provide valuable insights but may not always capture the intricate patterns of anomalies due to their reliance on predefined equations that might not account for all variability in groundwater data. However, the complexity of groundwater dynamics, influenced by various environmental factors, demands more sophisticated approaches like Machine Learning (ML) to enhance detection accuracy [43]. Machine learning (ML) is a form of artificial intelligence (AI) focused on building systems that can learn from processed data or use data to improve performance [28].

The Machine Learning process consists out of the following three steps [28]:

1. **Training** the algorithm on a dataset to calibrate its parameters, aiming to align the model's output closely with reality.
2. **Testing phase** where the model's performance is assessed on a separate dataset to ensure its predicting precision.
3. **Modelling** involves leveraging machine learning algorithms to identify patterns and correlations within the data. Finalizing in a model that can predict or classify new instances based on its training.

Machine Learning models detect anomalies by comparing new data points against the learned patterns. When a data point significantly deviates from the model's expectations, it is flagged as an anomaly. This process involves setting thresholds based on the model's confidence intervals or statistical metrics derived from the training data. Machine Learning models require substantial data for training and validation. This data includes historical groundwater levels, atmospheric pressure, precipitation rates, and any other relevant environmental factors that influence groundwater dynamics.

But before the machine learning process starts it is important to choose how the machine learning algorithm is trained broadly categorized into [40]:

- **Supervised learning** where the model learns from labelled data.
- **Unsupervised learning** involves discovering patterns in unlabelled data.

Both are applicable in automatic monitoring for anomalies in groundwater measurements. The selection of an appropriate algorithm depends on the nature of the problem and the data, with a wide range of options. The algorithms discussed in this research will be k-Nearest Neighbors (kNN), k-means, Support Vector Machines (SVM) and the Isolation Forests (iF). The way these algorithms function is explained in Appendix B5. These algorithms can be further divided into two types of models where kNN, SVM and k-means are all geometric models. This entails that data is represented as points in a multidimensional space, where the number of attributes of the dataset also represents the number of dimensions of this space [42]. In its simplest form it can be a two-dimensional graph for example, the time plotted against the water level. The Isolation forests (iF) algorithm however, is an ensemble model which entails that the algorithm combines multiple machine learning algorithms to collectively arrive at a prediction [29]. The following paper [19] observes that the best results are generally achieved with model ensembles.

The above-named algorithms were compared to one another in two different studies [29,30], where the performance of the model is described by an AUC score. AUC ranges in value from 0 to 1. A model whose predictions are 100% wrong has an AUC of 0.0; one whose predictions are 100% correct has an AUC of 1.0 [41].

Table 2: Comparing the different machine learning algorithms on AUC score.

Algorithm	**Type**	**Learning Method**	**Results [29]**	**Results [30]**
KNN	Geometric	Supervised	0.9662	0.9324
k-Means	Geometric	Unsupervised	**0.9719**	-
(One Class) SVM	Geometric	Unsupervised	0.9691	**0.9566**
iForest	Ensemble	Unsupervised	0.8764	0.9419

The study [29] on anomaly detection of groundwater micro dynamics in Chengdu, China, utilized Machine Learning algorithms such as the kNN method, the SVM method, the Isolated Forest and One-Class SVM demonstrating their efficacy in identifying outliers in groundwater levels with high precision and recall rates here the k-Means method outperforms the others slightly. The study [29] used data from monitoring wells, including rainfall and barometric pressure, to train ML models for detecting anomalies in groundwater levels. Furthermore, in the second study [30] the results show that, one class SVM achieves the best detection performance while iForest and KNN are good candidates for anomaly detection.

As the results show the level of correct predictions can be high ranging from 87% – 97% of the measurement series. This high percentage is also dependent on the quality of data that is presented to the method when training them. Moreover, the process also involves careful consideration of potential pitfalls. One of those being sampling bias, where the data used may not be representative of the broader context [28]. And overfitting, where the model becomes too tailored to the training data and loses its ability to generalize to new data [28]. Additionally, it's crucial to distinguish between correlation and causality to avoid erroneous conclusions about the relationships between variables.

5.2. Manual measurement

In modern times where groundwater measurements are increasingly dominated by telemetric sensor technology, manual measurements stand as a critical counterbalance, providing a means to validate and read out data loggers. When it comes to monitoring groundwater levels, the practice of manual control measurements has the following benefits:

- Improving the measurement accuracy and reliability of groundwater level series [31].
- Detecting any measurement errors and correcting them [32].

Frequency: For the frequency of manual checks there is no specified or mandatory number but if automatic sensors are used, it is advisable to manually check the groundwater monitoring wells about four times a year, spread over the seasons [31]. These manual checks verify the accuracy of the recorded data from the past three months and ensure the validity of the upcoming three-month period, which is crucial for data validation.

Reliability: Errors can also occur in manual measurements, which is why multiple manual readings are required to correctly validate the quality of the data from pressure sensors. From an analysis performed at Vitens [33] +/- 15% of the manual measurements taken as manual control measurements had to be corrected or deleted, the most common types of correcting interventions were:

- Meter Error: Manual measurement differs exactly by 1 or more meters from nearby measurements.
- Zero Value Correction: Manual measurement is exactly 0 cm relative to the Reference Ground Level and clearly deviates from nearby measurements.

5.3. Plausibility

The following step after monitoring for anomalies is to determine what to do when anomalies are present in the measurement series. Anomalies, whether found through automated means or visual inspection, must always be investigated. This is to determine whether it is an actual incorrect measurement, or if it was a unique groundwater level that was accurately measured. So the anomaly has to be checked if it is possible and if so, what the explanation is behind the anomaly. If explainable it is time to correct the anomaly, if that is not the case, then it is maybe better to remove the data point from the series. The plausibility check consists of a visual check and hydrological reasoning.

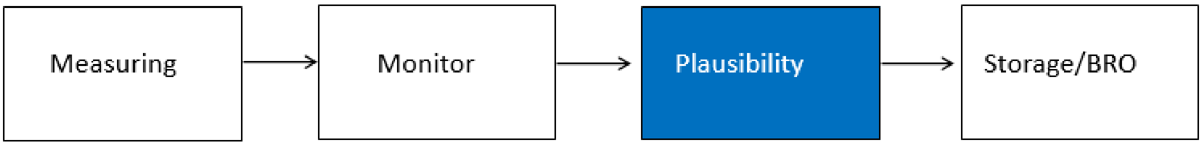

Figure 16: Plausibility check phase.

The plausibility check process follows the structure depicted in figure 17:

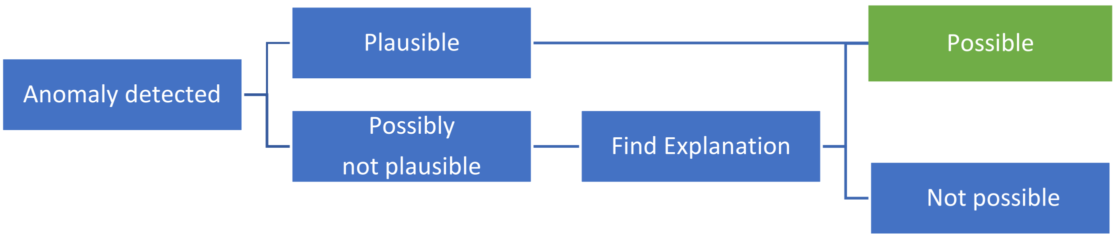

Figure 17: Schematic representation of a plausibility check [34].

