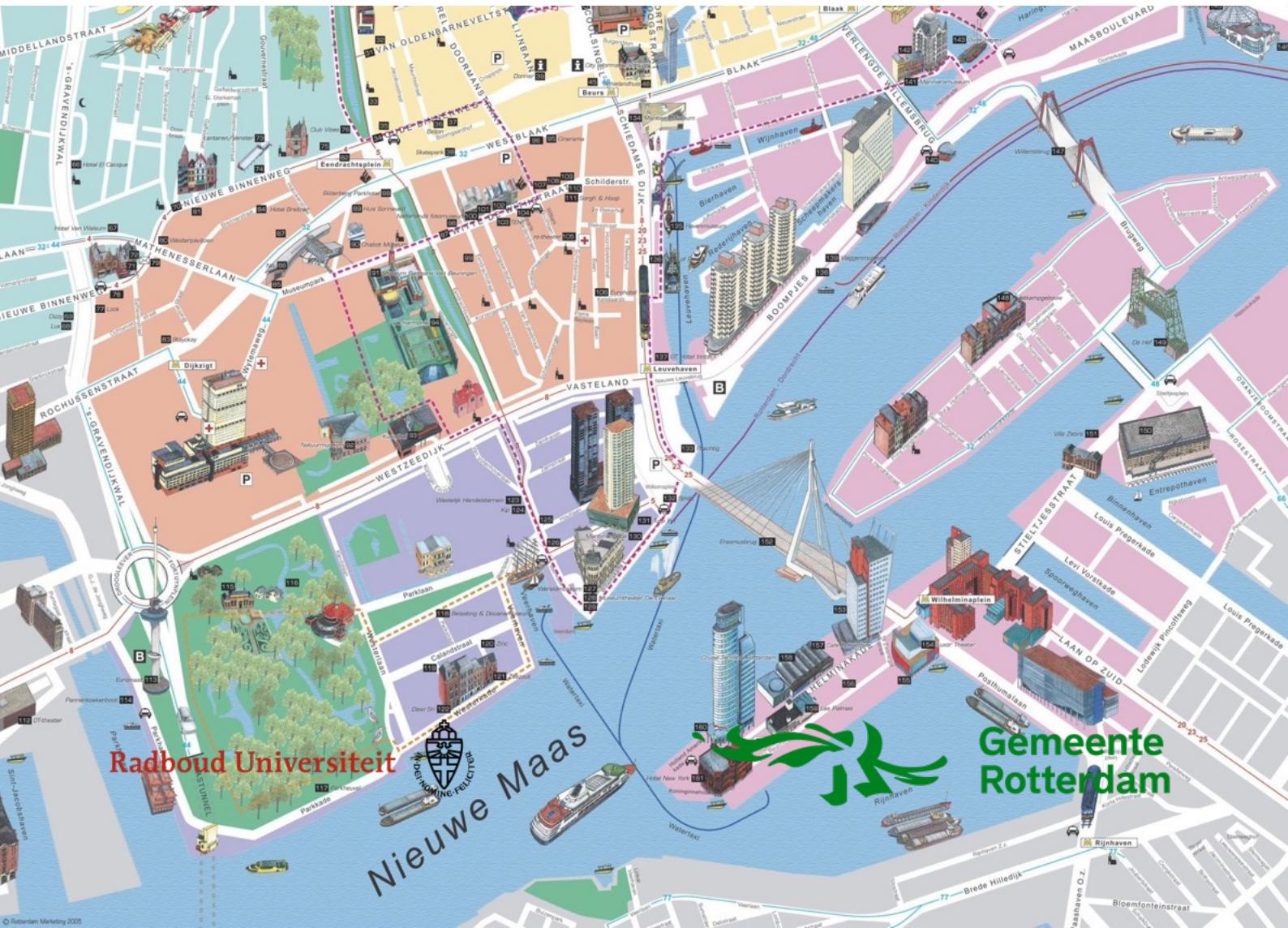


# Rotterdam's Groundwater Monitoring Network

A Critical Examination of the  
Groundwater Monitoring Network  
and its Optimization Potential

Master Thesis  
H.W. Elzinga



# Rotterdam's Groundwater Monitoring Network

A Critical Examination of the  
Groundwater Monitoring Network  
and its Optimization Potential

by

H.W. Elzinga

Hannah Willemijn Elzinga

Student Number 1070923

First Assessor: Dr. Gertjan Geerling  
Second Assessor: Dr. Ekkehard Christoffels  
External Supervisors: Bert de Doelder, Koen Geul  
Project Duration: September 20, 2023 - April 9, 2024  
Faculty: Faculty of Science, Radboud University - Faculty of Biology, University Duisburg-Essen

Defense: Wednesday April 17 2024. From 13.30-14.30 at the Rotterdam City Hall

Cover: Rotterdam City Map (2021)

This Master Thesis has been prepared and written according to the Examination Regulations of the Master Program Environmental Sciences at the Radboud University Nijmegen and the Examination Regulations of the Master Program Transnational ecosystem-based Water Management (TWM) at University Duisburg-Essen. This includes all experiments and studies carried out for the Master Thesis.

The Thesis has been written at Gemeente Rotterdam/Municipality of Rotterdam.

First Assessor: Dr. Gertjan Geerling

Second Assessor: Dr. Ekkehard Christoffels

Head of Examination Board at 1st Assessor's University:

**Declaration**

I declare that I have prepared this Master Thesis self-dependent according to §16 of the Examination Regulation of the Master's Program Transnational ecosystem-based Water Management (TWM) published at 9 August 2005 at the Faculty of Biology at the University Duisburg-Essen.

I declare that I did not use any other means and resources than indicated in this thesis. All external sources of information have been indicated appropriately in the text body and listed in the references.

An electronical version of the master thesis, the model, and the results of the model are available in Github.

*H.W. Elzinga  
Rotterdam, April 2024*

# Preface

In front of you lies my master thesis with the title: 'Rotterdam's Groundwater Monitoring Network: A Critical Examination of the Groundwater Monitoring Network and its Optimization Potential'. This master thesis has been written to fulfill the graduation requirements of the double degree program Transnational ecosystem-based Water Management at Radboud University Nijmegen, The Netherlands, and University Duisburg-Essen, Germany. My research and writing for this thesis spanned from September 20, 2023, to April 9, 2024.

This thesis highlights the remarkable moment of submitting my master thesis on April 9 2024, when I started my initial master thesis exactly one year ago. Due to unfortunate circumstances the thesis was paused. I am very grateful I could continue the end phase of my studies at City of Rotterdam on the topic of geohydrology.

I would like to thank my in-company supervisors Bert de Doelder and Koen Geul for their guidance and support during the process. As well as the colleagues at the department of Urban Water, who supported and advised me. Besides that, external advice from CWG Engineers, Deltares, and Bewonersorganisatie Hillegersberg-Schiebroek. Lastly, Dr. Gertjan Geerling as first assessor from Radboud University Nijmegen and Deltares as well as the second assessor Dr. Ekkehard Christoffels from University Duisburg-Essen and IBC Ingenieure for their participation.

Finally, many thanks to my family and friends who supported me throughout the 2.5 years of the masters program.

Enjoy your reading.

*H.W. Elzinga  
Rotterdam, April 2024*

# Summary

City of Rotterdam oversees an extensive network of approximately 2000 groundwater monitoring wells, which are crucial for monitoring the groundwater dynamics across both urban and rural settings. The network is essential for addressing issues on foundations, subsidence, and (ground)water nuisance in basements throughout the municipal area, but is not intended to be used on a household level. Despite the distribution of the wells across the neighborhoods, the uniformity of the monitoring wells differs across them. City of Rotterdam only investigates the coverage of the neighborhoods when a well is expired: the necessity of well replacement is only considered. However, a reduction or extension optimization approach is not considered. This resulted in a stagnant number of monitoring wells over the past years, prompting questions regarding the level of optimization of the network in capturing comprehensive groundwater level data. Groundwater plays a pivotal role in the urban hydrological cycle. Consequently, the research aims to examine the level of optimization, questioning the density of the network and the location of the monitoring wells throughout the case study areas Rozenburg and Heijplaat within City of Rotterdam.

The methodology for evaluating Rotterdam's groundwater monitoring network (GWMN) employs a two-part design: descriptive and quasi-experimental. At first, the approach characterizes the current state of the GWMN by evaluating a hydrogeological system description and by visualizing the well distribution and groundwater variables without manipulating the existing groundwater level data. Subsequently, the network is manipulated by simulating groundwater level data based on the Python package *Pastas* that is used to incorporate time series modeling of groundwater level data. QR factorization optimizes the network by identifying the informational relevance of the monitoring wells, based on a hierarchical list. The approach enables a data-driven approach to enhance monitoring network efficiency across the case study areas Rozenburg and Heijplaat.

The research study critically examines the optimization of Rotterdam's GWMN through a time-focused analysis and reduction-optimization model. The observed data from 2010 to 2024 highlights the importance of choosing an appropriate time frame for capturing groundwater dynamics accurately. Utilizing *Pastas*, the study bridges data gaps and employs QR factorization for strategic well reduction. The study finds the statement that an optimal GWMN does not necessarily rely on a high data quantity, but on strategic placement of monitoring wells and data quality. A 25 percent reduction emerged as a suitable approach, enhancing the network's efficiency without compromising data integrity. The methodology offers a model for optimizing GWMNs, applicable across Rotterdam's diverse neighborhoods, aligning with the Pareto Optimum principle for maximizing informational relevance and minimizing expenses. Overall, the study underscores the balance between monitoring network density and data quality, revealing an approach to a more optimized GWMN, but also addresses implications that extend beyond the municipal area and aims towards future-oriented urban water management.

# Samenvatting

Gemeente Rotterdam houdt toezicht op een uitgebreid netwerk van ongeveer 2000 peilbuizen die cruciaal zijn voor het monitoren van de grondwaterdynamiek in zowel stedelijk als landelijk gebied. Het netwerk is essentieel voor de aanpak van vraagstukken met betrekking tot funderingen, verzakkingen en (grond) wateroverlast in kelders in het gehele gemeentegebied, maar is niet bedoeld voor gebruik op het niveau van huishoudens. Ondanks de verdeling van de peilbuizen over de wijken verschilt de uniformiteit per wijk. Gemeente Rotterdam onderzoekt alleen de dekking van de buurten wanneer een peilbuis is verlopen: er wordt gekeken naar de noodzaak van vervanging. Mogelijkheden voor reduc tie of uitbreiding van de peilbuizen wordt niet onderzocht. Dit resulteerde de afgelopen jaren in een stagnerend aantal peilbuizen, wat vragen oproep over de mate van optimalisatie van het netwerk bij het vastleggen van uitgebreide grondwaterdata. Grondwater speelt een cruciale rol in de stedelijke hydrologische cyclus. Het onderzoek heeft daarom als doel het niveau van optimalisatie te onderzoeken, waarbij de dichtheid van het netwerk en de locatie van de peilbuizen in de onderzoeksgebieden Rozenburg en Heijplaat binnen de Gemeente Rotterdam onderzocht worden.

De methodologie voor het evalueren van het Rotterdamse achtergrondmeetnet maakt gebruik van een tweedelig ontwerp: beschrijvend en quasi-experimenteel. Als eerste wordt de huidige toestand van het netwerk behandeld, waar een hydrogeologische systeembeschrijving en verdeling van de peilbuizen beschreven en gevisualiseerd wordt zonder de bestaande grondwaterdata te manipuleren. In de tweede stap wordt het netwerk gemanipuleerd door grondwaterdata te simuleren op basis van het Python pakket *Pastas*, dat gebruikt wordt voor tijdreeksanalyse van grondwaterdata. QR factorisatie gebruikt een optimalisatie methode door de informatie relevantie van de peilbuizen te identificeren op basis van een hiërarchische lijst. De aanpak maakt gebruik van een data gestuurde aanpak, mogelijk om de efficiëntie van het meetnet in de onderzoeksgebieden Rozenburg en Heijplaat te analyseren en te verbeteren.

Het onderzoek analyseert kritisch de mate van optimalisatie van het Rotterdamse achtergrondmeetnet door middel van tijdreeksanalyse en een reductie-optimalisatiemodel. De geobserveerde data van 2010-2024 benadrukt het belang van het kiezen van een geschikte onderzoeksperiode voor het nauwkeurig vastleggen van de grondwaterdynamiek. Met behulp van *Pastas* overbrugt het onderzoek datalacunes, waarna in de methode gebruik gemaakt wordt van QR factorisatie voor een strategische peilbuisreductie. De studie komt tot de conclusie dat een optimaal GWMN niet noodzakelijkerwijs afhankelijk is van een grote kwantiteit data, maar van datakwaliteit en strategische plaatsing van peilbuizen. Een reductiepercentage van 25 procent bleek een geschikte aanpak, waardoor de efficiëntie van het netwerk verbeterd kan worden zonder de data integriteit in gevaar te brengen. De methodologie biedt een model voor het optimaliseren van het GWMN, toepasbaar binnen de Gemeente Rotterdam en in lijn met het Pareto Optimum principe voor het maximaliseren van de informatie relevantie en het minimaliseren van de kosten van het GWMN. De studie markeert de balans tussen het monitoren van de netwerkdichtheid en de gegevenskwaliteit, waardoor een aanpak voor een meer geoptimaliseerd GWMN worden onthuld, maar ook de implicaties worden behandeld die verder reiken dan het gemeentelijk gebied en gericht zijn op toekomstgericht stedelijk waterbeheer.

# Contents

<b>Preface</b>	<b>iii</b>
<b>Summary</b>	<b>iv</b>
<b>Samenvatting</b>	<b>v</b>
<b>Nomenclature</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Problem statement . . . . .	1
1.3 Research objective and questions . . . . .	2
1.4 Social and scientific relevance . . . . .	3
1.5 Scope and delimitations . . . . .	3
1.6 Reading guide . . . . .	3
<b>2 Literature Review</b>	<b>4</b>
<b>3 Theoretical Background</b>	<b>6</b>
3.1 System description . . . . .	6
3.1.1 Geohydrological system description . . . . .	6
3.1.2 Ecohydrological system description . . . . .	8
3.2 Hydrological cycle in urban areas . . . . .	8
3.3 Collection of groundwater level data . . . . .	9
3.4 Groundwater level monitoring . . . . .	11
3.5 Groundwater management and policies . . . . .	11
<b>4 Research Methodology</b>	<b>13</b>
4.1 Research design . . . . .	13
4.1.1 Descriptive design . . . . .	13
4.1.2 Quasi-experimental design . . . . .	13
4.2 Data collection and preparation . . . . .	13
4.2.1 QGIS and PROWAT . . . . .	13
4.2.2 Pastas time series modeling . . . . .	14
4.3 QR factorization . . . . .	15
4.4 Data analysis . . . . .	15
<b>5 Case Study</b>	<b>16</b>
5.1 Rozenburg . . . . .	16
5.2 Heijplaat . . . . .	19
<b>6 Results</b>	<b>22</b>
6.1 Rozenburg . . . . .	22
6.1.1 QGIS and PROWAT . . . . .	22
6.1.2 Pastas time series modeling . . . . .	23
6.1.3 QR factorization . . . . .	26
6.2 Heijplaat . . . . .	34
6.2.1 QGIS and PROWAT . . . . .	34
6.2.2 Pastas time series modeling . . . . .	34
6.2.3 QR factorization . . . . .	38
<b>7 Discussion</b>	<b>47</b>
<b>8 Conclusion</b>	<b>50</b>

<b>9 Recommendations</b>	<b>53</b>
<b>References</b>	<b>54</b>
<b>A Pastas Time Series Modeling</b>	<b>57</b>
A.1 Results of Pastas time series modeling . . . . .	57
A.1.1 Stress Model 1: Rozenburg . . . . .	58
A.1.2 Stress Model 1: Heijplaat . . . . .	60
A.1.3 Stress Model 2: Rozenburg . . . . .	62
A.1.4 Stress Model 2: Heijplaat . . . . .	64
<b>B Mathematical Background</b>	<b>66</b>
B.1 Python Packages . . . . .	66
B.1.1 Pastas . . . . .	66
B.1.2 PySensors . . . . .	66
B.2 Statistics . . . . .	67
B.2.1 Welch's t-Test . . . . .	67
B.2.2 QR factorization . . . . .	68
B.2.3 Genetic Algorithm . . . . .	68

# List of Figures

1.1 Main research question and five sub questions. . . . .	2
3.1 Regional groundwater flow in urban areas [41]. . . . .	7
3.2 Schematic graph of an unconfined (left) and confined aquifer (right) [26]. . . . .	9
3.3 Hydrological situation and process in the western part of the municipal area of Rotterdam translated, based on Dufour [10]. . . . .	10
3.4 Overview of the urban water system according to the Municipal Sewage Plan of Rotterdam [18]. . . . .	12
5.1 Digital terrain model (DTM 0.5 m) for Rozenburg. Areas marked in red indicate locations with the highest elevation, while orange marked areas indicate locations with the lowest elevation. . . . .	17
5.2 Profile of drill site profile with geological formations for the location x: 75250; y: 436180 [47]. . . . .	17
5.3 Precise location for the drilling profile by the coordinates x: 75250, y: 436180 [47]. . . . .	18
5.4 Groundwater monitoring network in Rozenburg. Based on EPSG:28992/Amersfoort RD New. . . . .	18
5.5 Digital terrain model (DTM 0.5) for Heijplaat. Areas marked red indicate the locations with the highest elevation, while blue marked areas indicate locations with the lowest elevation. . . . .	19
5.6 Profile of drill site profile with geological formations for the location x: 88529; y: 433870 [47]. . . . .	20
5.7 Precise location for the drilling profile by the coordinates x: 88529, y: 433870 [47]. . . . .	20
5.8 Groundwater monitoring network in Heijplaat. Based on EPSG:28992/Amersfoort RD New. . . . .	21
6.1 Observed data for 29 monitoring wells in Rozenburg. . . . .	22
6.2 Recharge [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands. . . . .	23
6.3 Precipitation [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands. . . . .	23
6.4 Potential evaporation [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands. . . . .	23
6.5 Reversed forecasting data based on datalogger data for monitoring well 112566-5. . . . .	24
6.6 Reversed forecasting data based on manual measurement data for monitoring well 112566-5. . . . .	24
6.7 Reversed forecasting data based on manual measurement data and datalogger data for monitoring well 112566-5. . . . .	24
6.8 Bar plot showing the metrics RMSE and $R^2$ for 29 monitoring wells. . . . .	24
6.9 Bar plot showing metrics RMSE and $R^2$ for 29 monitoring wells with a division in data logger and manual collection method. . . . .	24
6.10 Bar plot showing metric EVP for 29 monitoring wells. . . . .	24
6.11 Overview of the performance of RMSE, $R^2$ , and EVP after the Welch's t-test and the difference between data loggers and manual collection method. . . . .	25
6.12 Scatter plot of simulated and observed data by data loggers for monitoring well 112566-5. The x-axis shows a period of 2020-2024 and the y-axis shows a range of -0.5 to 0.6 meters NAP. . . . .	26