5.3.1. Visual Checks

Visual inspection of the data series is a valuable tool to discover any unusual situations in a sequence (jumps, drift, outliers) [24]. Since manual efforts by experts are still crucial and necessary because data quality checks are integral. The study [13] reveals that complete automation has not yet been achieved, which necessitates human intervention.

Graphical representation of series is a suitable method for this visual inspection. Including additional information to a groundwater level graph can improve the visual inspection. Extra information can include filter depth, ground level, top of the monitoring well, as well as statistical information like the expected range of groundwater levels [24].

Finally, each series is visually inspected. What the human eye sees is difficult to program. It is advised to conduct this evaluation at least annually and keep a detailed logbook of the validation and applied corrections so that changes are always reproducible and traceable [31].

The visual check can be performed as follows:

When an anomaly is found in the measurement series, the first step is to consider the weather as an input. Therefore the measurement series is compared with daily precipitation and evaporation figures to see if the response is logical. Other points to consider are [34]:

1. Soil description: Is the response logical in relation to the soil structure.
2. Filter position of the monitoring well: Is the response logical in relation to the filter position.
3. Surface water levels / polder levels: Are the groundwater levels logical compared to the surface water level.

Another step [34] can be to compare the measurement series with an occurring anomaly to other nearby groundwater monitoring wells. Since under similar conditions, it is expected that nearby monitoring wells will react similarly. Another step can be comparing the measurement series to historical measurement data that has been validated previously. Here the data is checked whether the groundwater regime of the recent period differs from the previous measurement period.

The result from these visual checks is classified into 2 types of measurement data [34]:

1. Measurement data that is plausible: These are logical measurement values, and there is no reason to reject the measurement values. This data is thus qualified as reliable.
2. Measurement data that may not be plausible: These measurement series exhibit abnormal behaviour, and it is the question if an explanation for the observed abnormal groundwater regime can be determined.

5.3.2. Hydrological Reasoning

After the anomalies are discovered in the visual check, it is time to give an explanation of why this event occurred. If an explanation is not found, then the measurement is rejected. A few examples of anomalies and their possible explanations are given below:

Pattern Change [24]

- Relocating the tube to a different location or depth.
- Raising the top of the monitoring well.
- Environmental interventions, such as installing drainage or starting extraction.
- Adjusting the measurement frequency. This does not change the actual groundwater level, but the measurement series does change.

Jumps [24]

- A jump can be caused by a hydrological intervention in the environment, such as an increase in extraction or setting up a surface water level. Usually, these types of interventions have a more gradual effect.
- A sudden jump is often caused by adjusting the level filter. If a level filter is extended, then the distance from the top of the monitoring well to the water surface increases. If the new height of the top of the monitoring well is not recorded in the metadata, it appears as if the groundwater level has dropped. The opposite happens if the level filter is shortened.
- A change in the hanging depth of a data logger that is not properly implemented also causes a jump.

- A sign change can cause a jump in the series. For example, if a reading in meters + top of the tube is entered instead of meters above the top of the tube.
- In a well with salty groundwater, fresh (rain)water can infiltrate. This causes jumps in the measured pressure due to the difference in density.
- Land subsidence causes the monitoring well to drop.

Temporary lowering: for example, due to temporary dewatering [24].

5.4. Comparison of Automatic Validation and Plausibility Check

Plausibility checks and automatic validation serve distinct but complementary roles in the process of validating groundwater measurement data. In this segment the way these two validation systems differ from one another and contribute to the overall validation protocol will be discussed.

5.4.1. Differences

Automatic validation involves the use of algorithms, mathematical and statistical models. These models include time series models and machine learning, to process large datasets efficiently on anomalies. These anomaly detection techniques primarily focus on identifying data that deviates from expected patterns, ranges, or statistical norms. This is done without necessarily understanding the cause of these deviations, because these anomaly detections techniques are based on predefined criteria and statistical boundaries. While plausibility checks include manual analysis. These analyses involve visual inspection of data trends and comparing them with meteorological events and nearby monitoring wells. This process helps evaluate the plausibility of data points. Therefore it goes beyond only identifying anomalies. It relies on expert judgment and understanding of the hydrological processes to assess whether identified anomalies are errors or accurately represent real hydrological events. It considers the broader context for example equipment failure and environmental factors such as rainfall, evaporation and human activities affecting groundwater levels.

5.4.2. Contributions

Automatic validation and hydrological validation greatly support one another, adding these two methods together gives a more accurate validation process of groundwater data. In the first layer of detection automatic validation methods are used. Which offers a fast and efficient way to process large volumes of data, making it scalable and effective for initial screenings. Identifying anomalies and patterns of errors and flagging these potential anomalies. Here it is important to note that automatic validation methods are not 100% accurate, so using automatic validation as the only filter is not advised since it will get rid of true extreme values. So the best way to incorporate it is for flagging and labelling potential anomalies. Afterwards an expert checks the data and provides the necessary contextual interpretation, ensuring that the validated data is both accurate and reflective of real hydrological processes and deciding on the appropriate corrective actions.

6. Assessment of validation methods

6.1. Overview of several quality control protocols

The information displayed in table 3 comes from the conducted Interviews [35,36,37,38,39] with the different water authorities. The interview questions can be found in Appendix B2 and the summaries of the interviews in Appendix C.

Table 3: Overview of quality control protocols

	Brabant Water [35]	AA & Maas [36]	Vitens [37]	Blik [38]	Artesia [39]
Estimated Number of Monitoring wells	4000	1000	7000	1700	550
Measuring Technique	Datalogger	Telemetric	Predominantly datalogger	Telemetric	Datalogger
Measuring Frequency	Hourly (Retrieval every 3 months)	Hourly	3 hours	Hourly	Hourly (Retrieval every 3 months)
Air pressure measurements	KNMI	Own Equipment	KNMI	Own Equipment	-
Manual Measurements	✓	✓	✓	✓	-
Visual Inspection	✓	✓	✓	✓	✓
Periodic Visual Checks of measurement series	Quarterly	1/2 times per year	Quarterly	-	Quarterly
Used Software	ArtDiver	Traval & Python	ArtDiver	Own software	ArtDiver
Use Of Timeseries Models	-	✓	-	-	-
Validation Step					
Physically impossible	Manual	Automatic	Manual	Automatic	Manual
Data Completeness	Manual	Automatic	Manual	Automatic	Automatic
Sensor Drift	Manual	Automatic	Manual	Manual (by companies that use Blik's software)	-
Air Pressure Compensations	Automatic	Automatic	Automatic	.	Automatic
Outlier Detection	Manual	Automatic	Manual	Manual (by companies that use Blik's software)	Automatic
Pattern Change	Manual	Automatic	Manual	Manual (by companies that use Blik's software)	Manual
Cross-reference	Manual	Automatic	Manual	Manual (by	Manual

with nearby well				companies that use Blik's software)	
Protocol					
	Brabant water emphasizes the thorough placement and maintenance of measurement points to ensure long-term reliability. They rely heavily on manual validation processes and historical data comparison for anomaly detection.	AA & Maas uses automated Python scripts for daily checks of groundwater data against physical boundaries and predictive models. This telemetric system allows for quick identification of anomalies and timely intervention.	Vitens focus on historical data trends, thorough equipment calibration technical feasibility checks, and comparison with nearby measurements to validate data.	Blik conducts preliminary validation checks on the completeness of the data packet (water pressure, air pressure, and temperature).	Artesia focuses on manual validation through hand measurements, acknowledging the inherent error possibilities in both hand measurements and sensor data. They advocate for the use of hand measurements as a critical validation step.

6.2. Comparison of the different methods

Comparing the five different groundwater measurement and validation protocols from Brabant Water, AA & Maas, Artesia, Vitens, and Blik. Provides insight into the variety of approaches used in monitoring and ensuring the quality of groundwater data. The methods have similarities, advantages, disadvantages and differences among each other.

6.2.1. Similarities

Measurement Frequency: All methods involve telemetry or dataloggers with an average frequency of one data entry every hour. Emphasizing the importance of continuous monitoring in modern groundwater management.

Manual Validation: Despite varying degrees of automation, there's a common reliance on manual checks and visual inspection to confirm the accuracy of sensor data.

Environmental Factors: Each method accounts for environmental influences like air pressure, temperature, and precipitation.

6.2.2. Advantages and Disadvantages

Brabant Water:

- Advantages: Through careful site selection and measurement setup placement Brabant Water ensures high reliability and ensures long-term data consistency.
- Disadvantages: The lack of automation and telemetry calls for a labour-intensive process. With less frequent data this leads to potential delays in identifying and addressing data issues and may not catch all anomalies as promptly as automated systems.

AA & Maas:

- Advantages: Rapid identification of issues through real-time monitoring. Automated scripts enable efficient data validation and quick response to equipment failures.

- Disadvantages: Potential overreliance on automation; complex statistical models may require significant expertise and could potentially miss anomalies not covered by the predictive models or physical boundary checks.

Artesia:

- Advantages: Flexible validation approach combining semi-automatic and manual methods; strong emphasis on manual control measurements for accuracy.
- Disadvantages: Semi-automatic process may not be as quick as fully automated systems; manual aspects can be resource intensive.

Blik:

- Advantages: Protocol allows for automated preliminary checks to ensure data integrity before further analysis.
- Disadvantages: strong reliance on manual processes

Vitens:

- Disadvantages: Strong reliance on manual processes and preliminary checks could also limit their ability to quickly process large volumes of data.

6.2.3. Differences

Air pressure data: The way the different water entities collect their data for air pressure differs. This differs between using the air pressure data provided by the KNMI and using their own equipment to measure the air pressure in situ. The problem with using KNMI data is that the KNMI only measures on a few selected areas. This implies that if the monitoring well is situated rather far from this KNMI point the air pressure data needs to be converted which can go hand in hand with anomalies. On the other hand organisations using their own equipment can lead to anomalies as well through faulty equipment.

Datalogger vs Telemetric systems: Although the different water entities measure the data with approximately the same frequency. Retrieving the data is done in different intervals, where telemetric systems retrieve the data hourly and datalogger systems retrieve the data every 3 months. This means there is a risk of data loss. Only knowing the measurement equipment and therefore the data series is faulty at the moment of retrieval, which can lead to significant loss of data.

Degree of automation and manual processes: The key differences between the methods lie in the balance between automated and manual processes. Here some organizations lean more towards automation for efficiency and others prioritizing manual checks for their depth and reliability. Automated systems offer efficiency and real-time monitoring capabilities, essential for managing large datasets and responding quickly to equipment failures or anomalies. However, manual validation processes provide depth, accuracy, and a nuanced understanding of data anomalies, which are invaluable for long-term groundwater management and analysis.

7. Discussion

The aim of this research was to gain insight into what validation steps can be automated, with the scope focussing on the anomaly detection phase, and the BRO requirements.

The first result was that the BRO does not have any requirements for the data entries. Since one of the pillars of the BRO is to have data entries of high quality in their database. It does not align when no requirements or validation protocols are suggested to get this desired result. Since different parties validate in different ways the quality of data may differ substantially. Leading to a not uniform quality of data entries.

The next result is concerning the automated checks excluding the consistency checks. Since these automated checks are not simple physical boundary checks like the consistency checks but are dependent on the quality of the training data and how the models are fitted by an expert. As these models try to imitate the validated data which is assumed the truth and is put into the model as training material. Here a problem arises namely manual validation is subjective. Because the quality of the plausibility and visual checks are determined by the experience and the local knowledge the expert has. Here two experts can identify different number of anomalies in the same measurement sequence. Which create different fits on the same dataset making it hard to say when an automated check is performed correctly.

It is also important to note that apart from the consistency test and the KRW case study using the value range method, the other automated validation methods were tested on different measurements sequences. Which can vary in quality, in quantity, in different events occurring in the data etc. making comparisons hard between methods.

Furthermore next to the subjectiveness of the plausibility and visual checks there are other points that are just outside the scope. But are part of the validation process and influence the reliability of the data. These points are as follows a proper measurement set up, the quality of the training data and the storage of this data. Afterwards the expertise of making the model is discussed.

Measurement set up

Before the anomaly detection begins the first step is to establish the measurement setup for data retrieval. To have a reliable quality of data, it is important to place the measurement setup correctly. The proper way of setting up will be achieved by considering the most suitable location, what measuring method is the most applicable for your research goal, what equipment you want to use and maintaining the measurement setup [32]. Selecting an optimal location is crucial for groundwater measurement, emphasizing setup depth, and distance from influencing structures to ensure accuracy. The method chosen (monitoring wells, open boreholes, or field estimates) depends on the research goals. Measurement frequency and the choice between manual, semi-automatic, and fully automatic methods are tailored to monitoring objectives, balancing cost, labour, and data precision needs. Filter selection is based on the target aquifer and subsoil characteristics, ensuring the filter's placement and length are optimal for accurate measurements. Regular maintenance, both minor and major, is essential to maintain equipment reliability and data accuracy, supporting effective groundwater management and research. A more detailed description of the measurement setup can be found in Appendix A1.

Quality of data

The quality of the automatic models (the value range, time series models and machine learning models) are dependent on the quality of groundwater level data. Ideally, the measurements should meet a pre-specified quality standard and if they don't meet that quality standard the data series should be revalidated.

Points to look after to ensure the training data is of decent quality are [27]:

- Making sure the measuring equipment is set up correctly and working well.
- Minimising the manual processing steps.
- Revalidating historical measurement series.
- Manual control measurements.

Outside of the quality of data that is used to train the models, the availability of data is also important, since the available data can have the following problems [28]:

1. Sparse Events: Making predictions on rarely occurring phenomena challenging due to insufficient historical data. Effective predictions require datasets with numerous instances of the phenomena.
2. Short Measurement Periods: Long-term accuracy needs extensive historical data to understand and predict system dynamics. Short data periods risk increasing prediction errors over time due to the inability to recognize long-term patterns.
3. Low Sampling Frequency: The frequency of data collection must match the system's dynamics and the scale of relevant phenomena.