---

6.13 Groundwater level [m NAP] across Rozenburg, using a color-coded system to represent the mean GWL observations for the 29 monitoring wells. Based on EPSG:28992/Amersfoort RD New. . . . .	26
6.14 Visualization of the hierarchical list of 29 monitoring wells across Rozenburg. Based on EPSG:28992/Amersfoort RD New. . . . .	27
6.15 RMSE based on a network reduction of 10%. The optimal number of wells is 26. . . . .	28
6.16 RMSE based on a network reduction of 25%. The optimal number of wells is 21. . . . .	28
6.17 RMSE based on a network reduction of 50%. The optimal number of wells is 14. . . . .	28
6.18 RMSE based on a network reduction of 75%. The optimal number of wells is 7. . . . .	29
6.19 RMSE based on a network reduction of 90%. The optimal number of wells is 2. . . . .	29
6.20 Overview of the RMSE for a range of reductions (10-25-50-75-90%). . . . .	29
6.21 Eliminated monitoring well: 114564-2. . . . .	30
6.22 Eliminated monitoring well: 113565-11. . . . .	30
6.23 Eliminated monitoring well: 112565-6. . . . .	30
6.24 Eliminated monitoring well: 113565-2. . . . .	30
6.25 Eliminated monitoring well: 113565-12. . . . .	30
6.26 Eliminated monitoring well: 113564-3. . . . .	30
6.27 Eliminated monitoring well: 112566-8. . . . .	30
6.28 Eliminated monitoring well: 112565-7. . . . .	30
6.29 Performance metric $R^2$ for Rozenburg. The orange line indicates the median value. The y-axis ranges from 0.785 to 0.952 [-]. . . . .	31
6.30 Mean Absolute Error [meters] across Rozenburg, using a color-coded system to represent the calculated MAE for the eliminated monitoring wells. Based on EPSG:28992/Amersfoort RD New. . . . .	32
6.31 Map depiction of the remaining monitoring wells in Rozenburg after following a 25% reduction of the network. Based on EPSG:28992/Amersfoort RD New. . . . .	33
6.32 Observed data for 14 monitoring wells in Heijplaat. . . . .	34
6.33 Recharge [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands. . . . .	35
6.34 Precipitation [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands. . . . .	35
6.35 Potential evaporation [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands. . . . .	35
6.36 Reversed forecasting data based on data logger data for monitoring well 125563-97. . . . .	35
6.37 Reversed forecasting data based on manual measurement data for monitoring well 125563-97. . . . .	35
6.38 Reversed forecasting data based on manual measurement data and datalogger data for monitoring well 125563-97. . . . .	35
6.39 Bar plot showing the metrics RMSE and $R^2$ for 14 monitoring wells. . . . .	36
6.40 Bar plot showing metrics RMSE and $R^2$ for 14 monitoring wells with a division in data logger and manual collection method. . . . .	36
6.41 Bar plot showing metric EVP for 14 monitoring wells. . . . .	36
6.42 Overview of the performance of RMSE, $R^2$ , EVP after the Welch's t-test and the difference between data loggers and manual collection method. . . . .	37
6.43 Scatter plot of simulated and observed data by data loggers for monitoring well 125563-97. The x-axis shows a period of 2010-2024 and the y-axis shows a range of +1.25 to 3.0 meters NAP. . . . .	37
6.44 Groundwater level [m NAP] across Heijplaat, using a color-coded system to represent the mean GWL observations for 14 monitoring wells. Based on EPSG:28992/Amersfoort RD New. . . . .	38
6.45 Visualization of the hierarchical list of 14 monitoring wells across Heijplaat. Based on EPSG:28992/Amersfoort RD New. . . . .	40
6.46 RMSE based on a network reduction of 10%. The optimal number of wells is 12. . . . .	41
6.47 RMSE based on a network reduction of 25%. The optimal number of wells is 10. . . . .	41
6.48 RMSE based on a network reduction of 50%. The optimal number of wells is 7. . . . .	41
6.49 RMSE based on a network reduction of 75%. The optimal number of wells is 3. . . . .	42

---

6.50 RMSE based on a network reduction of 90%. The optimal number of wells is 1 . . . . .	42
6.51 Overview of RMSE for a range of reduction (10-25-50-75-90%) . . . . .	42
6.52 Eliminated monitoring well: 125563-111 . . . . .	43
6.53 Eliminated monitoring well: 125564-1 . . . . .	43
6.54 Eliminated monitoring well: 125563-113 . . . . .	43
6.55 Eliminated monitoring well: 125564-3 . . . . .	43
6.56 Performance metric $R^2$ . The orange line indicates the median value. The y-axis ranges from 0.928 to 0.987 [] . . . . .	44
6.57 Mean Absolute Error [meters] across Heijplaat, using a color-coded system to represent the calculated MAE for the eliminated monitoring wells. Based on EPSG:28992/Amersfoort RD New . . . . .	45
6.58 Map depiction of the remaining monitoring wells in Heijplaat after following a 25% reduction of the network. Based on EPSG:28992/Amersfoort RD New . . . . .	46
A.1 Monitoring well: 112566-3 . . . . .	58
A.2 Monitoring well: 112566-5 . . . . .	58
A.3 Monitoring well: 112566-1 . . . . .	58
A.4 Monitoring well: 113566-1 . . . . .	58
A.5 Monitoring well: 112565-3 . . . . .	58
A.6 Monitoring well: 112565-5 . . . . .	58
A.7 Monitoring well: 113565-6 . . . . .	58
A.8 Monitoring well: 113565-1 . . . . .	58
A.9 Monitoring well: 112565-8 . . . . .	58
A.10 Monitoring well: 113565-2 . . . . .	59
A.11 Monitoring well: 112565-4 . . . . .	59
A.12 Monitoring well: 113565-7 . . . . .	59
A.13 Monitoring well: 113565-3 . . . . .	59
A.14 Monitoring well: 113565-4 . . . . .	59
A.15 Monitoring well: 113565-5 . . . . .	59
A.16 Monitoring well: 114565-1 . . . . .	59
A.17 Monitoring well: 113564-3 . . . . .	59
A.18 Monitoring well: 114564-2 . . . . .	59
A.19 Monitoring well: 113565-11 . . . . .	59
A.20 Monitoring well: 113565-10 . . . . .	59
A.21 Monitoring well: 113565-12 . . . . .	59
A.22 Monitoring well: 112565-7 . . . . .	59
A.23 Monitoring well: 114565-4 . . . . .	59
A.24 Monitoring well: 112565-6 . . . . .	59
A.25 Monitoring well: 1145654-3 . . . . .	60
A.26 Monitoring well: 112566-6 . . . . .	60
A.27 Monitoring well: 112566-8 . . . . .	60
A.28 Monitoring well: 114564-7 . . . . .	60
A.29 Monitoring well: 114565-5 . . . . .	60
A.30 Monitoring well: 125564-1 . . . . .	60
A.31 Monitoring well: 125564-3 . . . . .	60
A.32 Monitoring well: 125564-4 . . . . .	60
A.33 Monitoring well: 125563-94 . . . . .	61
A.34 Monitoring well: 125563-95 . . . . .	61
A.35 Monitoring well: 125563-96 . . . . .	61
A.36 Monitoring well: 125563-97 . . . . .	61
A.37 Verander: Monitoring well: 125564-114 . . . . .	61
A.38 Monitoring well: 125563-113 . . . . .	61
A.39 Monitoring well: 125563-92 . . . . .	61
A.40 Monitoring well: 125563-93 . . . . .	61
A.41 Monitoring well: 125563-111 . . . . .	61
A.42 Monitoring well: 125563-112 . . . . .	61

---

A.43 Monitoring well: 125563-115.	61
A.44 Monitoring well: 112565-3.	62
A.45 Monitoring well: 112565-4.	62
A.46 Monitoring well: 112565-5.	62
A.47 Monitoring well: 112565-6.	62
A.48 Monitoring well: 112565-7	62
A.49 Monitoring well: 112566-1.	62
A.50 Monitoring well: 112566-3.	62
A.51 Monitoring well: 112566-5.	62
A.52 Monitoring well: 112565-6.	62
A.53 Monitoring well: 112565-7.	63
A.54 Monitoring well: 112566-1.	63
A.55 Monitoring well: 112566-3.	63
A.56 Monitoring well: 112566-5.	63
A.57 Monitoring well: 112566-6.	63
A.58 Monitoring well: 112566-8.	63
A.59 Monitoring well: 113564-3.	63
A.60 Monitoring well: 113565-1.	63
A.61 Monitoring well: 113565-2.	63
A.62 Monitoring well: 113565-3.	63
A.63 Monitoring well: 113565-4.	63
A.64 Monitoring well: 113565-5.	63
A.65 Monitoring well: 113565-6.	63
A.66 Monitoring well: 113565-7.	63
A.67 Monitoring well: 113565-8.	63
A.68 Monitoring well: 113566-1.	64
A.69 Monitoring well: 114564-2.	64
A.70 Monitoring well: 114564-3.	64
A.71 Monitoring well: 114564-7.	64
A.72 Monitoring well: 114565-1.	64
A.73 Monitoring well: 125563-111.	64
A.74 Monitoring well: 125563-92.	64
A.75 Monitoring well: 125563-93.	64
A.76 Monitoring well: 125563-94.	65
A.77 Monitoring well: 125563-95.	65
A.78 Monitoring well: 125563-97.	65
A.79 Monitoring well: 125563-112.	65
A.80 Monitoring well: 125563-113.	65
A.81 Monitoring well: 125563-96.	65
A.82 Monitoring well: 125563-114.	65
A.83 Monitoring well: 125563-115.	65
A.84 Monitoring well: 125564-3.	65
A.85 Monitoring well: 125564-4.	65
A.86 Monitoring well: 125564-1.	65
B.1 Formula for a basic time series model [3].	66
B.2 Formula of the Sparse Sensor Placement Optimization for Reconstruction algorithm [30].	67
B.3 Formula to calculate t in the Welch's t-test [27].	67
B.4 QR factorization or decomposition of the matrix A into an orthogonal matrix Q and triangular matrix R [20].	68
B.5 Formula to determine the genetic algorithm in the optimization model [29].	68

# List of Tables

6.1	Summary of sampling data. . . . .	27
6.2	Overview of statistical results of the eliminated monitoring wells with a network reduction of 25%. The overview explains the p-value of the monitoring wells and if they experience a significant difference between the observed and reconstructed data. . . . .	31
6.3	Summary of sampling data. . . . .	39
6.4	Overview of statistical results of the eliminated monitoring wells with a network reduction of 25%. The overview explains the p-value of the monitoring wells if they experience a significant difference between the observed and reconstructed data. . . . .	43
8.1	Overview of the ratio monitoring well to surface area across study areas Rozenburg and Heijplaat. . . . .	52

# Nomenclature

## Abbreviations

Abbreviation	Definition
ANN	Artificial Neural Network
EPSG	European Petroleum Survey Group
EVP	Explained Variance Percentage
GA	Genetic Algorithm
GIS	Geographical Information System
GWL	Groundwater Level
GWMN	Groundwater Monitoring Network
GWMNO	Groundwater Monitoring Network Optimization
KIT	Karlsruhe Institute of Technology
KNMI	<i>Koninklijk Nederlands Meteorologisch Instituut</i> - Royal Netherlands Meteorological Institute
MAE	Mean Absolute Error
MSP	Municipal Sewage Plan
NAP	<i>Normaal Amsterdams Peil</i> - Amsterdam Ordnance Datum
NEN	<i>Nederlandse Norm</i> - Dutch Standard
PCA	Principal Component Analysis
$R^2$	R-squared, coefficient of determination
RMSE	Root Mean Square Error
rBIAS	relative Bias
SIKB	<i>Stichting Infrastructuur Kwaliteitsborging Bodembeheer</i> - Soil Management Infrastructure Quality Assurance Foundation
SSPOC	Sparse Sensor Placement Optimization for Classification
WWTP	Waste Water Treatment Plant

## Symbols

Symbol	Definition	Unit
$A$	Matrix	[ $\cdot$ ]
$\delta$	System inertia	[t]
$d$	Base level of the model	
$E_a$	Actual evaporation	[m/day]
$E_p$	Potential evaporation	[m/day]
$f$	Evaporation factor	[ $\cdot$ ]
$H_{average}$	Average groundwater level	[m NAP]
$hm(t)$	Contribution of influence $m$	[t]
$M$	Total influence that contributes to groundwater level fluctuations	[ $\cdot$ ]
$n$	Size of sample	[ $\cdot$ ]
$P$	Precipitation	[m/day]
$Q$	Orthogonal matrix	[ $\cdot$ ]

$r$	Model residues	[ $\cdot$ ]
$R$	Recharge rate	[m/day]
$R$	Triangular matrix	[ $\cdot$ ]
$s$	Variance of the sample	[ $\cdot$ ]
$u$	Population mean	[ $\cdot$ ]
$\omega$	Recharge sensitivity at time $\tau$	[ $t^{-1}$ ]
$x$	Mean of the sample	[ $\cdot$ ]

---

# 1

## Introduction

### 1.1. Background

Over the years, many cities have gone through a transformation regarding surface and subsurface measures with the aim to mitigate and adapt to the consequences of climate change. Due to this transformation, the already heterogeneous and complex urban groundwater system faces an additional challenge regarding climate change. Two aspects that characterize the behavior of the urban groundwater system include surface sealing and systematic drainage of urban areas, resulting in surface runoff of precipitation to the sewage system next to replenishing the groundwater quantity through infiltration. (Ground)water nuisance can occur in urban areas, because the process of urbanisation does not include sustainable management of groundwater levels throughout the city [50].

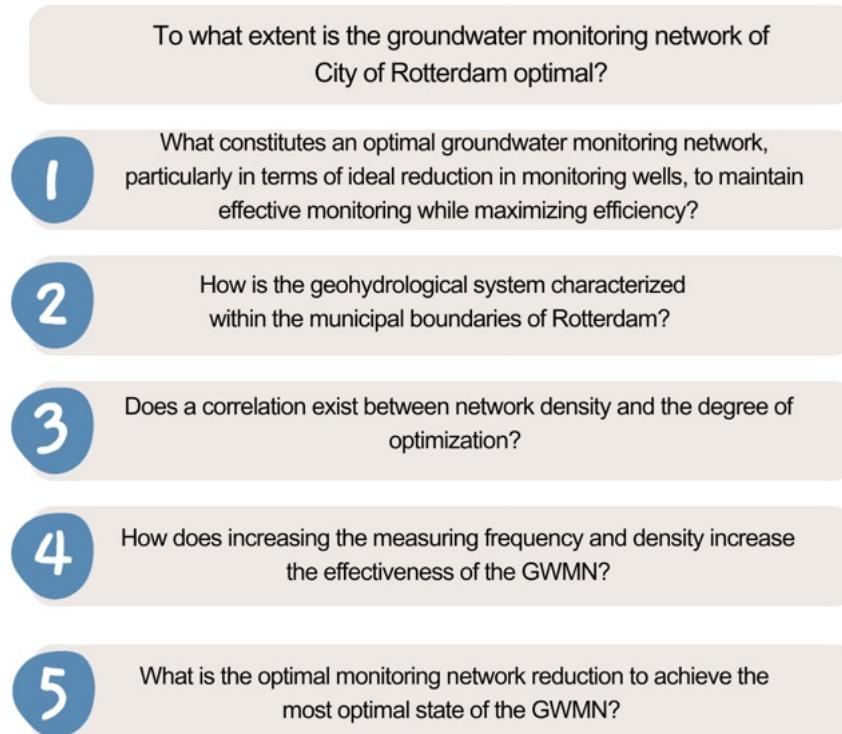
City of Rotterdam manages an extensive groundwater monitoring network, consisting of approximately 2000 groundwater wells. The wells cover the entire municipal area of Rotterdam, reaching from urban to rural locations. The network is essential for addressing issues on for example foundations, subsidence, and (ground) water nuisance in basements throughout the municipal area, but is not intended to be used on a household level. Besides, monitoring groundwater data is crucial for understanding the groundwater dynamics within the municipality. The observed data by the municipality is used for civil technical measures and asset management. The construction of monitoring wells in public areas falls under municipal jurisdiction. Given its responsibility, it is essential not only to monitor groundwater levels but also to evaluate if the network density and well locations are efficient for optimal results [19].

### 1.2. Problem statement

Despite the distribution of the monitoring wells throughout the neighborhoods of Rotterdam, the uniformity of the monitoring wells also differs across them. One of the reasons is that City of Rotterdam only investigates the coverage of a subarea when a well is expired: They determine the necessity of well replacement and if the new well can be placed on a more optimal location. Consequently, the number of groundwater wells has been stagnant throughout the years and it is not known if the current number of monitoring wells is the most suitable number to retrieve groundwater information [19]. Since groundwater is an integral part of the hydrological cycle, it is key to monitor and anticipate on the groundwater levels. Therefore, it is important to research the design of the network, analyze whether the network is optimal, and if monitoring wells need to be removed or reconstructed in a later stage [13]. An insufficient groundwater monitoring network could lead to consequences for urban planning, infrastructure, and environmental pollution. An insufficient network might form an obstacle regarding Rotterdam's ability to mitigate and adapt to climate change impacts such as increased precipitation variability [12].