Storage of data

Measurement data are crucial for research, making documentation and storage vital [31]. According to [24], it's crucial to digitally archive both manual observations and raw readings from automatic instruments, ensuring that any data corrections can be accurately traced and reproduced. This archival process should clearly document who made specific corrections and when they were made, enhancing transparency and accountability.

Since at the moment consistency across different databases is not the standard. The study performed in [27] highlighted the discrepancy that can occur across two distinct databases housing the same measurement series. By comparing groundwater levels from the same period in both databases, the study uncovered that 79 out of 81 series displayed discrepancies, including both technical data and groundwater measurement series. In the groundwater measurement series differences reached up to 6 cm between the databases and a consistent difference of 1 cm. These findings highlight the importance of maintaining uniformity in data to prevent such inconsistencies, which can significantly impact data reliability and analysis outcomes.

Expertise of making model

Creating an automated model is a useful and powerful method of anomaly detection, however developing an elaborate model takes time. The designer within the company wanting to model most likely has to get familiar with the software program and develop a complicated tool. This takes a substantial amount of time and for a company this is translated into labour costs. The cost-benefit ratio must therefore be explored before the development of a tool. An investigation must be conducted to find if the contribution of the developed tool outweighs the investment costs. A cost-benefit analysis would have been an interesting insight for this research.

8. Conclusions

In conclusion the conducted research tries to answer the main question:

What validation steps can be automated in the new validation process of groundwater level measurements?

This is an important question in current water management because of the new BRO legislation. Here the BRO wants the data to be of high quality, complete and up to date. And because of the shift to telemetry which also increases the number of data entries in the systems it is sought after to go through the validation process automatically. This is discussed in this research by looking at the BRO requirements, methods that can be used to automate anomaly detection, manual anomaly detection and at last looking at current validation protocols of different water entities.

Although the BRO wants the delivered data to be of high quality, complete and up to date. The BRO sets no requirements surrounding how the quality of this data is inspected and does not recommend a particular validation method. Furthermore, the BRO set a limit of 20 days to deliver the data when completely validated. This implies when each data entry has been inspected and labelled with the status quality categories provided by the BRO.

For methods to automatise validation, this research delves into basic tests, statistical tests, time series models and machine learning algorithms. Basic tests ensure the logical coherence of data, such as verifying that measurement timestamps are plausible and that spatial coordinates of measurement points remain consistent. These checks have a high-efficiency rate and are straightforward in automatically filtering out glaring inconsistencies before further analysis. Therefore a viable step in automatically checking the data if physically possible.

Statistical tests play a significant role in automated validation processes. They include evaluating data for statistical anomalies, by use of value ranges determined by a moving average and a standard deviation. Based on the quality of the training data these tests can be an efficient way to filter out spikes or trends that deviate significantly from established patterns.

Time series models and machine learning algorithms represent advanced methods for automating the validation process. These models can predict expected data values based on historical trends and external environmental factors, such as weather conditions. Anomalies are detected when actual measurements significantly deviate from these predictions. Again the quality of the training data here is important. The case studies discussed in this paper show the accuracy of the time series model and the machine learning algorithms in correctly identifying anomalies from the measurement series. This makes them both very viable steps to implement in automating the whole anomaly detection process.

Despite the advancements in automation, this research emphasizes the importance of human oversight. Expert judgment is crucial for interpreting the results of automated tests, especially in complex situations where the context of data anomalies needs to be understood. For instance, an automated system might flag a sudden change in groundwater level as an anomaly, but an expert can determine whether this is due to a natural event, such as heavy rainfall, or an error in data collection.

Furthermore, by comparing the different protocols used currently by the different water authorities it is seen that a large portion of the validation process is still predominantly performed manually. Each approach has its strengths and is tailored to the specific needs, resources, and objectives of the organization. The choice of validation protocol depends on the balance each organization wishes to achieve between efficiency, accuracy, and the specific demands of their groundwater monitoring projects.

9. Recommendations

The recommendations section aims to provide actionable insights based on the findings of this research, focusing on enhancing the efficiency and reliability of automated groundwater data validation processes. These suggestions are directed towards organizations involved in water management, encouraging a balanced approach that incorporates both automation and expert knowledge. By adopting these recommendations, stakeholders can improve data validation methods, contributing to more informed decision-making and effective groundwater management.

- What can be applied in practice from this study:

Implement a Hybrid Validation Framework: Organizations should integrate automated validation tools with visual expert checks to create a robust validation framework. This framework would accelerate the anomaly detection process while retaining the oversight for complex data interpretations.

Standardization of Validation Protocols: There is a pressing need for the standardization of validation protocols across different organizations. Standardized protocols would enhance the consistency and reliability of groundwater data, facilitating more effective data sharing.

Collaborative Efforts and Knowledge Sharing: Encourage collaborative efforts between academic institutions, government bodies, and industry players to share knowledge, data, and best practices related to automated data validation. This could accelerate innovation and improve the effectiveness of water management practices.

- What is open for future research:

Meerwaardig Validation Approaches: Explore the concept of 'meerwaardig' validation, where data is not merely classified as correct or incorrect but is assessed on a spectrum. This approach could offer a more nuanced understanding of data quality and anomaly significance.

Interplay Between Manual and Automated Methods: A deeper examination of the interaction between visual checks and automated validation methods is needed to optimize their integration. This includes identifying specific scenarios where one method is preferable over the other and how they can complement each other even better.

10. References

[1] Statistiek, C. B. voor de. (2016, February 25). Minder landbouw, meer natuur. Centraal Bureau Voor de Statistiek. https://www.cbs.nl/nl-nl/nieuws/2016/08/minder-landbouw-meernatuur#:~:text=Ruim%20twee%20derde%20van%20oppervlakte%20Nederland%20is%20groene%20ruimte&text=Nederland%20bestaat%20daarnaast%20voor%2019

[2] Waterstaat, M. van I. en. (2015, September 16). Waterbeheer in Nederland - Water - Rijksoverheid.nl. Www.rijksoverheid.nl. https://www.rijksoverheid.nl/onderwerpen/water/waterbeheer-in-nederland

[3] Rijkswaterstaat. (n.d.). Grondwatertaak gemeente. Informatiepunt Leefomgeving. Retrieved December 7, 2023, from https://iplo.nl/thema/water/stedelijk-afvalwater/stedelijkwaterbeheer/grondwatertaak-gemeente/

[4] Waterstaat, M. van I. en. (2020, August 11). Grondwater - Ons water. Www.onswater.nl. https://www.onswater.nl/onderwerpen/grondwater

[5] Het grondwatersysteem | RIVM. (n.d.). Www.rivm.nl. Retrieved December 7, 2023, from https://www.rivm.nl/landelijk-meetnetgrondwaterkwaliteit/grondwatersysteem#:~:text=Belang%20van%20grondwater

[6] Grondwater - Geologie van Nederland. (n.d.). Www.geologievannederland.nl. Retrieved December 7, 2023, from https://www.geologievannederland.nl/ondergrond/afzettingen-endelfstoffen/grondwater

[7] von Asmuth, J., & van Geer, F. (2015). Kwaliteitsborging grondwaterstands- en stijghoogtegegevens : Systematiek en methodiek voor datakwaliteitscontrole (QC). In Wageningen University & Research eDepot. KWR. https://edepot.wur.nl/345246

[8] What is Data Validation: Definition | Informatica Netherlands. (n.d.). Www.informatica.com. Retrieved December 7, 2023, from https://www.informatica.com/nl/services-andtraining/glossary-of-terms/data-validation-definition.html

[9] Bouma, J. (2023). Afstudeeropdracht Hydrologische validatie grondwaterstanden in grondwatermeetnetten. Afstudeervoorstel hydrologische validatie grondwaterstanden. https://www.bzim.nl/vacature_pdf/Afstudeervoorstel_hydrologische_validatie_grondwaterstanden.pdf

[10] Wet Bro. (n.d.). Basisregistratieondergrond. Retrieved December 7, 2023, from https://basisregistratieondergrond.nl/inhoud-bro/wetbro/#:~:text=Vanaf%201%20januari%202018%20is

[11] Singh, C. K., & Katpatal, Y. B. (2020). A Review of the Historical Background, Needs, Design approaches and Future Challenges in Groundwater Level Monitoring Networks. Journal of Engineering Science and Technology Review, 13(2), 135–153. https://doi.org/10.25103/jestr.132.18

[12] Vries, D., & Seshan, S. (2021). Data validation implementation pilot. In Wageningen University & Research eDepot. KWR. https://edepot.wur.nl/579226

[13] Panagopoulos, Y., Konstantinidou, A., Lazogiannis, K., Papadopoulos, A., & Dimitriou, E. (2021). A New Automatic Monitoring Network of Surface Waters in Greece: Preliminary

Data Quality Checks and Visualization. Hydrology, 8(1), 33. https://doi.org/10.3390/hydrology8010033

[14] On the Experiences of Adopting Automated Data Validation in an Industrial Machine Learning Project | IEEE Conference Publication | IEEE Xplore. (n.d.). Ieeexplore.ieee.org. Retrieved December 7, 2023, from https://ieeexplore.ieee.org/abstract/document/9402048

[15] Zio, M., Fursova, N., Gelsema, T., Gießing, S., Guarnera, U., Petrauskienė, J., Quensel-Von Kalben, L., Scanu, M., Ten Bosch, K., Van Der Loo, M., & Walsdorfer, K. (2016). Methodology for data validation 1.0 Essnet Validat Foundation. https://croslegacy.ec.europa.eu/system/files/methodology_for_data_validation_v1.0_rev-201606_final.pdf

[16] S. Shrivastava et al., "DQA: Scalable, Automated and Interactive Data Quality Advisor," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 2913-2922, doi: 10.1109/BigData47090.2019.9006187.

[17] Esling, P., & Agon, C. (2012). Time-Series Data Mining. *ACM Computing Surveys*, *45*(1), 1–34. doi:10.1145/2379776.2379788

[18] Maimon, O., & Rokach, L. (2010). Data Mining and Knowledge Discovery Handbook, 2nd ed. *ResearchGate*. https://www.researchgate.net/publication/236005657_Data_Mining_and_Knowledge_Discovery_Handbook_2nd_ed

[19] Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87. https://doi.org/10.1145/2347736.2347755

[20] *4 plichten uit de wet*. (n.d.). Basisregistratieondergrond. https://basisregistratieondergrond.nl/inhoud-bro/wet-bro/4-plichten-wet/

[21] *Grondwaterstandonderzoek (GLD)*. (n.d.). Basisregistratieondergrond. https://basisregistratieondergrond.nl/inhoud-bro/registratieobjecten/grondwatermonitoring/grondwaterstandonderzoek-gld/

[22] Basisregistratie Ondergrond Catalogus Grondwaterstandonderzoek. (2023, December 20). https://broprogramma.github.io/GLD/#grondwaterstandonderzoek

[23] Simmelink, E., personal communication, January 12, 2023

[24] Leunk, I. (2014). Kwaliteitsborging grondwaterstands- en stijghoogte Retrieved from https://edepot.wur.nl/317083

[25] Waaijenberg, J., 2009. Numerieke verdroging. Onderzoek naar numerieke verdroging in het beheergebied van Waterschap Regge en Dinkel. Rapport Waterschap Regge en Dinkel, Almelo.

[26] Brakenhoff , D. (2020). Automatische foutherkenning met Tijdreeksanalyse. Retrieved from https://edepot.wur.nl/524849

[27] von Asmuth, J. R., & van Geer, F. C. (2013). Kwaliteitsborging grondwaterstands- en stijghoogtegegevens Op weg naar een landelijke standaard . Retrieved from https://edepot.wur.nl/345245

[28] Vonk, E., & Vries, D. (2016). BTO Rapport Datamining voor assetmanagement – inventarisatie en voorbeelden uit de watersector. Retrieved from https://edepot.wur.nl/424144

[29] Redyuk, S., Kaoudi, Z., & Markl, V. (2021). Automating data quality validation for dynamic data ingestion. Retrieved from https://sergred.github.io/files/edbt.reds.pdf

[30] Shi, H., Guo, J., Deng, Y., & Qin, Z. (2023). Machine learning-based anomaly detection of groundwater microdynamics: case study of Chengdu, China. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-38447-5

[31] Bouma, J., Maasbommel, M., & Schuurman , I. (2012). Retrieved from https://edepot.wur.nl/242920

[32] Ritzema, H., Heuvelink, G., Heinen, M., Bogaart, P., van der Bolt, F., Hack-ten Broeke, M., Vroon, H. (2012). Retrieved from https://edepot.wur.nl/215081

[33] van Vliet , F. (2014). Retrieved from https://edepot.wur.nl/392105

[34] Bouma, J., Groot, E. (2023). Validatie meetdata meetnetten. Unpublished manuscript.

[35] Castelijns, J. , personal communication, January 9, 2023

[36] van Santen,P., personal communication, January 11, 2023

[37] Poulie, C., personal communication, January 23, 2023

[38] Heimans, D., personal communication, January 11, 2023

[39] van Vliet, F., personal communication, January 10, 2023

[40] Delua, J. (2021). Supervised vs. unsupervised learning: What's the difference? Retrieved from https://www.ibm.com/blog/supervised-vs-unsupervised-learning/

[41] Bhandari, A. (2024). Guide to AUC ROC curve in machine learning : What is specificity? Retrieved from https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/#:~:text=The%20Area%20Under%20the%20Curve,the%20positive%20and%20negative%20classes.

[42] Taylor, M. (2019). Geometric deep learning - convolutional neural networks on graphs and manifolds. Retrieved from https://medium.com/@mtlazul/geometric-deep-learning-convolutional-neural-networks-on-graphs-and-manifolds-c6908d95b975#:~:text=Geometric%20deep%20learning%20is%20a%20new%20field%20of%20machine%20learning,3D%20objects%2C%20graphs%20and%20manifolds.