### 1.3. Research objective and questions

Even though the number of monitoring wells has been stagnant over the past years, the network is currently an actively monitored groundwater network through the measurement of data loggers and manual collection. However, it is not known if the network is ideal yet. Therefore, the objective of the research study is to investigate the degree of optimization of the groundwater monitoring network of Rotterdam, see figure 1.1. Incorporating the specified research questions into a methodology that blends theoretical insights with empirical investigations, the following research questions are outlined:



**Figure 1.1:** Main research question and five sub questions.

Summarizing, in the sub questions the following topics are researched and discussed:

- The definition of optimal in the context of a groundwater monitoring network;
- Characteristics of the geohydrological system of the municipal area of Rotterdam through visualization of geological information;
- The correlation between network density and the degree of optimization;
- Implications of optimizing the measuring frequency and coverage area of the monitoring wells;
- Comparison of reduction percentages to determine the optimal reduction rate to achieve an optimal state of the GWMN.

## 1.4. Social and scientific relevance

The hypothesis driving this research suggests that the GWMN of Rotterdam has not reached its ideal state yet. This study is relevant as it explores the GWMN with a focus on an urban and industrial context. The primary goal is to evaluate the level of optimization of the GWMN through the development of a generic model using QR factorization. The findings of the research have the potential to serve as a valuable point of reference for other urban areas that face similar groundwater challenges. Accordingly, it is imperative to examine the GWMN to determine if sufficient and essential groundwater data are collected to make informed decisions for integrated groundwater management. A synergy between policy and execution is necessary to reach optimal groundwater management [23]. The examination should include a critical examination on the spatial distribution and network density of the GWMN as well as the monitoring frequency. Ultimately, the research seeks to develop an approach for monitoring and optimizing Rotterdam's GWMN, offering contributions to societal and scientific fields.

## 1.5. Scope and delimitations

The study focuses and explores on the following:

- The inclusion of active, phreatic monitoring wells;
- The exclusion of project monitoring wells;
- Case study areas Rozenburg and Heijplaat chosen from the municipal area of Rotterdam;
- Project period of 2010-2024 with an optimization approach for 2020-2024;
- Development of a generic model on neighborhood scale.

## 1.6. Reading guide

The study begins with chapter 2 "Literature Review", analyzing various methodologies to select the most suitable approach for this research. Followed by the chapter 3 "Theoretical Background", focusing on the hydrogeological context and management policies in The Netherlands. Chapter 4 discusses the approach, data collection and preparation, followed by two case studies of Rozenburg and Heijplaat. Chapter 6 "Results", details results for both study areas. The study concludes with chapter 7, "Discussion", on the robustness of the model and reliability of the results, followed by chapter 8 "Conclusion" with aims to answer the research questions. The study ends with chapter 9 "Recommendations", highlighting the study limitations and suggesting directions for future research, aiming to enhance the understanding of Rotterdam's GWMN.

# 2

## Literature Review

The importance of monitoring groundwater quantities and collecting comprehensive data on the spatiotemporal dynamics of aquifers is highlighted. Understanding the movement and characteristics of groundwater over time allows the development of sustainable strategies [35]. A perception is the design of an optimal groundwater monitoring network. The ideal network achieves the maximum possible information content at the lowest feasible expenses. The perception is in close alignment with the Pareto Optimum, which can be described as a standard that illustrates a situation without any further improvements of society's well-being [43]. The optimum can be made through a re-allocation of resources that makes at least one person better-off without making someone else worse-off [43]. In the context of groundwater management, the principle suggests creating a system that maximizes efficiency and benefits without incurring additional costs or disadvantages to any of the concerning stakeholders. As written by STOWA, optimizing the density of monitoring wells in a groundwater monitoring network is a process that requires a clear definition of the term 'optimization' [45]. STOWA provides a comprehensive framework, outlining a six-step methodology in order to determine the degree of optimization in monitoring networks [45].

1. The network is optimal when it is able to continuously provide accurate data on the status of the groundwater. *If that is true*, the use of spatiotemporal models as process modeling and geostatistical modeling are recommended.
2. The network is optimal when it is able to translate data into accessible information. *If that is true*, mapping of historical, current, and future data are recommended.
3. The network is optimal when every monitoring well is representative for the water quantity and quality in the area or for a specific water body. *If that is true*, the representativeness of the research area has to be assessed through project-based studies.
4. The network is optimal when one or more area covering images of the water quantity and quality can be obtained. *If that is true*, based on geographical information and system knowledge it can be researched to what extent the measuring points are researched throughout the research area. The use of a Geographic Information System (GIS) is required.
5. The network is optimal when every monitoring well provides unique information without overlapping neighboring monitoring wells. *If that is true*, based on statistical tests it can be researched to what extent the monitoring wells show overlapping in data.
6. The network reaches an optimum when the above-mentioned steps are effectively combined and aligned with the other monitoring objectives.

The development of monitoring networks has been an occasional process that is shaped by varying demands and interests, specific to the time of their construction. Monitoring networks grow historically based on ad-hoc installation of monitoring wells. Installed monitoring wells have acquired the right to exist, but all have their own advantages and disadvantages. Practically, a combination of the previously mentioned six-step methodology is required [45]. It is argued that refining the frequency of measurements at a monitoring well is ineffectual if the monitoring wells lack a rational for its existence. Certain

---

monitoring wells may prove to be suitable for trend analysis, because of its statistical characteristics. Adjustments in measuring density must consider a range of monitoring objectives beyond optimization.

From a research perspective, Mirzaie-Nodoushan proposed paired objectives for the establishment of a groundwater monitoring network [33]. Firstly, to minimize the number of monitoring wells, and secondly, to minimize the root mean square error (RMSE) between observed and simulated data in the aquifer under study. The latter objective focuses on minimizing the average of the squares of all prediction errors.

In a study by former waterboard "Maaskant", presently known as waterboard Aa en Maas, the intention was to complement the regional water system, focusing on the unconfined groundwater monitoring network [31]. In this study, Massop and van der Gaast detail the distribution of the network within the study area and evaluate it against specific criteria. The criteria include the representativeness of the GWMN in relation to the spatial characteristics of the research area and its capability to distinguish between different clusters. The study represents the network into clusters based on observable characteristics, assigning average values for hydrotype, groundwater depth classes, soil type, land use, water supply, and distance ranking. Logically, clusters with analogous characteristics are grouped. For each cluster, key hydrological parameters such as average groundwater level [ $h_{\text{average}}$ ], system inertia [ $\delta$ ], and recharge sensitivity [ $\omega$ ] are calculated, which are influenced by drainage resistance and the storage coefficient, affecting groundwater level reactions to precipitation and surface water level changes. More specifically, the system inertia helps to create an understanding of the relationship between precipitation and the influence on groundwater recharge, while the recharge sensitivity describes the sensitivity of the groundwater system to changes through external influences as precipitation or local pumping activities. The minimum and maximum values for these three parameters are estimated, as well as the median values (50% higher and 50% lower) for every cluster of monitoring wells in the study area. The median value is chosen for the estimation, because the median is less sensitive to outliers in data. The most representative location for each cluster of monitoring wells is determined by the smallest deviation for the cluster [31].

Reproducing the approach by Massop and van der Gaast [31], Hosseini and Kerachian argue that the purpose of groundwater monitoring is to deepen the understanding of the hydrogeological system through continuous data collection of the groundwater quantity and/or quality [24]. The approach divides the research area into clusters and employs hexagonal gridding with use of the Thiessen polygon approach in order to examine spatial sampling patterns. They integrate three criteria to set a base for priority areas for monitoring, based on aquifer traits and groundwater level fluctuations. The approach combines three criteria to identify priority areas for monitoring which are based on aquifer characteristics and groundwater fluctuations (marginal entropy of water table levels; estimation error variances of mean values of water table levels; estimation values of long-term changes in groundwater level). Spatiotemporal kriging and ANN are used for predicting groundwater level variations. Data fusion techniques are deployed to refine prediction precision, reducing the RMSE. The Value Of Information technique is instrumental in identifying monitoring sites that yield maximum informational value within each priority area and suggests varying sampling frequencies. The approach by Hosseini and Kerachian presents a cost-effective strategy for redesigning a GWMN [24].

Overall, the literature on GWMN accentuates a transition from historical development towards optimized configurations. The literature by STOWA [45] and Hosseini and Kerachian [24] reflect consensus on the need for the interpretation of monitoring wells and their data collection in terms of frequency and data utility. While Mirzaie-Nodoushan focuses on minimizing both the number of monitoring wells and the prediction errors [33], Massop and van der Gaast prioritize the representativeness and clustering of monitoring wells to visualize the heterogeneity of the study area [31]. Insights by Ohmer emphasize the role of monitoring in groundwater management that is in line with the Pareto Optimum [35], to achieve the maximum efficiency of the GWMN [43]. There is a clear perspective for the future of groundwater monitoring: Networks that are not only cost-effective and efficient, but also robust in terms of their capacity to provide high quality data. The literature review provides a foundation for the development of a generic model that evaluates the performance of the GWMN on neighborhood scale and is proficient in addressing and handling challenges in groundwater management.

# 3

## Theoretical Background

### 3.1. System description

A system description can be regarded as the verbal model of distinguishable hydrological systems, as they influence nature reserves [14]. Next to the hydrological system, the geological and ecological system will also be included. A system description can be built according to: 1) The description of the local geohydrological system; 2) The description of the regional geohydrological system; 3) The interaction between the local and regional system; 4) The relationship between the groundwater and surface water system [28]. As well as the quantity and quality aspects of the area. Within the system description, emphasis lies on particle and water flow as a function of time and place to the root zone of valuable habitats [14].

#### 3.1.1. Geohydrological system description

To start with the geological formations within the municipal boundaries. Throughout the municipal area, the formation is complex [49]. Meaning that there are strongly heterogeneous formations with local gully erosion and cuts. A heterogeneous formation refers to geological layers that differ in parameters as structure, composition, density, and permeability [49]. The subsoil mainly consists of clay and sandy soils, the distinction is made regarding the permeability of the soils. The presence of creeks has influenced the erosion of peat and sandy channel depositions, resulting in the creation of levees [25]. By the influence of wind throughout the municipal area, river dunes advanced along old river courses. River dunes and levees form sand channels and ensure different groundwater flows. The municipal area of Rotterdam is located in the western part of the Dutch delta. Over the past 10.000 years, the area has been influenced by the North Sea. This can be seen through a layer of 20 meters of mudflat and lagoon deposits from the Holocene period. While the mudflat deposits consist of sand, silt, and clay, the lagoon deposits consist of silt, clay, and peat. Because of the influence of the North Sea, there is much spatial variation in the Holocene cover layer. Previous intrusions of the North Sea within the municipal area resulted in a boundary between fresh and salt groundwater that is found in the upper part of the Holocene layer. The distribution of fresh water and salt water is also influenced by water management, for example land reclamation of polder areas and water abstraction. The morphology of the municipal area is influenced by the presence of the river Meuse, the large scale embankment and excavation of the port and industrial area. Rivers in the lower polder area that are located within the dike ring, cross the polder areas and characterize the groundwater flow situation in Rotterdam. Seepage from the river mainly infiltrates into the first aquifer. River water from the first aquifer occurs as seepage into the lower polder areas. The infiltrating river water flows into the deeper aquifers or is caught for other purposes such as industry [25].

As shown in figure 3.1, the top-soil layer is influenced by human interventions. Consequently, the top-soil layer, or also called the Anthropocene layer, varies in soil composition. The distribution of the top layer is derived from the height of the ground level and top layer of the Holocene cover layer. Underneath the Holocene phreatic package, the overburden layer (Westland) is present. The overburden layer consists of sandy material that is associated with the occurrence of channel fillings or old river

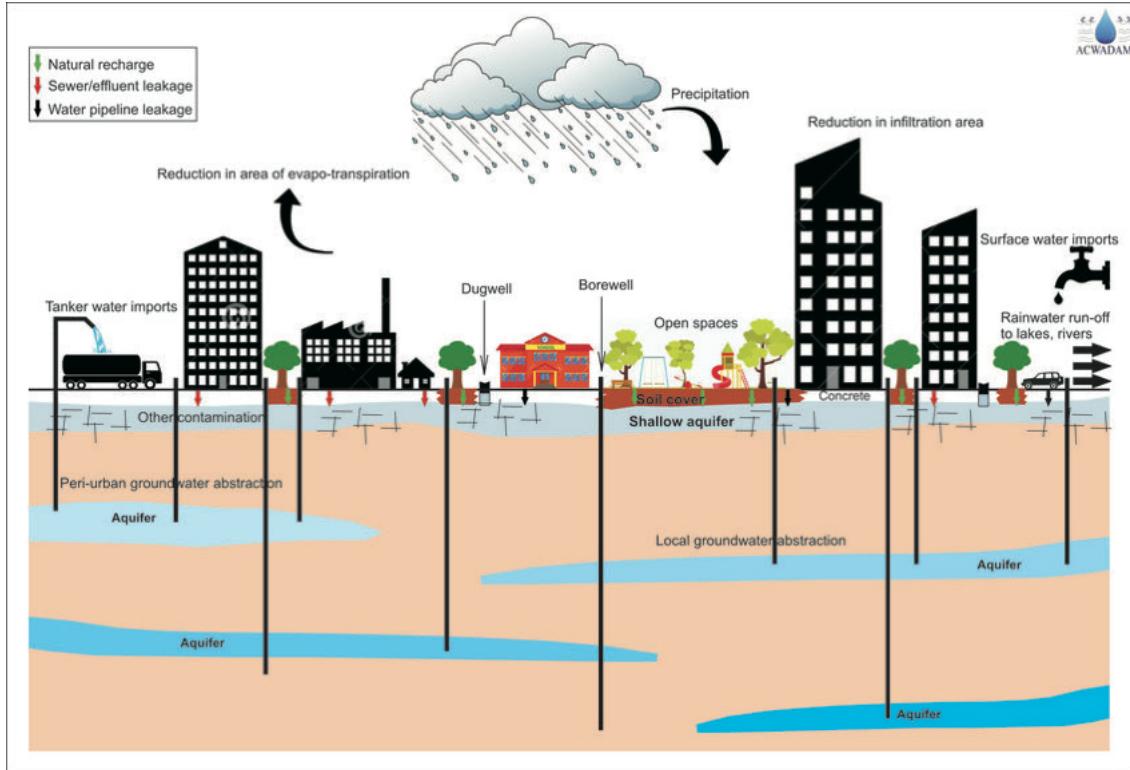


Figure 3.1: Regional groundwater flow in urban areas [41].

dunes, which means that the covering layer has a greater permeability [25]. The Holocene overburden is bounded at the bottom by the first Pleistocene aquifer (Twente, Eem, Urk, and Krefthenhey). The layer consists of fine and coarse sands and gravel with intercalations of sand. In the middle of the first aquifer is a compacted peat layer. The aquifer system is thinner in the south than in the north of the municipal area. Below the first aquifer lies a poorly permeable layer (Kedichem and Tegelen) which is formed by a package of sandy clay's with intercalations of sand. In the middle of the permeable layer is a compacted layer of peat. Below the first permeable layer, a second aquifer is present (Tegelen). The second aquifer consists of fine-medium sandy layers with a high permeability. The second aquifer has a thickness of 30-100 meters. An additional second poorly permeable layer (Tegelen) is present between the second and third aquifer (Maassluis). This layer is made of clay deposits and has a varying thickness between 50-100 meters. The third aquifer characterizes itself by fine to medium sandy layers. The thickness of the third aquifer varies from 50-100 meters. The bottom of the geohydrological system (Pliocene: Formation of Oosterhout) consists of impermeable clay deposits. The second and third aquifer systems are thick in the center, but thinner on the eastern and western side of the municipal area [25]. As previously mentioned, an additional Anthropocene top layer is supplemented on the top layer of the Holocene deposition. Sand-filled vertical drains are used to add vertical drainage, which can accelerate the discharge of infiltrated water into the deep groundwater. Flow occurs from a high hydraulic head to a low hydraulic head. In this situation, flow might occur from the dunes on the western side of the municipal area to the polders farther inland. Higher lying areas can serve as infiltration areas where groundwater flows through the deeper layers to the deeper polders in addition to the hydraulic head differences that causes groundwater flow [49]. Seepage and infiltration patterns can only be considered at regional scale, not at local scale [28].

The geological and hydrological complexity of Rotterdam displays a diversity of soil types, but is characterized by sandy soils, clay, and peat. The heterogeneity is a result of depositions of mudflat and lagoon sediments during the Holocene period. Also, erosion and deposition corresponding to ancient river systems and the influence of the North Sea. A landscape is created in which groundwater flow patterns are influenced by the spatial variability in soil type and the corresponding hydraulic properties. Presently, the Anthropocene layer consists of permeable sand and the thickness varies between centimeters to

approximately 30 meters at Maasvlakte, at the western side of the municipal area. Within the city, the process of soil consolidation takes place and in order to prevent further consolidation, supplementary sand is added. Through this process, the top layer is strongly affected by Anthropocene influences. Outside of the road cunet, the soil is more clayey and likely to be less permeable. The geohydrological dynamics, that are influenced by hydraulic properties of the soil types, affect the groundwater system and thus the strategies that are suitable regarding optimal groundwater management [25].

### 3.1.2. Ecohydrological system description

The interplay between ecology and hydrology can be defined as the effects an ecological system has on the groundwater system. For example: Evapotranspiration from shallow groundwater [14]. Ecological factors help to define spatial and temporal scales, which are smaller than geohydrological characterizations. Groundwater perceived ecological functions extend beyond encompassing stream flows, moderating water level fluctuations in lakes and wetlands reliant on groundwater resources, enrichment of ecosystems with nutrients, and hydration of riparian zones and groundwater-dependent vegetation.

The main objective for the description of ecohydrology within the municipal area of Rotterdam is in what way the ecohydrological system can be influencing the groundwater system. Important parameters are the overall structure of the study area (urban or rural), the characteristic ecosystems and gradients that are present, the amount of precipitation and evaporation that could influence the groundwater level in the study area. It has to be noted that specific local conditions, species composition, and community structure will not be discussed in this research study [28].

Commencing, the municipal area is located next to a rural region filled with grassland that comprises peat meadows and agricultural land. In the western part of the area, a broad coastal zone that consists of dunes, beaches, and forests in the inner dune area are present. The area is characterized by: 1) A transition of high lying peat meadows to reclaimed land; 2) A transition of coastal dunes to beach walls to peat meadows to reclaimed land; 3) A transition of river areas to peat meadows and reclaimed land. Overall, the characteristic ecosystem of the municipal area can be identified as river and brook valleys, coastal dunes, marshes, grass, and arable land in combination with urban development [28]. Groundwater and surface waters interact through ecohydrological processes. The groundwater level correlates with the ground level and influences the vegetation composition at local scale. Since the height of the groundwater level influences the oxygen concentration and extent of moisture that is available in the soil, surface waters interchange with nearby groundwater [53].

## 3.2. Hydrological cycle in urban areas

Urbanisation is a far-reaching process that has influence on the hydrological cycle in an area [5]. The hydrological cycle in urban areas is different than the hydrological cycle in rural areas. A key distinction in the cycle among urban and rural landscapes lies in surface sealing and the transport of precipitation through surface runoff. It is proven that the water balance transforms when urbanisation takes place. The effects of urbanisation on the hydrological cycle are: 1) Decrease of evaporation; 2) Decrease of percolation and capillarity, because of a loss in permeable surfaces [11]; 3) Decrease of the net groundwater recharge with optional processes as seepage; 4) Increase of discharge to surface waters, because of surface sealing and drainage measures; 5) Discharge to a WWTP.

A major element of the hydrological cycle is groundwater recharge as a result of percolation and capillarity. The consequence of urbanisation is the reduction of groundwater recharge, because of surface sealing. Typically, in urban environments, precipitation rarely contributes to groundwater recharge. This is true for heavy rainfall events, during which no time is available to infiltrate into the soil. Conversely, light rainfall events result often in groundwater recharge.

Urban areas have a dual hydrological system: A modified natural drainage system and an artificial water supply and wastewater disposal system [9]. Throughout City of Rotterdam, the modified natural drainage system includes canals and river diversions such as the "*Nieuwe Waterweg*".

### 3.3. Collection of groundwater level data

Measurements of groundwater levels are necessary for understanding subsurface dynamics [21]. This section discusses the practice of groundwater level measurements, types of monitoring wells, and the influence of groundwater flow dynamics on measurements.

Groundwater levels can be identified through monitoring wells, open boreholes, and field estimates. A monitoring well is the common name for a tube or comparable construction that consists of perforated, permeable, and impermeable sections in which a groundwater level or pressure head can be measured or where groundwater sampling can take place. Monitoring wells can be divided into groundwater wells and piezometers [38]. In The Netherlands, monitoring wells have to be placed according to the NEN standard. The standard is correlating with the requirements of Dutch law in order to ensure safety and quality of certain activities and products. The standard that applies for the placement of monitoring wells is NEN 5766[1]. According to the NEN standard, several requirements have to be followed regarding the placement of the monitoring well [1]. More specifically, the borehole has to be drilled underneath the groundwater table. Depending on the monitoring objective of the contractor, the diameter of the bore hole, the depth of the filter, and the length of the filter varies. Besides, determination of the length of the monitoring well has to be at least 1 meter. The length of the filter can be extended, depending on the type of subsoil and contamination in the subsoil [1]. A groundwater well has a relatively short filter length that varies between 0.5-1 meter. The bottom side of the well has a short distance underneath the groundwater table. Conversely, a piezometer can be described as a tube with a short perforated filter, where the pressure head can be measured. Groundwater wells can be distinguished in unconfined and confined wells. Unconfined wells are called phreatic wells. Unconfined aquifers are located underneath the water table in a permeable layer that is located on top of an impermeable layer. Confined aquifers can be described as an enclosed, water-bearing layer [46]. Figure 3.2 displays the difference between an unconfined and confined aquifer. The flow of water in the

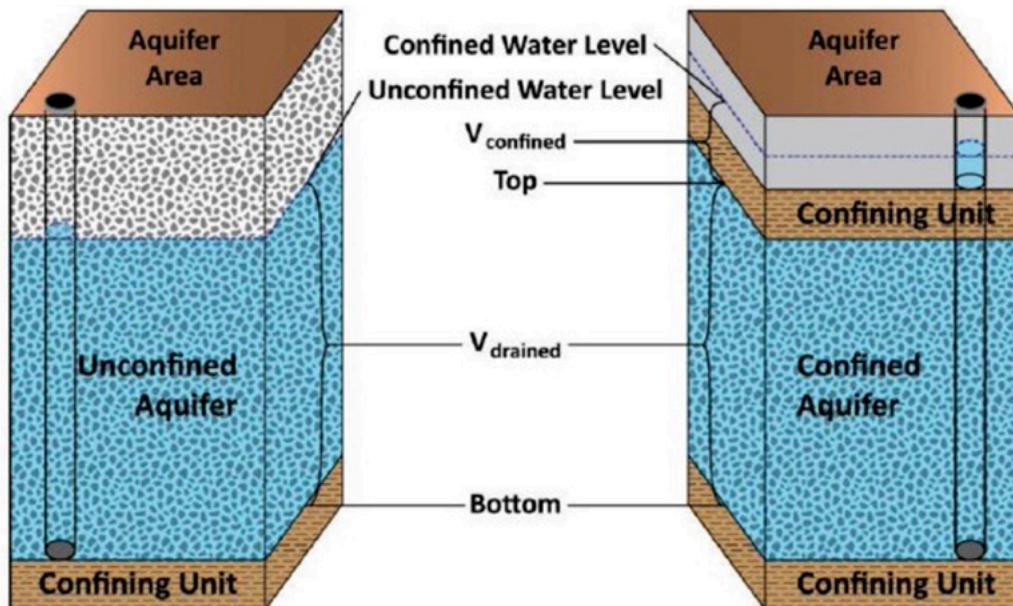
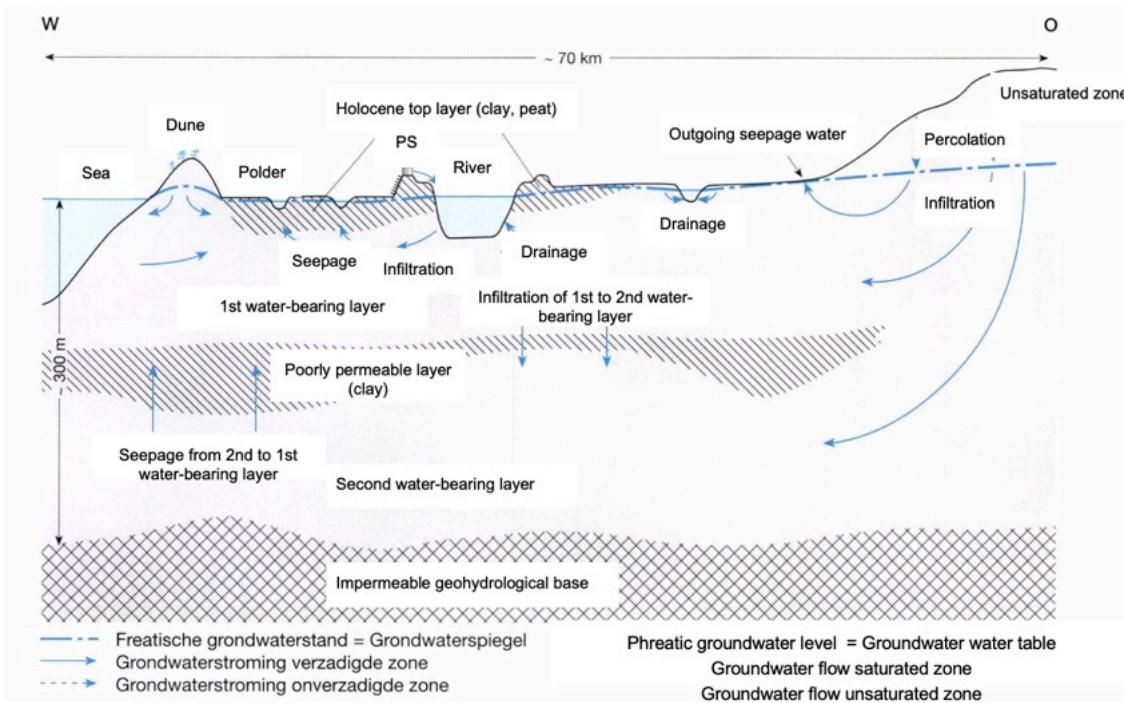


Figure 3.2: Schematic graph of an unconfined (left) and confined aquifer (right) [26].

subsoil is a 3D-spatial process. Spatial differences in ground level height, degree of permeability of the soil, hydrological environment, relative humidity, weather conditions and precipitation, and atmospheric pressure play a significant role in the potential distribution and, in the end, the flow patterns of groundwater [10]. The groundwater level in a monitoring well can be investigated by measuring the weighted average pressure head over the entire perforated length of the well. The extent of measuring depends on parameters such as the type of monitoring well, soil composition and the hydrological situation. The hydrological situation in this context can be defined as: 1) Hydrostatic equilibrium; 2) Non-hydrostatic

equilibrium with upward and downward seepage; 3) Stationary and non-stationary situation [38]. The hydrostatic equilibrium can be illustrated as no movement of groundwater in the subsoil. If a monitoring well is present in the subsoil, water in the well is in hydrostatic equilibrium. The water level in the well is at a constant height and there is no pressure head gradient. While a non-hydrostatic equilibrium can be defined as water movement in the subsoil with an additional distinction between downward and upward seepage. Generally, air pressure influences the fluctuation of the groundwater well. For the quantification of measuring data with a high frequency (< 24 hours). The abstraction of soil moisture through the roots of vegetation can also influence the daily fluctuation of the groundwater level [38]. Figure 3.3 illustrates the hydrological situation with among others the processes of infiltration and seepage to the first and second aquifer.



**Figure 3.3:** Hydrological situation and process in the western part of the municipal area of Rotterdam translated, based on Dufour [10].

### 3.4. Groundwater level monitoring

A GWMN includes a cluster of monitoring wells that are strategically placed to measure groundwater levels. The network is crucial for effective groundwater management and serves as a tool for groundwater management and, inadvertently, for integrated water management [52]. A GWMN acts as a tool for gathering information about a groundwater system. A monitoring well provides information on the state of the groundwater. Single measurements nor isolated monitoring wells provide additional value, because of the spatio-temporal deviations over time, an adequate monitoring plan is required. The objective of a GWMN and the additional interests depend on local themes, this can even differ across neighborhoods in the municipal area. If groundwater nuisance is occurring, it is needed to determine the phreatic groundwater level. A GWMN can provide a prediction of the groundwater dynamics in the area. Differences between calculations and monitoring can give an indication of the quality of the model [55]. It has to be noted that groundwater level monitoring with the use of monitoring wells is one of the essential parts of groundwater management. Yet, it is not possible to install monitoring wells at every corner of the street, but for effective groundwater management it is proven to be a prerequisite to explore the status of groundwater at 'forgotten' locations. To reduce the costs of (re)construction and installation of additional monitoring wells, it is key to identify the most appropriate locations [31].

A potential approach for identifying a representative monitoring location in an existing network involves the computation of the parameters average groundwater level, system inertia, recharge sensitivity, which describe the groundwater dynamics together, and through field assessments [31]. The existing monitoring wells are examined to determine if the well qualifies as representative. In the study by Massop and van der Gaast, the objective is not to find the most suitable monitoring location, but to assess the existing GWMN based on the representativeness and the potential for remediation and expansion [31]. The result of the assessment is a set of locations that is representative for the distribution of properties that occurs within the entire research area as well as distinguished clusters to achieve a representative reflection of the variation in groundwater level [m].

### 3.5. Groundwater management and policies

Groundwater management is an integrated cluster of water system management. The government, provinces, water boards and municipalities execute tasks within groundwater management. Regulating groundwater management can be divided into quantitative and qualitative groundwater management. Both quantitative and qualitative management include strategic and operational management, which comprises tasks as policy planning, framework construction, granting and enforcing permits [39]. Since 2009, the National Water Act ensures a municipal obligation of care regarding groundwater management. Although the municipality of Rotterdam is not liable for groundwater problems, they are accountable [16]. Important objectives in the context of groundwater policies are included in the "*Gemeentelijk Rioleringsplan*" (Municipal Sewage Plan) [17]. The Municipal Sewage Plan ensures an overview of developments that occur now and possibly can in the future within the urban water system of Rotterdam. The system consists of a sewage system, public area, and surface water system. The urban water system gathers, transports, processes, and discharges domestic wastewater, groundwater and precipitation. Within the sewage plan, there is a rough overview of what the urban system looks like, which developments are going to take place in the future, and which changes are likely to occur. Nevertheless, it is crucial to note that City of Rotterdam is not able to implement such policies by itself. Therefore, the Municipal Sewage Plan consists of three clusters: 1) The urban water system; 2) Strategy; 3) Program of the Municipal Sewage Plan of Rotterdam. The sewage plan has a valid time period of 5 years and comprises the period of 2021-2025. Within this period, the "*Omgevingswet*" (National Environmental Act) was also initiated. With the start of the National Environmental Act, the Municipal Sewage Plan is not mandatory for the city anymore. Nevertheless, they choose to extend the program. Partly because of the desires of the city and its residents, but also because of the costs that are connected with the urban water system. The Municipal Sewage Plan provides insight into the expenses and justifies the municipal sewage fees [17], as shown in figure 3.4. The tasks within the green boxes fall under the municipality's jurisdiction, while responsibilities like water purification, drinking water provision, and surface water management are handled by Rijkswaterstaat, water boards, and drinking water companies, respectively.

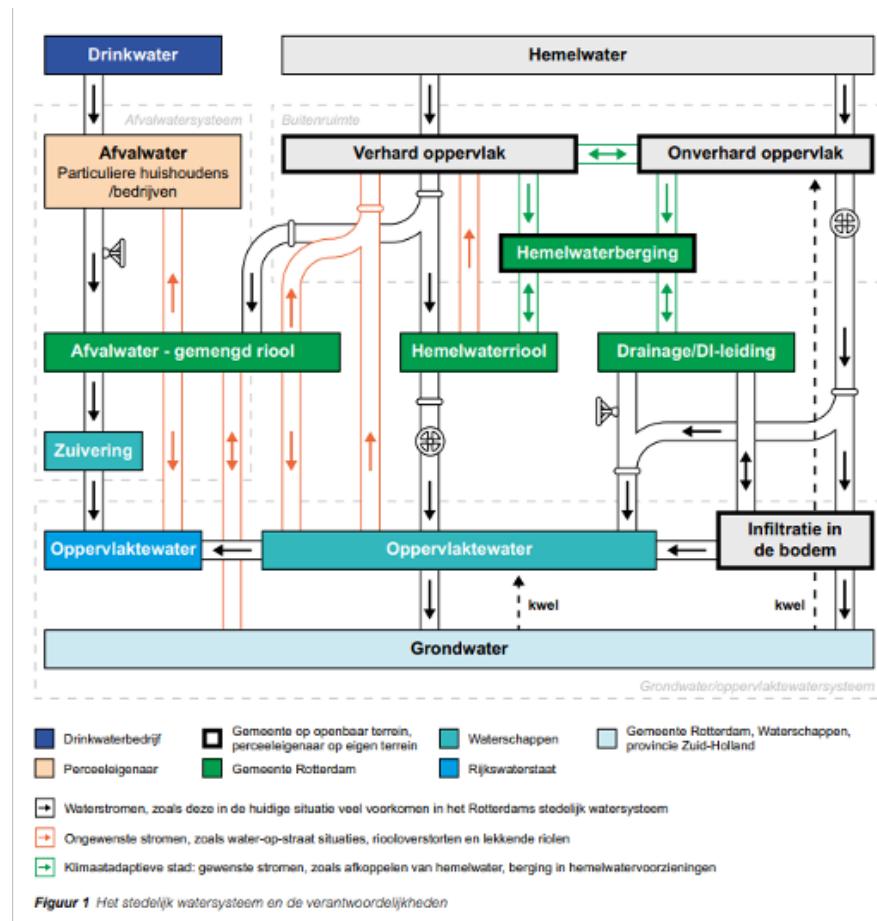


Figure 3.4: Overview of the urban water system according to the Municipal Sewage Plan of Rotterdam [18].

# 4

## Research Methodology

### 4.1. Research design

The research design outlines the framework guiding the research, aligned with the objectives. The study aims at developing a generic model on neighborhood scale and intends to assess the optimization level within Rotterdam's GWMN. The methodology is structured into a dual-part research design:

#### 4.1.1. Descriptive design

The initial phase aims to characterize the present state of the system, utilizing hydrogeological and geographical data of the research areas. The system description involves analyzing the distribution and coverage of monitoring wells within the municipality without manipulating variables. The descriptive design develops a detailed visualization of the existing GWMN, including parameters as distribution, coverage density, and groundwater level measurements.

#### 4.1.2. Quasi-experimental design

The second phase includes an approach that manipulates the GWMN by simulating data and removing monitoring wells. These actions have the purpose to observe changes regarding network optimization. Initially, the research study includes 14 city districts and 4 industrial areas, however, a generic model is set-up to assess the optimization on neighborhood scale. Only two specific research areas are included in the chapters 4 and 6. The choice of the two specific research areas is based on factors such as data collection type, monitoring frequency, data availability, and density of the coverage area. Data loggers ensure that inaccessible and vulnerable locations within the city districts still can be frequently monitored through this approach. Certain monitoring wells are on inaccessible locations within the municipal area, however, through the use of data loggers these locations can still be actively monitored. The design continues with a quantitative method to analyze the impact of changes regarding the reduction-optimization method over a longitudinal period.