[43] Shi, H., Guo, J., Deng, Y., & Qin, Z. (2023, September 7). *Machine learning-based anomaly detection of groundwater microdynamics: Case study of chengdu, China*. Nature News. https://www.nature.com/articles/s41598-023-38447-5

[44] Vries, D., & Seshan, S. (1970, January 1). *Data validation implementation pilot*. De bibliotheek van KWR. https://library.kwrwater.nl/publication/65850678/data-validation-implementation-pilot/

Appendices

A. Measurement

A.1. Measurement Setup

Location

As a starting point it is important to consider the location of placement of the measurement setup, here it is important to focus particularly on the arrangement of the measurement site, such as depth, position, and length of the filter in relation to the soil structure and the hydrological system (Van der Bolt et al., 2010). In addition to the vertical positioning of a filter, its spatial positioning is also important, for instance, the distance to a watercourse, road, or house, or the location in a pit/ditch, verge, road, cutting, under trees, or on a local elevation (as reported by Kleijer, communicated by Massop & Van der Gaast, 2003). The local groundwater situation is influenced by these factors. For many measurement objectives, it is important to place the monitoring wells outside the sphere of influence of these local effects to ensure that the ground situation measured in the monitoring well reasonably corresponds with the ground situation in the surrounding area. Therefore, do not place a monitoring well too close to such objects, unless it is necessary for the measurement objective. The final location is determined in the field. At last, it is also important that the monitoring well is placed in a safe and protected spot.

Soil, Hydrological system and Topography

To effectively monitor groundwater, it's essential to integrate knowledge from soil science, hydrology, and topography. Understanding soil structure (layering, permeability, and composition) is crucial for accurately describing measurement data and optimally placing monitoring wells. This includes knowledge of the depth and variations of the water table, as well as the hydrogeological conditions like aquifers and groundwater flow. Additionally, the hydrological system's complexity, influenced by spatial variations in surface elevation, permeability, and environmental factors such as rainfall and atmospheric pressure, dictates groundwater flow patterns and level fluctuations. Topographical considerations are equally important; avoiding areas prone to runoff interference, maintaining safe distances from large trees and structures, and considering the effects of groundwater extraction are essential for precise groundwater level measurements. These considerations, combined with an awareness of the surrounding land usage, are foundational for effective groundwater monitoring, ensuring accurate data collection and the protection of water resources.

Measuring method

When the location of placement is determined, the next step is to evaluate the measuring method. The method that is chosen is mostly dependent on the research objective, these objectives can be regional, urban or project based. In the Netherlands, groundwater levels are estimated using the following methods:

Monitoring wells, which are subdivided into.

Groundwater level pipes

Piezometers

Open boreholes

Field estimates

A groundwater measuring network can exist out of multiple different estimating methods.

Groundwater level pipes: These are shallow monitoring wells that measure a hydraulic head that deviates little from the phreatic groundwater level.

A **piezometer** is a monitoring well which is used to measure the head at the location of the filter. These are monitoring wells that measure the hydraulic head in deeper soil layers.

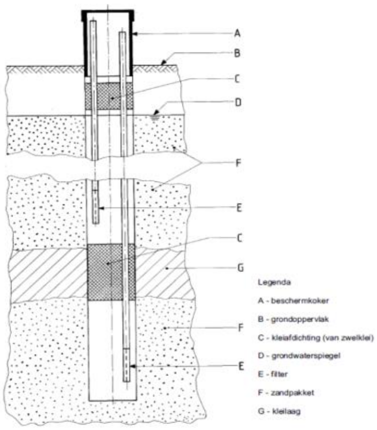

Figure 18: Monitoring Well

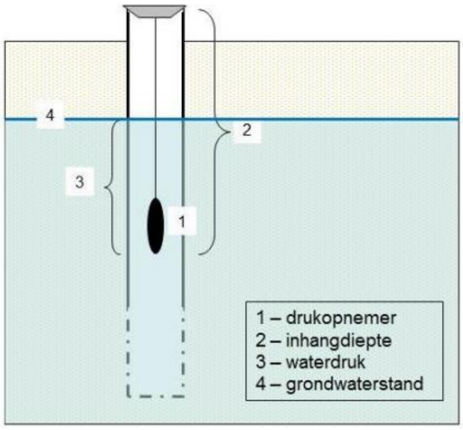

Figure 19: Pressure Sensor

Open borehole: Open boreholes are created by drilling a vertical hole with a diameter of 8-12 cm using a soil auger, down to about 10 cm below the groundwater level (Hooghoudt, 1952). After a settling period (1-2 days), the groundwater level in the boreholes is measured.

Field estimates: Field estimates are primarily a tool to select the location of open boreholes. Field estimates are based on profile and field characteristics and measured groundwater levels in boreholes and monitoring wells. Profile characteristics are caused by the annual fluctuation of the groundwater level.

As seen in the table in below open boreholes and field estimates are less then ideal so that is why in practice, it is not recommended to use these 'non-monitoring wells' for groundwater level measurements, because their filters are usually located in multiple aquifers or the pipes have a diameter that adversely affects accurate measurements.

Criterium	Beoordeling van de meetmethode			
	Grondwater-standbuis	Piëzometer	Open boorgaten	Veldschattingen
Detailniveau	+	++	+	-
Doorrekenen scenario's	n.v.t.	n.v.t.	n.v.t.	n.v.t.
Doorlooptijd	+	+	+/-	-
Nauwkeurigheid	+/-	++	+/- tot +[a]	-
Gebruik aanvullende gegevens	+/-	+/-	+	+
Fluxen te bepalen	-	++	-	-
Extrapoleerbaarheid	n.v.t.	n.v.t.	n.v.t.	n.v.t.
Reproduceerbaarheid	++	++	+/-	-
Objectiviteit	+	+	+	-
Ontwikkel- en gebruiksgemak	+	+	+	+
Kwantificering onzekerheid	+/-	+	+/-	-
Toepassing bij weinig gegevens	n.v.t.	n.v.t.	n.v.t.	n.v.t.
Interpoleren als tussenstap	n.v.t.	n.v.t.	n.v.t.	n.v.t.

Figure 20: Comparison off different measuring methods

Frequency and Measuring Equipment

Choosing the appropriate frequency and method for groundwater level measurements is crucial, this chose is also influenced by the objectives of monitoring. The distinction between manual, semi-automatic, and fully automatic measurements outline the varied approaches to data collection. Manual measurements involve field visiting the site to record groundwater levels, suitable for locations with variable measurement frequency needs. Semi-automatic systems utilize pressure sensors and data loggers to record groundwater levels at set intervals, requiring manual data retrieval periodically, whereas fully automatic systems transmit data directly to a central station for immediate processing, enhancing real-time monitoring efficiency.

The decision on measurement frequency impacts the clarity of data, with higher frequencies better capturing short-term effects like rainfall or evaporation, and lower frequencies suited for observing long-term trends. Equipment selection is driven by the measurement frequency, offering different benefits and costs. Manual measurement is cost-effective but labour-intensive and prone to errors. Semi-automatic measurement offers a balance, reducing labour while providing higher frequency data at the risk of data gaps dependent on how often the data gets retrieved. Fully automatic systems offer real-time error detection and alerts but at higher costs and maintenance needs. Pressure sensors, such as Keller and Divers, are commonly used in semi and fully automatic systems, measuring total pressure (water plus atmospheric) to determine groundwater levels. Keller sensors measure atmospheric pressure directly, facilitating immediate groundwater level calculation, while Divers require atmospheric pressure compensation, either through a separate sensor or external data from often the KNMI (Koninklijk Nederlands Meteorologisch Instituut), for accurate groundwater level determination. This technological framework ensures accurate, efficient, and adaptable groundwater monitoring to meet various research and management objectives.