### 4.2. Data collection and preparation

Within the research study, the time period includes 01-01-2020 to 01-01-2024. From 2010 to 2014, the groundwater data was collected by City of Rotterdam. In the period of 2014-2018, however, the field measurements were carried out by an external contracting company. The field measurements take place at the end of every month, executing a measurement every month. Since 2019, groundwater data is again collected by City of Rotterdam, meaning that multiple monitoring wells throughout the municipal area are transferred to data loggers instead of the manual collection method. Because of this transition, inaccessible locations were still monitored and data is collected more frequently; every 4 hours instead of every month.

#### 4.2.1. QGIS and PROWAT

Starting with the first step of the data collection, the open-source QGIS and Sensor Management application of City of Rotterdam allows displaying monitoring data for all unique monitoring wells within the

municipal area. In QGIS, study areas of preference can be selected and transferred to the PROWAT system. PROWAT allows the data to be converted to an Excel data file. After which, data can be imported in the programming language Python.

#### 4.2.2. Pastas time series modeling

The quasi-experimental research design starts with Pastas, a Python package made to analyze a sequence of data points that are collected over specific interval of time [7]. In the original dataset, data gaps are present, resulting in potential problems later on in the development of the model. At first, a basic model is generated to visualize a simple time series model to simulate groundwater levels in the research areas.

##### Dependent time series

A time series of groundwater levels is imported through the Pastas package. The dependent time series data describe the observed time series monitored by City of Rotterdam. The results of the dependent time series describe the groundwater level [m NAP] over a period of 01-01-2010 to 01-01-2024. Unique monitoring wells of the neighborhoods Rozenburg and Heijplaat are visualized. The purpose is to depict figures that visually represent seasonal changes and variabilities as well.

##### Independent time series

The process includes incorporating an independent time series from an external database. The integration is being done for both precipitation and potential evaporation data. By utilizing both datasets, it is possible to estimate the recharge to the groundwater [m/day]. An additional stress model can be created through the calculation of the evaporation factor, the actual evaporation as a factor. The recharge is created with a linear recharge model, a combination of precipitation and evaporation series and so a parameter for the evaporation factor. The data is sourced from KNMI at weather station 344 in Rotterdam, The Netherlands. The depicted figures illustrate the recharge rate [m/day] spanning from 01-01-2010 to 01-01-2024, as well as the individual time series for precipitation [m/day] and potential evaporation [m/day].

##### Time series model

Once the dependent time series (observed GWL data) and independent time series (precipitation and potential evaporation data) are developed and plotted, the actual time series model can be created. The dependent and independent time series are imported to the actual time series model. The imported time series are verified for data inconsistencies and missing values. The model is improved through assessing optimal model parameters. The solution of the calculation can be plotted in a time series plot, which describes the observed data as a scatter plot and the simulation data, according to the independent data in a line plot. The actual time series model includes "Stress Model 1". The depicted data translates the groundwater level [m NAP] over a time period of 01-01-2010 to 01-01-2024. Besides the time series model, it is possible to accommodate an overview of the model. In the overview plot, information regarding the residuals, noise, recharge, and stresses are displayed. A full description of "Stress Model 1" is stated in Appendix A.

##### Statistical substantiation

In the end, a summary of the developed performance metrics is provided. The performance of the models that are created previously, is visualized in a bar plot with performance metrics of RMSE,  $R^2$ , and EVP. Nevertheless, based on statistical data only, it can not be provided if a model performs good or bad. Therefore, an additional statistical test is executed. The Welch's t-Test compares data groups: data logger and manual collection method without assuming equal data variances [2]. The t-Test is carried out to criticize whether a dataset with a combination of measurement types: data logger, manual collection or a combination of these two is recommended to be used further in the research study. A Welch's t-Test is a suitable approach, because the denominator provides the possibility that two data groups occur to have unequal variances. The t-Test is determined by the mean of the two data groups, the variances of both data groups, and the sample sizes of the two data groups. Furthermore, the Welch's t-Test suits, since a difference occurs between the data quantity of the data loggers and manual collection method. Nonetheless, in the context of environmental sciences, a perspective beyond statistics is key: Other factors such as the reliability and quality of the data and preference of stakeholders can shape the final decision regarding the overall substantiation of the data group that might

be included in the research study. In summary, adopting a comprehensive approach that goes beyond statistical significance is crucial. The decision has been taken to proceed with the research using the data logger group. The observed and simulated data logger measurements are combined into a new data frame, with which the research study will be proceeded with.

### 4.3. QR factorization

A data-driven algorithm, based on the Python package PySensors, is designed to optimize groundwater monitoring networks by determining the optimal number and location of monitoring wells for temporal and spatial level reconstruction. The algorithm can be explained by the dependency of data in order to make a certain decision, prediction or optimization process. Existing data from City of Rotterdam is used to study patterns, trends, and relationships between parameters. The method was first obtained by Ohmer at the KIT and applied the methodology to a study area in Germany [34]. The approach is included in the optimization approach regarding City of Rotterdam. The network reduction-optimization method is utilized on the GWMN, using one-dimensional hydrograph data. The main process of the algorithm is the use of QR factorization, a mathematical approach that breaks down matrix  $A$  into the product of two matrices:  $Q$  and  $R$ .  $Q$  is an orthogonal matrix that transforms the original dataset into a new dataset with independent columns.  $Q$  is essential for evaluating the unique contribution of each monitoring well.  $R$  is a triangular matrix that ranks the monitoring wells based on their informational relevance. In the algorithm, matrix  $A$  encompasses the dataset that is created through Pastas time series modeling. In the dataset, every row corresponds to a different observation and every column represents a unique monitoring well. Appendix B describes a more in-depth mathematical background of the QR factorization approach.

The optimization approach enables the network to be reduced along the Pareto front, balancing information loss against cost savings. To achieve the most accurate reconstruction with the least number of monitoring wells, the algorithm tests various reduction scenarios (10-25-50-75-90%) to identify the lowest reconstruction error for the monitoring network and so the optimal number of monitoring wells. The number of monitoring wells is indicated as a so-called threshold value, explaining the trade off between the maximum information value of the monitoring wells and the minimum costs, in which the maximum information content is characterized by the reconstruction error and the minimum cost is equivalent to the number of monitoring wells [54]. It can, however, be possible that the optimal reduction percentage differs between the study areas. A reduction percentage of at least 25% should be applied to describe the system dynamics in the study areas, while a reduction percentage exceeding 75 percent would influence the accuracy of the reduction results [34].

In Github, a detailed script regarding the QR factorization method in the model is described. The Python package "PySensors" contains a data-driven genetic algorithm for sparse sensor selection and signal reconstruction with dimensional reduction, see appendix B.

### 4.4. Data analysis

A quasi-experimental design is employed to critically examine, analysis, and compare data, derived from data loggers and the manual collection method by City of Rotterdam. Firstly, the dataset undergoes a preprocessing phase in which a division is made into a training and test dataset, designed to test the predictiveness and performance of the model. Continuing, a QR factorization approach is applied and offers insight into the impact of different data collection strategies on model performance. Based on the informational relevance of the monitoring wells within the study area, a hierarchical list can be generated. The QR factorization approach applies a quasi-experimental evaluation of the impact of well reduction on the overall model performance. A systematic reduction of the network enables to assess the correlation between groundwater monitoring network reduction and the accuracy of the model. Assessing the reduction scenarios ensures to gain understanding on how different data collection methods affect the robustness and accuracy of groundwater level predictions. After a reduction scenario is chosen, the scenario eliminates groundwater monitoring wells with less informational relevance from the network. The eliminated monitoring wells are visualized in hydrographs and performance metrics, in order to substantiate and compare the performance of the observed and reconstructed data.

# 5

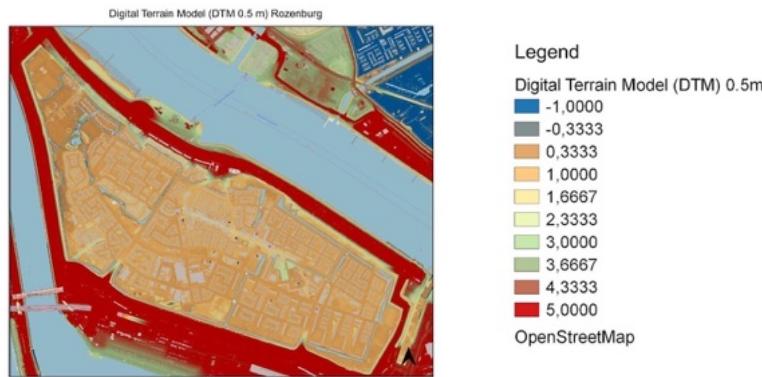
## Case Study

The substantiation to focus on Rozenburg and Heijplaat as case study areas, stems from the data availability and collection methods that are employed in these areas. With a dense network of data loggers and a comprehensive historical database, the study areas enable a thorough examination of the distribution of monitoring wells and the variation in groundwater levels across.

### 5.1. Rozenburg

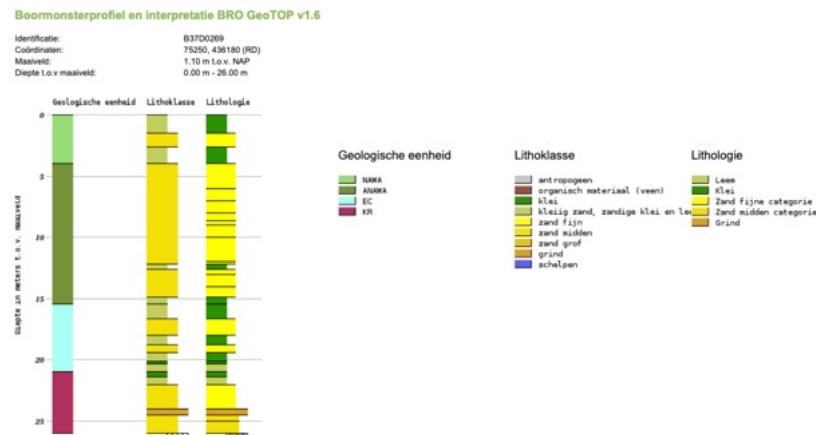
The headland Rozenburg is notable for its eponymous village. Part of the headland has undergone extensive excavation measures and now serves the purpose of an industrial landscape. Prior to 2010, the village maintained its autonomy as Gemeente Rozenburg, but eventually merged into the larger City of Rotterdam. Positioned on the headland, the village finds itself in the inner dike area, enclosed by the industrial zones of *Botlek* and *Europoort*. The worlds of port and industrial activities as well as nature coexist. Therefore, part of the headland is called the “*Groene Gordel*” [36]. In the 1950s, the headland primarily functioned as an agricultural area with a dune and nature reserve “*de Grote Beer*” on the western side. However, in the 1960s, major excavation measures took place to accommodate the development of *Maasvlakte* and *Europoort* industrial areas [8]. Given its location, Rozenburg faces a notable risk for flood events. The headland has an average ground level of +5.7 meters NAP [51], while the neighborhood itself has a much lower ground level, see figure 5.1. The strategic elevation differences reduces the village’s vulnerability to potential flooding, particularly from the canal “*Nieuwe Waterweg*” on the northern side of the district. Figure 5.1 shows an elevation model of Rozenburg, based on a digital terrain model with a resolution of 0.5 meters. As can be seen from the figure, the neighborhood Rozenburg is located in the orange marked area with a dike ring in the red marked areas.

Regarding the geology of the area, the first 22.91 meters are a Holocene formation that consists of sandy clay, medium-fine sand, clay, peat, and coarse sand, see figure 5.2. The presence of these sediments suggest a varied deposition environment; fluvial and marine processes. The diverse grain sizes of the sediment and the presence of an organic peat layer could indicate that infiltration rates can be varying and thus affect groundwater recharge and storage. Underneath the Holocene layer, a layer of about 1.46 meters of the formation of Kreftenheye is present, see figure 5.2. The formation consists of medium-coarse sand, little sandy clay, fine sand and cobbles with traces of clay and peat. The composition of the second formation indicates a more permeable layer, in comparison with the top Holocene formation. This can also be explained by larger grain sizes, which allows for groundwater movement and possible occurrence of an aquifer [25].



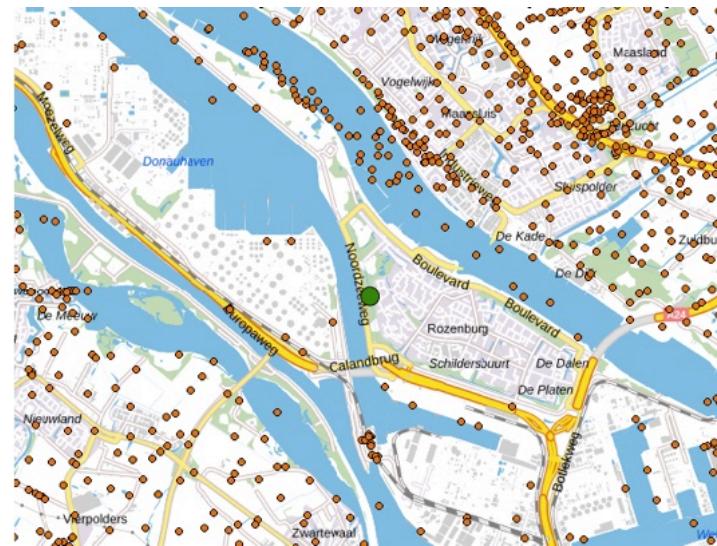
**Figure 5.1:** Digital terrain model (DTM 0.5 m) for Rozenburg. Areas marked in red indicate locations with the highest elevation, while orange marked areas indicate locations with the lowest elevation.

Collectively, the upper Holocene formation introduces a varied texture, ensuring heterogeneous permeability that affects groundwater flow and storage. The heterogeneity arose from the varied sediment composition, impacting the infiltration rate and groundwater dynamics. On the other hand, the Kreftenhaye formation characterizes itself as a permeable formation, which is the result of the medium-coarse sand and cobbles. Given the focus of the research study on phreatic monitoring wells, only the Anthropocene and Holocene formation are considered. However, for this specific drilling location, no information regarding the Anthropocene formation is available through BRO GeoTOP v1.6. This can also be seen in figure 5.2, the drill site profile called "*Lithoklasse*" in the middle of the figure explains a combination of clayey-sand, sandy-clay and loam in the first 20 meters of the profile.



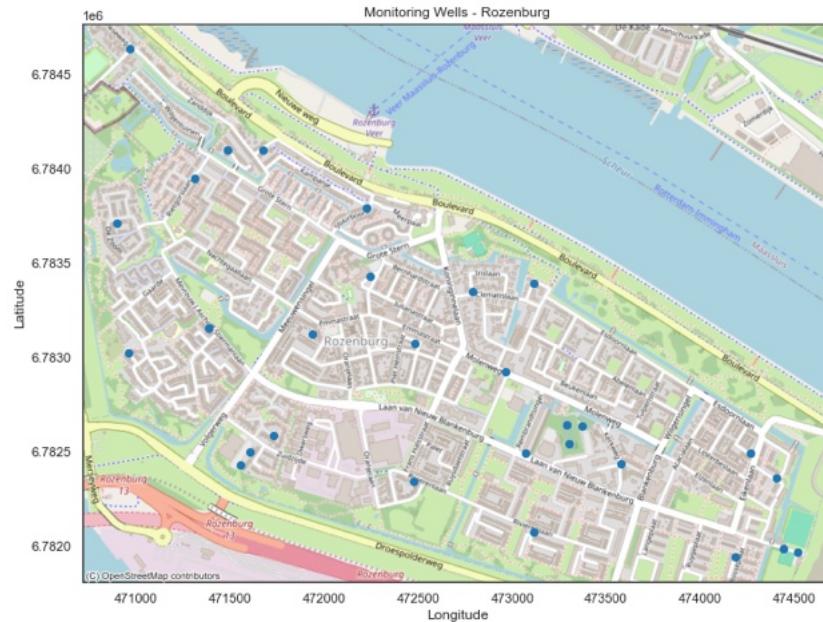
**Figure 5.2:** Profile of drill site profile with geological formations for the location x: 75250; y: 436180 [47].

The location of the exact drill profile is shown in figure 5.3. As can be seen in the figure, the location is in the southwestern side of the neighborhood, close to the connecting road. Only one drilling profile is considered, since this is the only drill profile that is present within the boundaries of the neighborhood.



**Figure 5.3:** Precise location for the drilling profile by the coordinates x: 75250, y: 436180 [47].

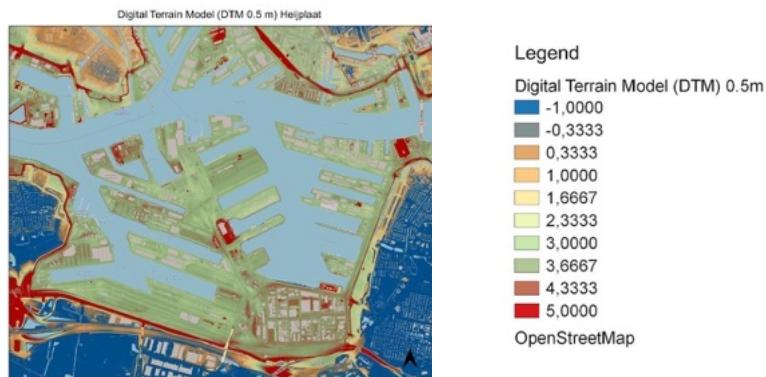
As stated by CBS, the total surface area of Rozenburg counts up to 411 hectares in total, of which 324 hectares is made of land. Rozenburg has 29 active, phreatic monitoring wells, meaning one monitoring well/11.17 hectares. In figure 6.1, a geographical representation of the neighborhood Rozenburg is displayed. Points of interest are the blue marked points, which indicate the available phreatic monitoring wells in the neighborhood.