Filter

When selecting a filter for groundwater measurements, the primary consideration is the target aquifer, guided by the objectives of the monitoring network. The filter depth and length should be based on subsoil data, with a standard length of 1 meter, unless the target layer is thinner. Anisotropy (movement of water) in soil can affect measurements, so filters should not be placed too deep. Different filter setups may be used at one location for comprehensive studies, such as seepage or hydrological modelling.

Maintenance

To ensure that the quality of data that is incoming is as reliable as possible it is important to perform maintenance checks on the equipment. There are minor maintenance routines and major ones and during the maintenance routine it is important to registrate what is done to the monitoring well. During fieldwork, the field technician conducts check on the monitoring well and may notice issues that require maintenance, such as:

- Damage to the monitoring well

- Blockage in the monitoring well

- ….

Minor/Major maintenance

Groundwater monitoring maintenance is split into minor and major tasks. Minor maintenance involves quick fixes and checks when systemic anomalies are detected in data readings, including sensor replacement for deviations of more than 10 centimetres. Major maintenance, scheduled every five years, focuses on extensive cleaning and recalibration of monitoring wells to ensure data accuracy, such as clearing blocked filters and recalibrating against a reference point like NAP. This systematic maintenance ensures reliable and precise groundwater measurements.

B. Monitoring

B.1. Frequency and importance of the anomalies

Nr	Soort fout of bijzonderheid	Frequentie	Mate van belang
1	Drift (trend in verschil tussen filters)	Weinig	Groot. Deze fout blijft makkelijk onopgemerkt en interfereert met werkelijk optredende trends.
2	Filterverwisseling	Weinig	Groot. Dit leidt tot foutieve waarden.
3	Patroonverandering van grondwaterstandsreeksen	Weinig	Klein/groot. Patroonverandering als gevolg van hogere meetfrequentie is geen probleem. Bij andere oorzaken is het belang wel groot, omdat het watersysteem dan gewijzigd is.
4	Sprong	Matig veel	Middelgroot. Deze is relatief eenvoudig visueel op te sporen. Een sprong kan echter zowel een foutief resultaat betreffen als werkelijk optreden in de grondwaterstand. Voor de borging is het hierbij dus van groot belang dat de beoordelaar over voldoende gebiedskennis beschikt.
5	Tijdelijke verlaging	Weinig	Middelgroot. Deze is relatief eenvoudig visueel op te sporen. Ook hiervoor geldt dat het zowel een fout in de meting kan zijn, als werkelijk kan optreden.
6	Tijdelijke verhoging	Weinig	Middelgroot. Deze is relatief eenvoudig visueel op te sporen. Ook hiervoor geldt dat het zowel een fout in de meting kan zijn, als werkelijk kan optreden.
7	Vreemde metingen	Veel	Groot. Vooral omdat niet duidelijk is wat de

Figure 21: Frequency and importance of the anomalies

Code	Groningen (aantal)	Noord-Brabant (aantal)	Zuid-Holland (aantal)
Totaal aantal reeksen	83 reeksen	65 reeksen	133 reeksen
Aantal reeksen met fouten	47 (= 57%)	24 (= 37%)	92 (69%)
Drift	15 (18%)	3 (5%)	0 (0%)
Filterwissel	7 (8%)	6 (9%)	6 (5%)
Patroon	4 (4%)	1 (2%)	18 (14%)
Sprong	16 (19%)	3 (5%)	20 (15%)
Verhoging	8 (10%)	0 (0%)	3 (2%)
Verlaging	5 (6%)	5 (8%)	7 (5%)
Vreemde meting	20 (24%)	12 (18%)	74 (56%)

Figure 22: Overview of the number of anomalies in the data series

B.2. Efficiency of the basic tests

nr	Toets	Effectiviteit	Opmerking
1	Bovenkant filter < onderkant filter	Groot	Vaak ontbrekende gegevens, waardoor toets niet mogelijk is.
2	Bovenkant filter > bovenkant peilbuis	Groot	Vaak ontbrekende gegevens, waardoor toets niet mogelijk is.
3	Filters van één peilbuis met verschillende coördinaten	Groot	Klein verschil (maximaal 5 meter) is toegestaan.
4	Metingen in de toekomst	Groot	Metingen in de toekomst zijn in geen enkele onderzochte reeks aangetroffen.
5	Metingen > bovenkant buis	Groot	Dit zijn niet perse fouten in de meting of in de metadata. Een put kan werkelijk overstroomd zijn (metingen met drukopneming, of handmeting met opzetstuk). Het is wel belangrijk om dit te controleren.
6	Metingen < onderkant filter	Groot	Dit kan een fout in de metingen zijn, maar ook een fout in de technische gegevens van de peilbuis.

Figure 23: Efficiency of the basic tests

B.3. Different standard deviations for the value range

Nr	Bandbreedte	Vals positief	Vals negatief	Type fout dat niet wordt opgespoord
1	Gemiddelde +/- 2 keer standaarddeviatie	97	0	
2	Gemiddelde +/- 3 keer standaarddeviatie	72	27	Drift (6) Filterwissel (3) Patroon (1) Sprong (7) Verhoging (1) Verlaging (4) Vreemd (5)
3	Gemiddelde +/- 4 keer standaarddeviatie	21	83	Drift (8) Filterwissel (13) Patroon (5)

4	Gemiddelde +/- 5 keer standaarddeviatie	19	104	Sprong (18) Verhoging (5) Verlaging (7) Vreemd (27) Drift (10) Filterwissel (15) Patroon (6) Sprong (19) Verhoging (5) Verlaging (8) Vreemd (41)

Figure 24: Results of using different sd on the value range

B.4. Rules used for the algorithms

- Standard Deviation Rule: If a measurement is more than N standard deviations away from the mean, it is considered suspicious.
- Maximum Gradient Rule: If the change between two measurements is greater than a certain threshold, the measurement is considered suspicious.
- Greater/Smaller Than Threshold Rule: If a measurement is greater/Smaller than a certain threshold value, it is considered suspicious.
- Spike Detection Rule: If there is a spike above a certain magnitude that is immediately followed by a jump of comparable size in the opposite direction, the measurement is considered suspicious.
- Flat Signal Rule: A comprehensive rule to signal a dead (flat) signal. This can involve checks such as maximum deviation over a period being below a certain value or the slope of a linear fit through the measurements in a period being below a certain value.
- Dry detection: A measurement is considered suspicious if the immersion depth relative to NAP (Normal Amsterdam Level) of the logger plus a tolerance (0.05 m) is higher than the NAP value of the measurement. For this, the immersion depth must be known. Or a measurement is considered suspicious if the raw water pressure series minus the raw air pressure series is less than a certain tolerance.