**Figure 5.4:** Groundwater monitoring network in Rozenburg. Based on EPSG:28992/Amersfoort RD New.

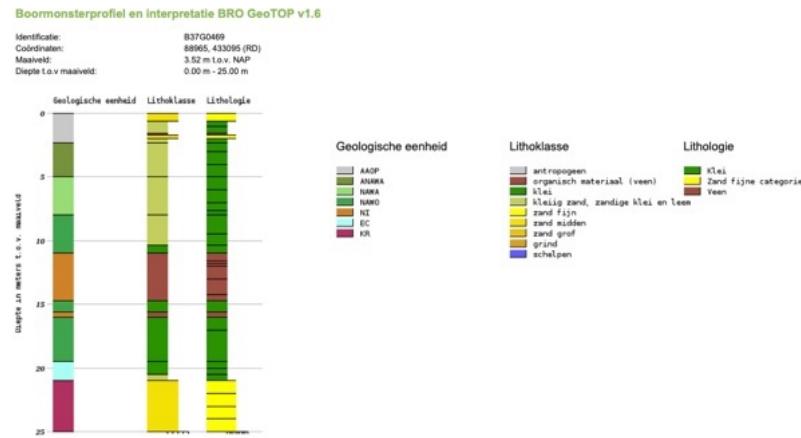
## 5.2. Heijplaat

Since the mid-15th century, the hamlet "de Heij" has been located at the northern side of the neighborhood Charlois. At the north of the hamlet, the creek "de Koedood" flew into the Meuse. However, years later, an industrial area called "de Heijsehaven" was constructed at that site, where an artificial sand formation was formed as well. Only at the beginning of the 20th century, the idea of the garden village arose [40]. Throughout the years, the village became more isolated because of the extension of the industrial areas [37]. Figure 5.5 shows that Heijplaat is still surrounded by industrial areas and water outside the dike ring. Access to Heijplaat is solely possible through one entrance road, the "Waalhavenweg" [32]. Figure 5.5 describes a digital terrain model with a resolution of 0.5 meters for the neighborhood. From the figure it can be seen that the neighborhood is located on a constant level of about 2-3 meters NAP.



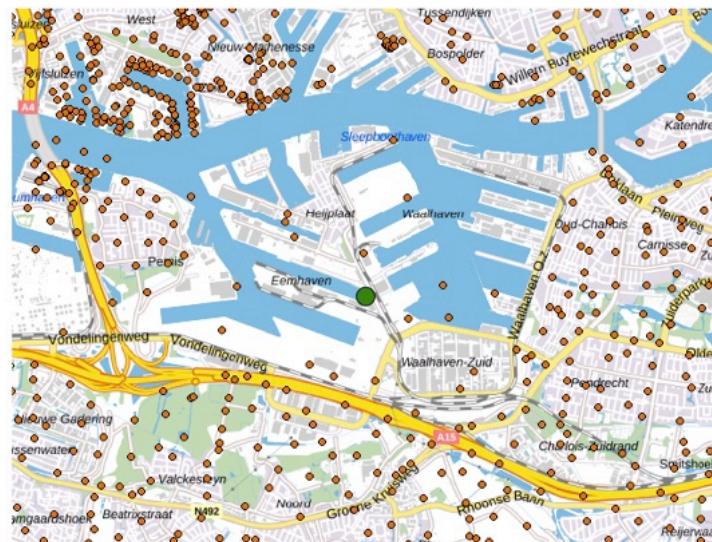
**Figure 5.5:** Digital terrain model (DTM 0.5) for Heijplaat. Areas marked red indicate the locations with the highest elevation, while blue marked areas indicate locations with the lowest elevation.

Figure 5.6 shows the lithology of Heijplaat. The subsurface in the eastern part consists of deposits from the formations Dunkirk III, Dunkirk I, and peat. A stratified geology is suggested with peat deposits present in areas with a high organic content. The peat deposits could influence water retention and filtration properties. On the western part of Heijplaat, however, the subsurface is characterized by the Dunkirk III formation atop a peat layer. The under-layer of peat indicates wet conditions in the past, referring to the relocation of the Riederwaard dike, which could influence the groundwater dynamics with layers of varying hydraulic properties [15]. In the period of the 10-12th century, the peat landscape has been drained which altered the landscape significantly. This showed itself through ground level reductions, necessitating dike construction. The manipulation of the landscape impacted and reshaped the groundwater recharge, flow and storage ability in the area. The upper Anthropocene layer is characterized by sand, fine to coarse clay, silty to sandy humic, and a layer of debris of about 2.3 meters thick in total, see the gray unit in the first site profile of figure 5.6 [47]. Overall, the area is marked by historical land alterations and complex, stratified geology, which influences the groundwater recharge rate, flow, and storage capacity in the neighborhood of Heijplaat. This specific drilling location is chosen, because it lies in the south of the area and is not located on the eastern side of the railway that crosses the neighborhood. The drill site profile called "Lithoklasse" in the middle of the figure explains a combination of fine-sand, clayey-sand, sandy-clay, and loam.



**Figure 5.6:** Profile of drill site profile with geological formations for the location x: 88529; y: 433870 [47].

The location of the exact drill profile is shown in figure 5.7. As can be seen in the figure, the location is on the southern side of the neighborhood.



**Figure 5.7:** Precise location for the drilling profile by the coordinates x: 88529, y: 433870 [47].

The total surface area of Heijplaat counts up to 39 hectares in total, of which 39 hectares are also made of land [6]. In Heijplaat, one monitoring well is present per one well/2.78 hectares according to [6]. Figure 5.8 visualizes a geographical representation of the neighborhood Heijplaat, part of Charlois. Points of interest are the blue marked points. They indicate the available phreatic monitoring wells in Heijplaat.



**Figure 5.8:** Groundwater monitoring network in Heijplaat. Based on EPSG:28992/Amersfoort RD New.

# 6

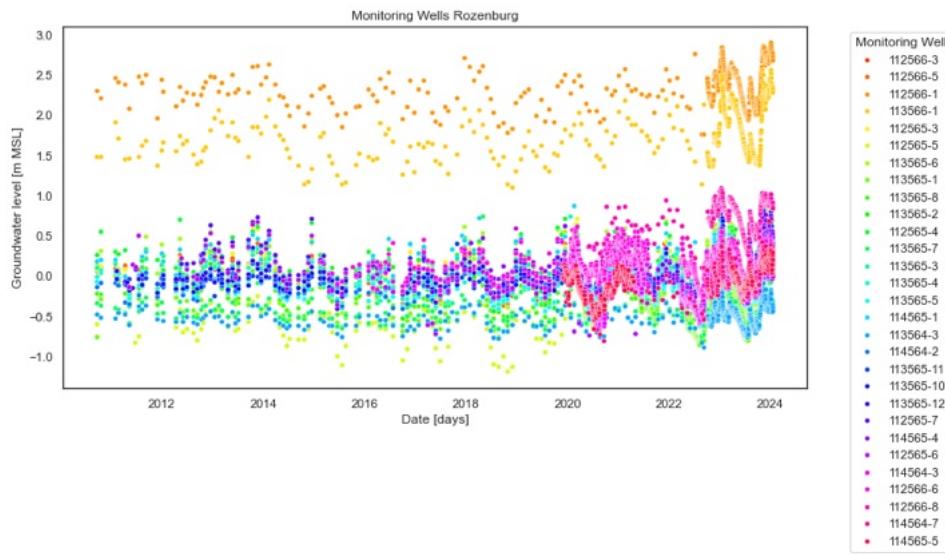
## Results

### 6.1. Rozenburg

#### 6.1.1. QGIS and PROWAT

Figure 6.1 represents a multi-colored scatter plot, showing the groundwater level [m NAP] over time [days]. Each colored dot represents an individual data point from a unique monitoring well, see legend on the right side of the figure. A wide array of groundwater levels is depicted, ranging from -1.0 to +3.0 m NAP.

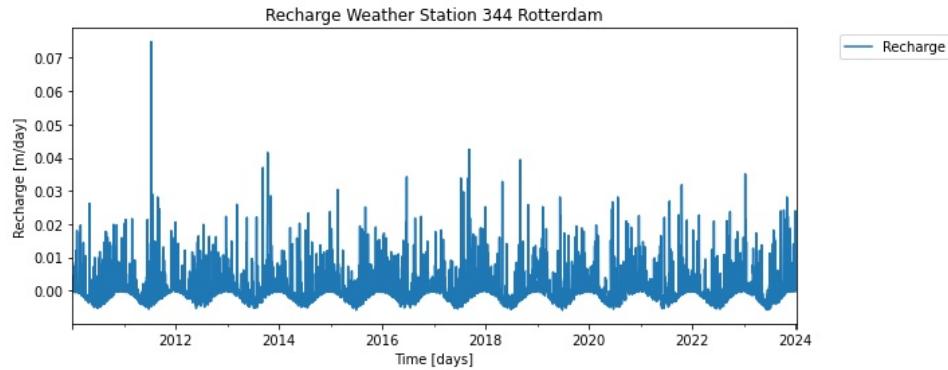
Plot 6.1 visualizes the variation in groundwater levels between different monitoring wells. The legend explains a range of colors for the unique monitoring wells, that counts to 29 in the study area Rozenburg. The function of the groundwater level is indicated by the colored dot for every monitoring well. When no dot is present, it means that no data is available for this specific measurement date. Several monitoring wells perform constant groundwater levels, while other monitoring wells have abundant fluctuations in their function. Dense clustering of data points might suggest that measurements were taken frequently over time. The variety of colors allows for the possibility to compare groundwater level fluctuations between monitoring locations within the neighborhood.



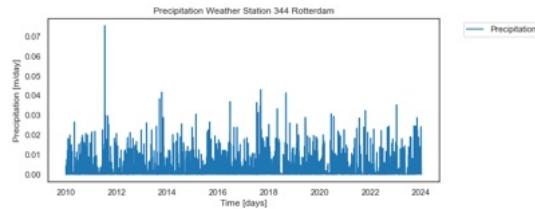
**Figure 6.1:** Observed data for 29 monitoring wells in Rozenburg.

### 6.1.2. Pastas time series modeling

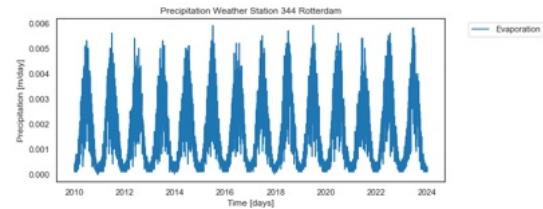
The bar plot displays the recharge rates [m/day] over a period from 2010-2024. The x-axis represents the time period from 2010-2024 with the dates labeled on the x-axis. The y-axis quantifies the recharge rate with values ranging from 0 to a maximum of 0.07 m/day. The bars shows the variability in recharge rates over time, where some years experience peaks in recharge (higher recharge rates), which can be due to increased precipitation. As can be seen from 6.2, the data is dense, indicating frequent measurements by KNMI. The pattern of the bars reflects seasonal changes. Lower values are more frequently observed, possibly indicating a baseline level of recharge over the years. Two additional figures (6.34 and 6.35) of precipitation and potential evaporation data are visualized as well. The data of these figures is combined to determine the recharge rate in the municipal area.



**Figure 6.2:** Recharge [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands.



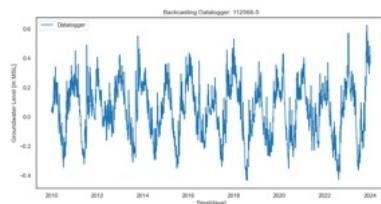
**Figure 6.3:** Precipitation [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands.



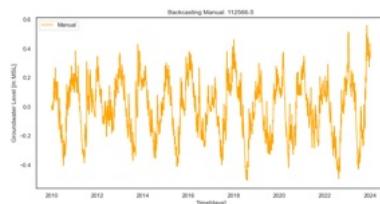
**Figure 6.4:** Potential evaporation [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands.

### Reversed forecasting

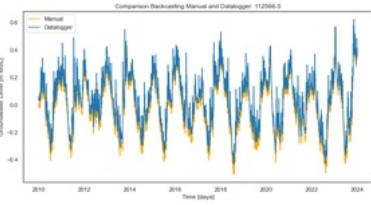
Reversed forecasting, or also known as backcasting, of the observed GWL data is a feature in the Pastas Python package. The process is applied for every unique monitoring well within the case study area. For every well, three types of visualizations are generated: 1) Backcasting for data logger data; 2) Backcasting for the manually collected data; 3) A combination figure of figures 1 and 2 combined, see figures 6.5-6.7. Monitoring well 112566-5 shows an example of the three visualized figures. The following applies for all three figures: The x-axis includes a period of 2010-2024 and the y-axis shows the simulated GWL data [m NAP].



**Figure 6.5:** Reversed forecasting data based on datalogger data for monitoring well 112566-5.



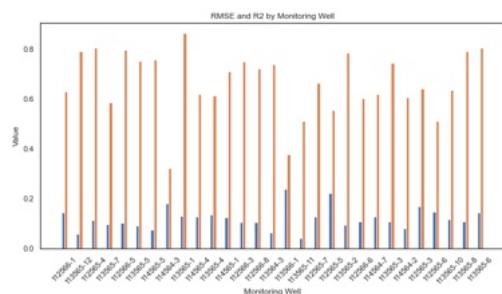
**Figure 6.6:** Reversed forecasting data based on manual measurement data for monitoring well 112566-5.



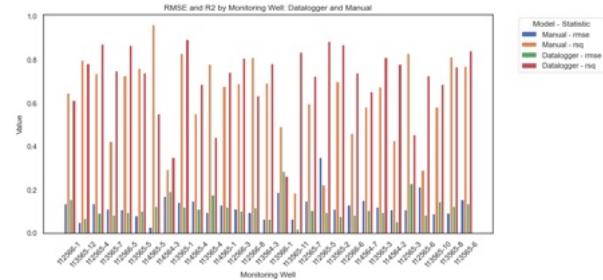
**Figure 6.7:** Reversed forecasting data based on manual measurement data and datalogger data for monitoring well 112566-5.

### Performance metrics

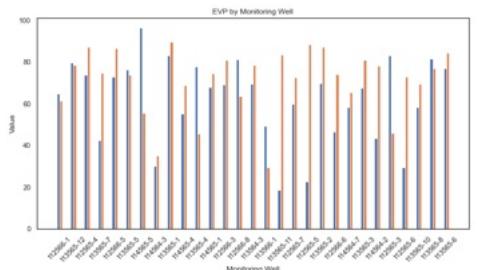
The bar plot 6.8 and 6.9 shows the two metrics RMSE and  $R^2$  for all unique monitoring wells. In figure 6.8, a first overview is given of the t-Test with a distinction between RMSE and  $R^2$  and an additional distinction between data loggers and manually collected wells. In figure 6.8, the  $R^2$  value (orange bar) of all monitoring wells seems higher than the values of the RMSE (blue bar) of the same monitoring wells. Figure 6.9 encompasses four categories: 1) manual-rmse; 2) manual-rsq; 3) datalogger-rmse; 4) datalogger-rsq. From this figure it is not clear if one of the datagroups has a better model performance, the performance of every model statistic differs between the monitoring wells. An additional bar plot, figure 6.10 describes the EVP [%] across the dataset. The EVP compares the variance of the observed data and the variance of the residual data across the 29 monitoring wells [3]. It is not clear how the values relate to each other, because there is a distinction between the datalogger and manual collection method. A low EVP might indicate that data could be missing in the dataset, the spatial pattern could be a possible reason. A statistical t-Test is a follow-up to identify and statistically substantiate the difference in performance between the datagroups. Based on the results of the barplot, the decision was made to execute further statistical tests. A Welch's t-Test was carried out, leading to the following results. The Welch's t-Test uses an  $\alpha = 0.05$ .



**Figure 6.8:** Bar plot showing the metrics RMSE and  $R^2$  for 29 monitoring wells.



**Figure 6.9:** Bar plot showing metrics RMSE and  $R^2$  for 29 monitoring wells with a division in data logger and manual collection method.



**Figure 6.10:** Bar plot showing metric EVP for 29 monitoring wells.

Figure 6.11 displays the result of the Welch's t-Test, based on  $\alpha = 0.05$ . The results of the performance metrics are labeled as RMSE,  $R^2$ , and EVP. Each performance metric provides the results of the t-statistic and p-value of the Welch's statistical test. The RMSE (blue section), explains that the t-statistic is negative. This result could indicate that the data loggers have a lower mean RMSE than the manual collection method. The p-value is higher than  $\alpha = 0.05$ . A p-value higher than the set alpha explains that no significant difference is present between the data loggers and manual collection method. The  $R^2$  (orange section) explains that the t-statistic is positive, indicating a higher mean  $R^2$  for the data loggers than the manual collection method. The p-value is slightly higher than  $\alpha = 0.05$ , indicating no significant difference. The last metric is the EVP (green section). The t-statistic of the EVP has a positive value, suggesting that the data loggers might have a higher mean EVP than the manual collection method. The p-value of 0.065 is very close to 0.05, but resulting in no significant difference.

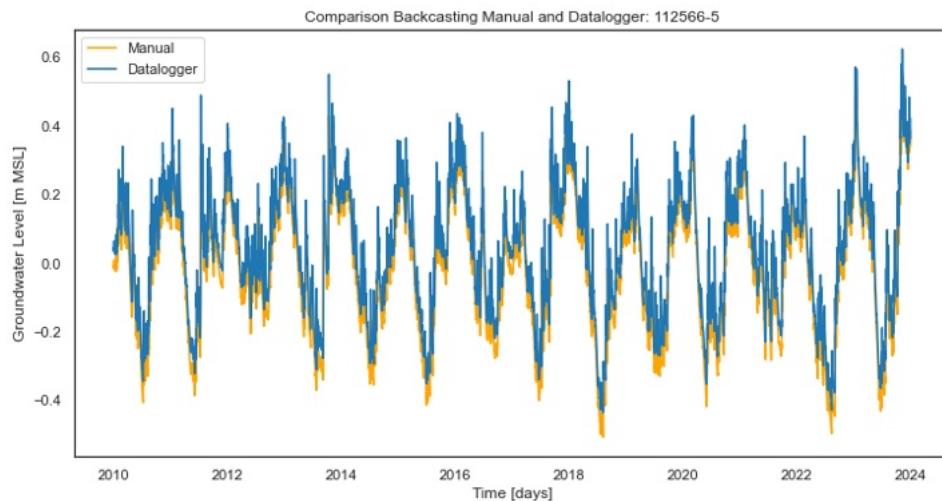
RMSE	T-stat: -0.6685651168754769 P-value: 0.5065504513682721
There is no significant difference between Datalogger and Manual for the rmse statistic (p >= 0.05).	
R2	T-stat: 1.8453668641964593 P-value: 0.07050974119938155
There is no significant difference between Datalogger and Manual for the rsq statistic (p >= 0.05).	
EVP	T-stat: 1.8820241005567129 P-value: 0.06530584247695573
There is no significant difference between Datalogger and Manual for the evp statistic (p >= 0.05).	