With these rules and a lot more that can be read in SOURCE. 5 algorithms were made and compared with each other the algorithms make use of different rules, but the different rules/checks are merged together at the end:

B.5. Explanations of the different Machine Learning Algorithms

K- Nearest Neighbors (KNN): For every data point in the feature space, the k-Nearest Neighbor (kNN) algorithm calculates the average distance to its k nearest neighbors and learns a threshold to decide what data points to consider inliers or outliers (Fabrizio Angiulli and Clara Pizzuti. 2002. Fast outlier detection in high dimensional spaces. In PKDD.

k-means clusting: K-means clustering is an unsupervised machine learning algorithm used to partition a dataset into 'k' distinct, non-overlapping clusters. It assigns each data point to the cluster with the nearest mean, optimizing the clusters based on minimizing the variance within each cluster. The process iteratively updates cluster centroids (the mean of all points in a cluster) until it converges, resulting in a set of separated groups with similar data points.

One Class SVM: One-Class SVM was proposed by Bernhard Schölkopf et al. in 2000, and the method is often used to monitor abnormal behavior in data51. The basic idea is to compute a hypersphere with the smallest radius in a sample and to include all samples inside this hypersphere. When this hypersphere is used to classify the dataset, the samples that fall inside the hypersphere are the first class (normal values) and the samples that fall outside the hypersphere are the second class (abnormal values).

Isolation Forrest: The Isolation Forest (iForest) algorithm is an unsupervised anomaly detection method based on random binary trees, which is suitable for continuous data. The algorithm defines an anomaly as "easily isolated anomalous values", which are points that are sparsely distributed and far away from highdensity populations in the feature space. In the feature space, areas with sparse distribution indicate that the probability of occurring the event is very low, so it is inferred that data points distributed in sparse regions are anomalous values.

C. Interview Questions

C.1. Interview Questions BRO

- Kunt u mij meenemen in het tot stand koming van de BRO (Het idee achter de BRO)?
- Hoe gaat de BRO te werk en dan specifiek over grondwaterstandenonderzoek?
- En wat voor data moet er worden opgestuurd naar de BRO (ruwe grondwaterreeksen?/historische reeksen)
- Binnen welk termijn moet de data zijn aangeleverd?
- Binnen welk termijn moet de data zijn gevalideerd (gecontroleerd)?
- Wanneer zijn meetreeksen nou definitief gevalideerd? [Sinds geen unifromiteit voor definitieve controle, misschien is de kwaliteit per check wel anders, sommige partijen alleen handmetingen binnen 5 cm is dan gevalideerd andere kijken er naar met een hydroloog]
- Hoe werken de verschillende betrouwbaarheid/beoordelings labels
- Is er niet een protocol die de BRO aanraadt om te gebruiken en is het misschien een idee om een uniform protocol te valideren
- Wat zijn punten waar de BRO nog op kan verder ontwikkelen

C.2. Interview questions water authorities

Protocol

Hoe ziet jullie validatie protocol eruit? (Kunnen tekenen?)

Metingen

- Hoe wordt de meting van grondwaterstanden uitgevoerd? Wat is de frequentie van metingen en hoeveel peilbuizen zijn in bezit/ worden gemonitord?
- Welke specifieke apparatuur wordt gebruikt voor de metingen, zoals Divers of Kellers?

Voorlopige validatie

- Op welke punten wordt de meetdata gevalideerd?

  (VOORBEELD PUNTEN)

- Gedupliceerde waarden
- Metingen met een NaN- of een zero waarde
- Correctie naar Tempratuur/Neerslag
- Correctie naar luchtdruk
- Bovenkant van de filter moet hoger zijn dan de onderkant van het filter
- Bovenkant van de filter mag niet hoger zijn dan de bovenkant van de peilbuis
- …

- Welke van deze (genoemde punten door partij) worden er al automatisch uitgevoerd?
- Hoe gaat die automatische validatie in zijn werk?
- Waar loop je tegen aan bij het voorlopige/ automatische valideren

Definitive validatie/ plausibilteitscheck

- Doen jullie een definitieve/ plausibiliteit check?
- Hoe doorlopen/welke stappen jullie je plausibiliteit check?

(Voorbeeld punten waarop gelet wordt)

- Metingen boven de bovenkant van de peilbuis liggen
- Metingen onder de onderkant van het filter liggen
- Geen drift/bias plaatsvind
- Geen filterwissel plaatsvind
- De verandering van het patroon gedurende de meetperiode
- Uitschieters
- Vreemde metingen

- Hoeverre is dit al geautomatiseerd?
- Zijn jullie van plan in de toekomst meer stappen te automatiseren?

D. Summary of the conducted interviews

D.1. Summary Interview Artesia

Artesia's semi-automatic validation system for groundwater measurements, focusing on manual validation as a primary method, supplemented by semi-automatic checks in ArtDiver. Key validation points include checking measurements against physical possibilities, such as the position above the monitoring well or below the filter, and looking for pattern changes, outliers, or unusual readings. Special attention is given to the measurement setup, especially in challenging environments like polders, and ensuring the accuracy of data like cable length and pressure. Artesia emphasizes the importance of accurate data collection and validation to maintain reliable groundwater monitoring, considering the specific challenges and errors that can occur in different measurement setups and environmental conditions.

D.2. Summary Interview Brabant Water

The Brabant Water validation protocol for groundwater measurement series involves a careful process including an initial quality check followed by a definitive validation or plausibility check. The initial validation examines the data for duplications, NaN or zero values, corrections for temperature/precipitation, air pressure, and ensuring the filter's top and bottom positions. Definitive validation involves checking for measurements outside plausible ranges, ensuring no drift or filter changes, and analysing pattern changes over the measurement period. While Brabant Water has extensive experience in groundwater measurements, the process still involves considerable manual work and assessment. The company emphasizes the importance of proper installation and maintenance of measurement points for long-term reliability and is exploring automation to enhance the validation process.

D.3. Summary Interview AA & Maas

The AA&Maas protocol involves monitoring 1000 telemetric points, with daily data checks using Python scripts. These scripts test against physical limits and predict groundwater levels by modelling precipitation and evaporation data, and comparing with nearby wells. Measurements falling outside expected ranges are flagged for further review. Real-time monitoring allows quick detection and response to issues, differing from quarterly data logger readings. The protocol lacks uniformity in handling erroneous data, with some being removed or interpolated. Data is pre-validated before being sent to the BRO (Dinoloket), using MOSGeo pressure sensors that compensate for air pressure, with each well having its own time series model. Challenges include insufficient staffing for visual checks of the measurement series and reliance on manual validation. The reliability of models is considered good. Further automation is sought, particularly in detecting dry measurements, but manual intervention remains necessary for issues like zero-point shifts.

D.4. Summary Interview Blik

The document describes Blik's protocol for validating groundwater measurement series. Blik's process involves preliminary validation and sharing the data through their dashboard. They utilize their measuring equipment and telemetry networks with hourly frequency. Upon receiving data, Blik checks for completeness and converts it into groundwater levels, considering the sensor's depth and location. Validation includes checking against physical impossibilities and visualizing data based on validation status, with errors highlighted in red. They face challenges in large-scale validation and aim for future automation, including a QC protocol for standardized data labelling. Blik monitors 1700 wells and uses proprietary software for installation, emphasizing calibration against air pressure and direct hand measurements for accuracy.