**Figure 6.11:** Overview of the performance of RMSE,  $R^2$ , and EVP after the Welch's t-test and the difference between data loggers and manual collection method.

The Welch's t-Test does not reveal no substantial discrepancies between the data groups. None of the p-values are below the threshold value of 0.05. Despite statistical significance, factors as data availability and reliability are decisive. Hence, the data logger group is chosen, merging observed and simulated data into one dataset and excluding the manual measurement from the research.

#### Creation of a new data frame

Scatter plot 6.12 below, visualizes the observed data logger data with the blue scatter and the simulated data of the data logger is visualized by the orange scatter. On the x-axis the time period [days] is shown and on the y-axis the groundwater level [m NAP]. Figure 6.12 visualizes only monitoring well 112566-5, the remaining monitoring wells of the neighborhood Rozenburg are available through Github.



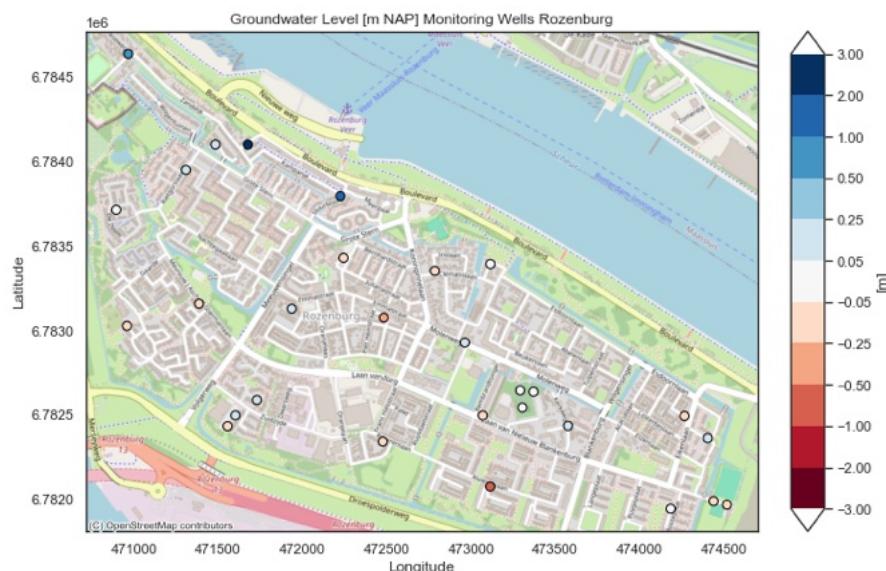
**Figure 6.12:** Scatter plot of simulated and observed data by data loggers for monitoring well 112566-5. The x-axis shows a period of 2020-2024 and the y-axis shows a range of -0.5 to 0.6 meters NAP.

### 6.1.3. QR factorization

Using QR factorization as a foundation, a hierarchical list of the monitoring wells in Rozenburg can be created. The optimal reduction percentage is determined and reduction tests can be executed, resulting in an overview of eliminated monitoring wells and their capability to reconstruct future groundwater levels in the network.

#### 1D hydrograph data

Figure 6.13 is a geographical representation visualizing the spatial distribution of monitoring wells in the neighborhood Rozenburg. Within the map, a scatter plot with a color scale represents the groundwater level measurements taken at the monitoring wells. The groundwater level has a range between -1.00 m to +3.00 m NAP. Each monitoring well represents a location as well as the color of monitoring well represents the value on the color scale. At the north side of the neighborhood, deviating groundwater level measurements can occur because of the elevation difference of the ground level, a digital terrain model of Rozenburg is displayed in figure 5.1.



**Figure 6.13:** Groundwater level [m NAP] across Rozenburg, using a color-coded system to represent the mean GWL observations for the 29 monitoring wells. Based on EPSG:28992/Amersfoort RD New.

### Sampling data

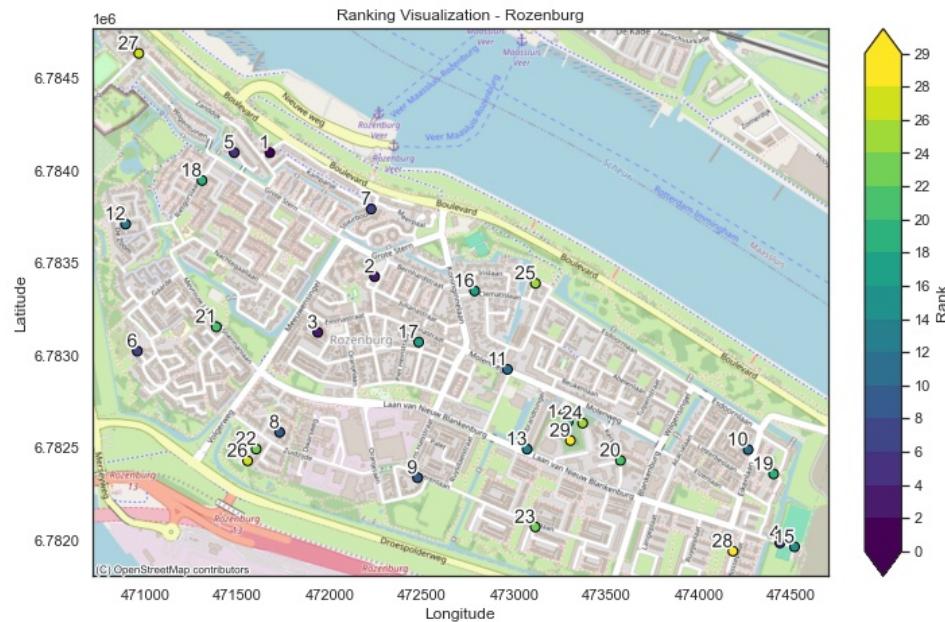
Preprocessing of the data includes transforming the data frame into an array, implementing both global and local centering methods, and splitting the dataset into an 80/20 ratio training and test set distribution. The result of the preprocessing is as follows, see table 6.1

**Table 6.1:** Summary of sampling data.

Parameter	Value
Sampling period	2020-01-02 00:00:00 to 2024-01-01 00:00:00
Number of samples	1462
Number of features (sensors)	29
Shape of X (samples, sensors)	(1462, 29)
Min. and max. value	-0.8794576895779862 [m] ; 2.9179266529665524 [m]
Train data, Test data	1169 , 293
Train data, Test data (percentage)	80%, 20%

### Hierarchy of monitoring wells

A geographical representation of data points plotted on a x,y coordinate system is visualized in figure 6.14. The horizontal axis is labeled as longitude and the vertical axis is labeled as latitude. The data points are scattered across the study area within the neighborhood Rozenburg. Each monitoring well is color-coded according to the color bar on the right side of the plot. The color indicates the rank of each monitoring well and has a range between dark blue for the lowest values (1) to yellow for the highest values (29). The distribution of the color code does not follow a clear pattern within the neighborhood boundary. Some clusters of higher or lower ranked monitoring wells are present. The monitoring wells with additional value are marked blue, while the monitoring wells that likely do not have additional value to the network are marked yellow. The most redundant monitoring wells appear to be located towards the southern edge of the neighborhood. At the southern side, and a single well in the north as well, monitoring wells are clustered and address a region of low data variability and flashiness.



**Figure 6.14:** Visualization of the hierarchical list of 29 monitoring wells across Rozenburg. Based on EPSG:28992/Amersfoort RD New.

### Determining the optimal reduction rate

In figures 6.15 - 6.19, the reconstruction error of Rozenburg is displayed, plotting the number of monitoring wells on the x-axis against the root mean square error (RMSE) in meters on the y-axis. The line graph shows a decreasing function as the number of monitoring wells increases and the RMSE decreases, suggesting that more monitoring wells contribute to a lower reconstruction error in the data. In the figures, the orange marked point marks the optimal number of monitoring wells where the reconstruction error reaches the threshold that is indicated by the orange line. The threshold is an acceptable level of RMSE for the reconstruction process. The original groundwater monitoring network of Rozenburg includes 29 monitoring wells.

Starting with figure 6.15, a reduction of 10% describes that 26 monitoring wells remain in the network and only 3 monitoring wells might be eliminated. The reconstruction error reaches a value below 0.02 meters. The RMSE in figure 6.16 with a reduction of 25%, explains that a number of 21 monitoring wells is the most optimal number with a removal of 8 monitoring wells. Figure 6.17 demonstrates a function based on a reduction percentage of 50%. The RMSE decreases as more monitoring wells are added to the network. At the point of n=14 on the x-axis, an orange dot is indicated. Beyond this point, the decrease in RMSE slows down, suggesting returns on error reduction after a specific number of monitoring wells is reached.

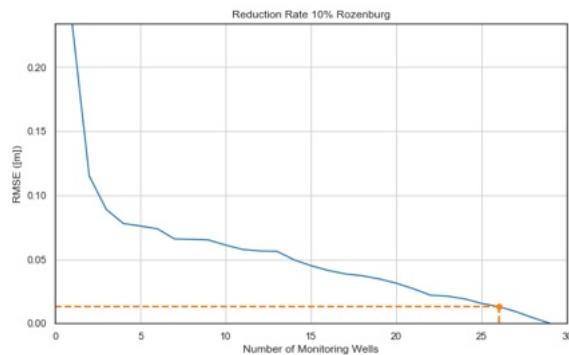


Figure 6.15: RMSE based on a network reduction of 10%. The optimal number of wells is 26.

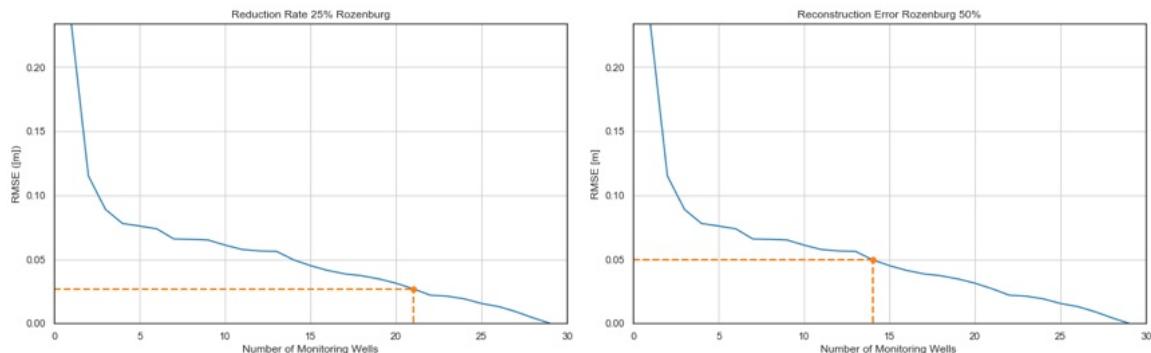


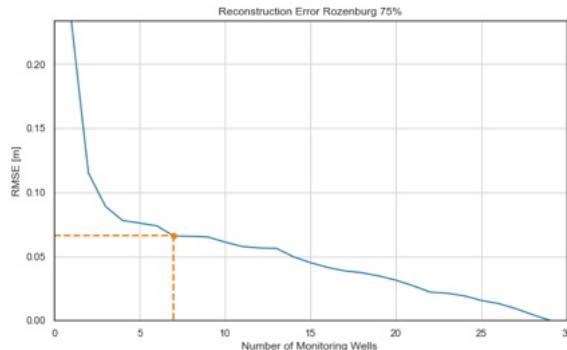
Figure 6.16: RMSE based on a network reduction of 25%. The optimal number of wells is 21.

Figure 6.17: RMSE based on a network reduction of 50%. The optimal number of wells is 14.

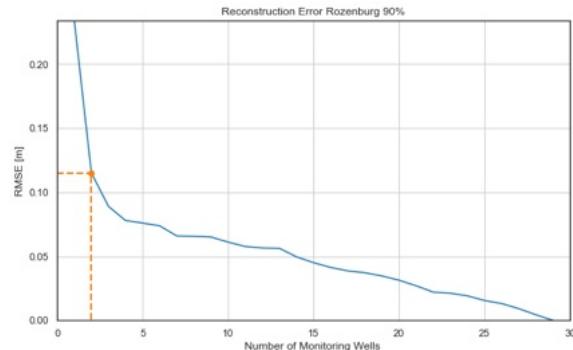
Figure 6.18 displays a function where the RMSE decreases as the number of monitoring wells increases. The RMSE sharply declines as the monitoring well count approach reaches the point above 10 monitoring wells. This point is marked by the orange dashed line. The line marks the optimal number of monitoring wells where the reconstruction error reaches the threshold value. The threshold value is determined by the trade off between the maximum information value and minimum cost, in which the maximum information value is characterized by the reconstruction error and the minimum cost is equivalent to the number of monitoring wells. Initially, Rozenburg includes 29 monitoring wells. The results of the RMSE explain that only 7 monitoring wells will remain, 22 monitoring wells have to be removed from the local network if a reduction of 75% is used in the approach. A reduction percentage of 75%

explains an equilibrium state between the number of monitoring wells and the reconstruction error that is achieved.

Continuing with figure 6.19, the reconstruction error for a reduction of 90% is displayed. The figure plots the number of monitoring wells on the x-axis against the RMSE in meters on the y-axis. The line graph shows a sharp decreasing function in the RMSE as the number of monitoring wells increase. The orange marked dot marks the optimal number of monitoring wells where the reconstruction error reaches the threshold value that is indicated by the orange line. The threshold value is an acceptable level of RMSE for the reconstruction process. Initially, Rozenburg has 29 monitoring wells. The results of the RMSE calculation explains that a number of 2 monitoring wells will remain with this reduction rate, 27 monitoring wells will be removed from the local network with a reduction rate of 90%.



**Figure 6.18:** RMSE based on a network reduction of 75%.  
The optimal number of wells is 7.

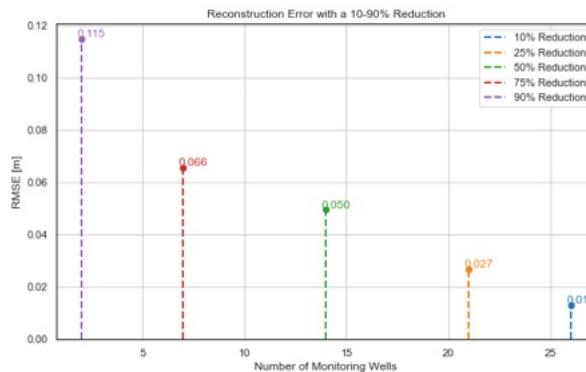


**Figure 6.19:** RMSE based on a network reduction of 90%.  
The optimal number of wells is 2.

### The optimal reduction rate

The optimal reduction rate for a monitoring network consisting of 29 monitoring wells counts up to a value where the RMSE reach stabilization. Figure 6.20 visualizes all reduction percentages between a range of 10-90% with their corresponding reconstruction error and mean absolute error.

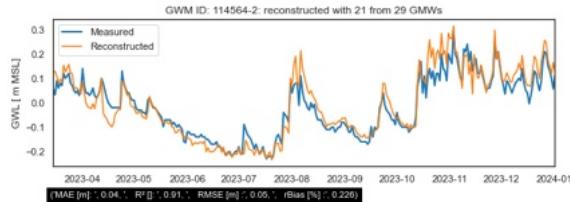
A reduction percentage of 25% implies that the optimal number of monitoring wells is 21, meaning that 8 monitoring wells might be eliminated, see figure 6.16. Additionally, the RMSE counts up to 0.027 meters. As stated in the chapter *Research Methodology*, the reconstruction capability of the



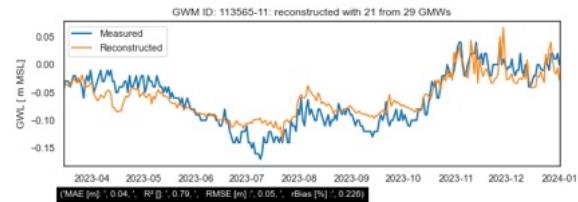
**Figure 6.20:** Overview of the RMSE for a range of reductions (10-25-50-75-90%).

eliminated monitoring wells is tested through the creation of hydrographs. In figures 6.21-6.28, hydrographs are shown. The hydrographs include a measured (blue line) and reconstructed (orange line) function. The reconstructed functions are based on observed data and the Goodness-of-Fit of the model. Performance metrics regarding the MAE, RMSE,  $R^2$ , and rBIAS determine the Goodness-of-Fit of the functions, they are available underneath the figures 6.21-6.28.

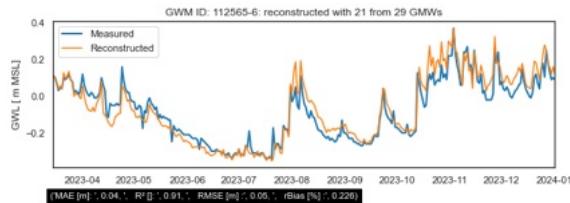
As can be seen from the figures, the plots have a comparable function; a decreasing groundwater level from April 2023 on and reaching the lowest groundwater level in July 2023. In August of the same year, the groundwater level increases again to high levels. At the end of August 2023, the groundwater level decreases again, but experiences peak values in October 2023. At the end of 2023, the groundwater levels are increasing again, but at this time of the year, the levels seem more constant. Based on the performance metrics of the potentially eliminated monitoring wells, it can be seen that the MAE and RMSE are constant values throughout the network. The MAE reaches a value of 0.04 meters and the RMSE counts up to 0.05 meters. The coefficient of determination differs for every monitoring well, but ranges between 0.79 and 0.95. An average of 0.905 [-] is calculated.



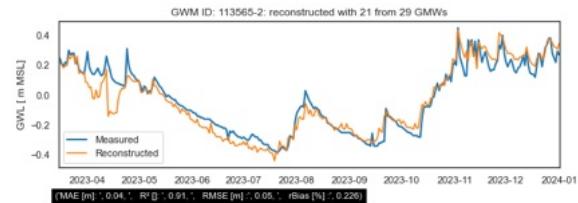
**Figure 6.21:** Eliminated monitoring well: 114564-2.



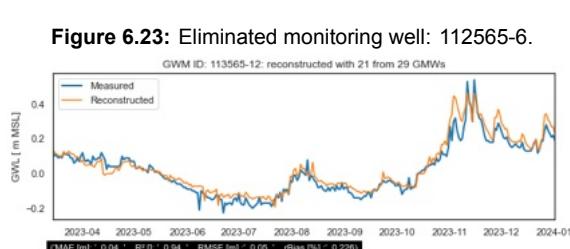
**Figure 6.22:** Eliminated monitoring well: 113565-11.



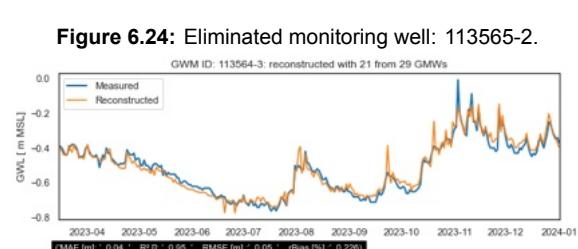
**Figure 6.23:** Eliminated monitoring well: 112565-6.



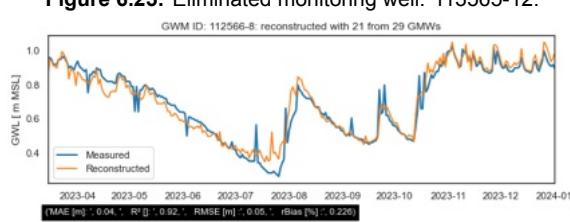
**Figure 6.24:** Eliminated monitoring well: 113565-2.



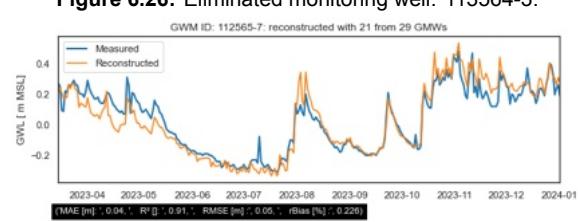
**Figure 6.25:** Eliminated monitoring well: 113565-12.



**Figure 6.26:** Eliminated monitoring well: 113564-3.



**Figure 6.27:** Eliminated monitoring well: 112566-8.



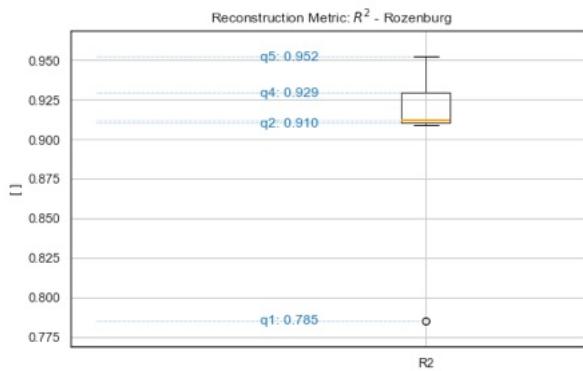
**Figure 6.28:** Eliminated monitoring well: 112565-7.

The measured and reconstructed groundwater levels are tested with the Welch's t-Test to determine whether there is a case of a significant difference between the measured and reconstructed groundwater levels. The results are displayed in table 6.2. Only 5 out of 8 eliminated monitoring wells experiences a significant difference between the measured and reconstructed groundwater level data. This means that the p-values of the monitoring wells are lower than  $\alpha = 0.05$ .

**Table 6.2:** Overview of statistical results of the eliminated monitoring wells with a network reduction of 25%. The overview explains the p-value of the monitoring wells and if they experience a significant difference between the observed and reconstructed data.

Monitoring Well	Result t-Test
112565-6	Significant difference (p-value = 0.002)
112565-7	No significant difference (p-value = 0.208)
112566-8	No significant difference (p-value = 0.115)
113564-3	No significant difference (p-value = 0.100)
113565-11	Significant difference (p-value = 0.036)
113565-12	Significant difference (p-value = 0.000)
113565-2	Significant difference (p-value = 0.002)
114564-2	Significant difference (p-value = 0.001)

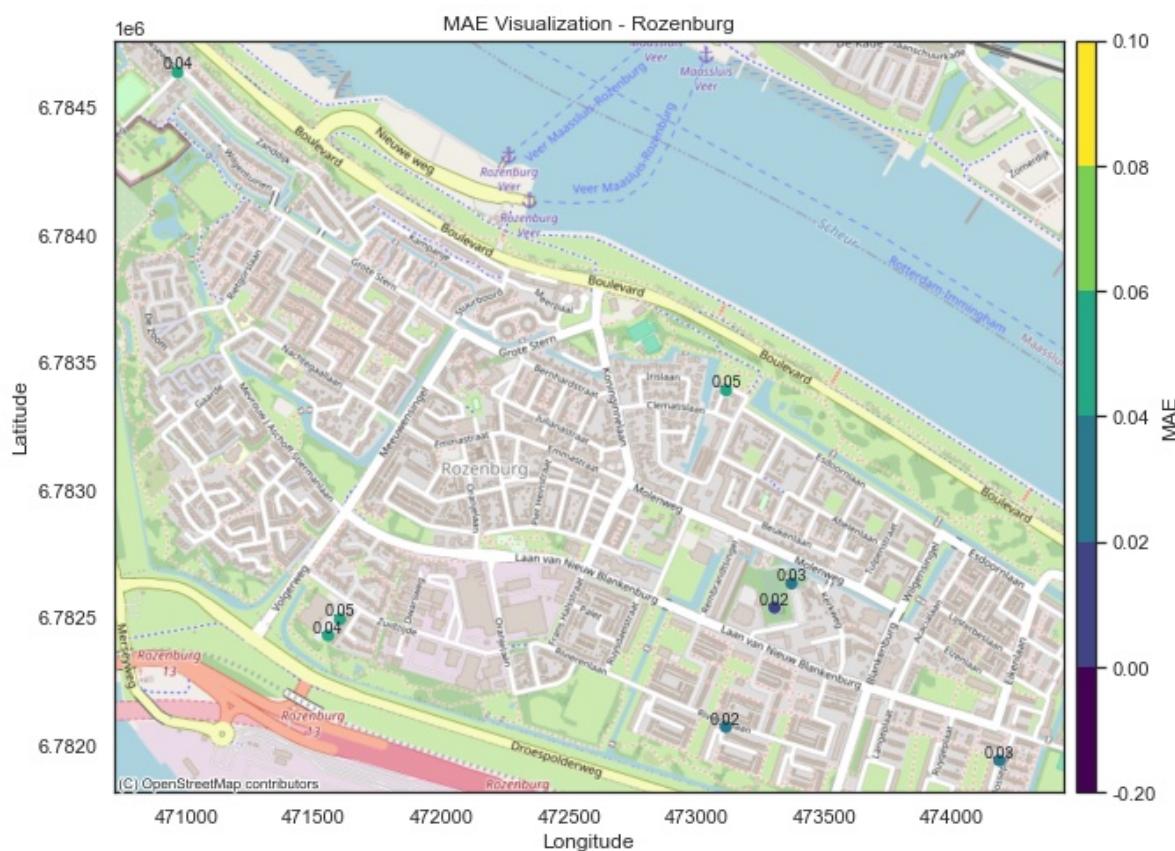
Figure 6.29 visualizes a box plot that displays the distribution of the performance metric  $R^2$ , the coefficient of determination, across the dataset after reduction of 25% took place. On the y-axis, the score in meters is shown, which ranges between 0.785 and 0.952. The orange line explains a median value of 0.912. The whiskers at the bottom and top of the box do not have equal length. The whisker at the bottom reaches a value up to 0.785 to a score of 0.910. The box itself ranges between 0.912 and 0.929 [-].



**Figure 6.29:** Performance metric  $R^2$  for Rozenburg. The orange line indicates the median value. The y-axis ranges from 0.785 to 0.952 [-].

#### Mean Absolute Error of Reduction

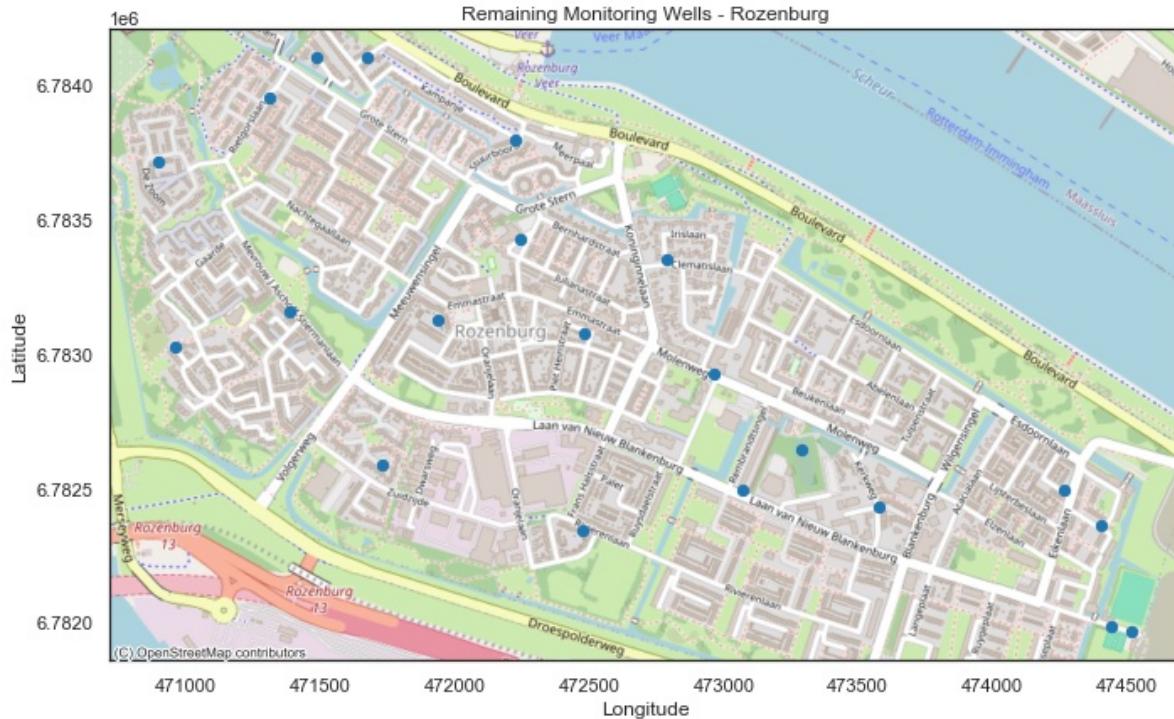
Based on a reduction percentage of 25%, reconstruction hydrographs could be plotted in figures 6.21-6.28. The extent to which the reconstructed groundwater level data corresponds to the actual observed data can be determined by the mean absolute error [meters]. The mean absolute error is calculated for every eliminated monitoring well. A color rank is visualized in figure 6.30 with the level of the calculated mean absolute error. Blue indicates a low MAE, while yellow indicates a high MAE. The visualization enables the identification of monitoring wells with higher or lower predictive accuracy, which could influence decision-making regarding the placement of future monitoring wells or the scope of maintenance efforts on existing wells. The MAE measures indirectly the performance of the monitoring well regarding data reconstruction. A high MAE in a network of lower MAE values can suggest a vulnerable monitoring well for external environmental factors. Monitoring wells with a high MAE can not be reproduced easily, because the prediction can be unreliable. The monitoring well that is marked with a yellow color indicates a higher MAE value compared to the other monitoring wells.



**Figure 6.30:** Mean Absolute Error [meters] across Rozenburg, using a color-coded system to represent the calculated MAE for the eliminated monitoring wells. Based on EPSG:28992/Amersfoort RD New.

### Remaining Network

After the network is reduced, 75% of the monitoring wells will remain in the GWMN after reduction takes place. A geographical representation of the monitoring wells is available in figure 6.31. The blue sites indicate the location of the existing monitoring wells. Initially, the GWMN of Rozenburg has a density of one monitoring well/11.17 hectares. After the reduction-optimization process, the network changes to one monitoring well/15.42 hectares.



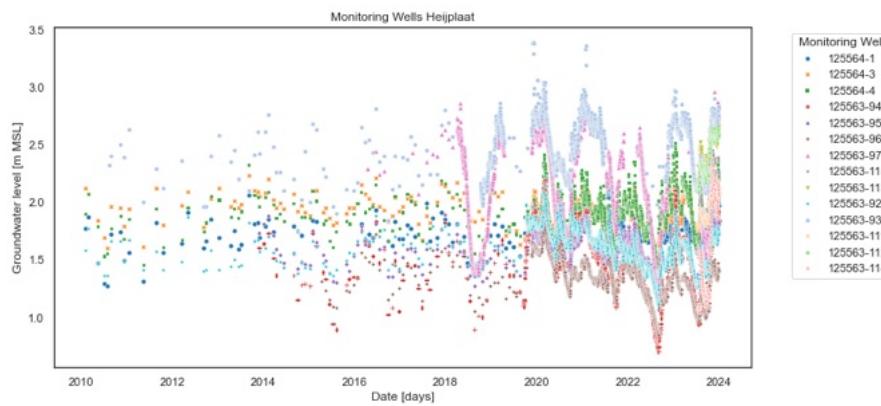
**Figure 6.31:** Map depiction of the remaining monitoring wells in Rozenburg after following a 25% reduction of the network.  
Based on EPSG:28992/Amersfoort RD New.

## 6.2. Heijplaat

### 6.2.1. QGIS and PROWAT

Figure 6.32 represents a multi-colored scatter plot, showing the groundwater level [m NAP] over time [days]. Each colored dot represents an individual data point from a unique monitoring well, see legend on the right side of the figure. A wide array of groundwater levels is depicted, ranging from +1.0 to +3.5 m NAP.

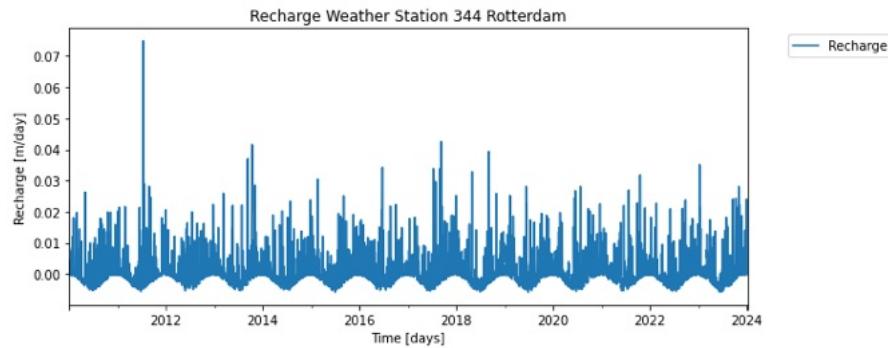
Plot 6.32 visualizes the variation in groundwater levels between different monitoring wells. The legend explains a range of colors for the unique monitoring wells that counts to 14 in the study area Heijplaat. The function of the groundwater levels is indicated by the colored dot for every monitoring well. When no dot is present, it means no data is available for this specific measurement date. Several monitoring wells perform constant groundwater levels, while other monitoring wells have abundant fluctuations in their function. Dense clustering of data points might suggest that measurements were taken frequently over time. The variety of colors allows for the possibility to compare groundwater level fluctuations between monitoring locations within the neighborhood.



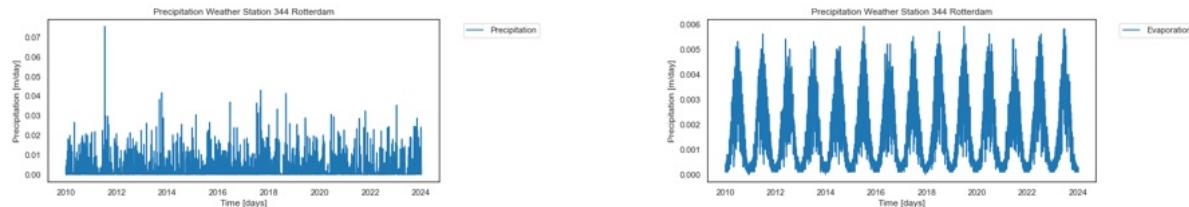
**Figure 6.32:** Observed data for 14 monitoring wells in Heijplaat.

### 6.2.2. Pastas time series modeling

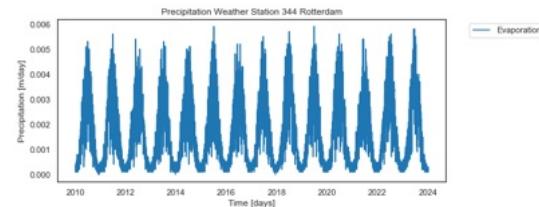
Similarly to the plot in the section of Rozenburg, the bar plot displays the recharge rates [m/day] over a period from 2010-2024, figure 6.33 . The y-axis quantifies the recharge rate with values ranging from 0 to 0.07 m/day. The bars show variability in recharge rates over time, where some years experience peaks in recharge (higher recharge rates), which can be due to increased precipitation. As can be seen from figure 6.33, the data is dense, indicating frequent measurements by KNMI. The pattern of the bars reflects seasonal changes. Lower values of recharge are more frequently observed than high outliers, possibly indicating a baseline level of recharge over the years. Two additional figures (6.34 and 6.35) of the precipitation and potential evaporation data are visualized as well. The data of these figures is combined to determine the recharge rate in the municipal area.



**Figure 6.33:** Recharge [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands.



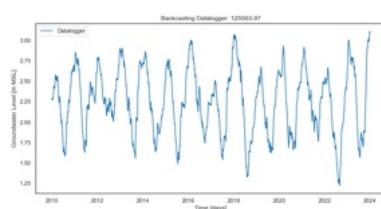
**Figure 6.34:** Precipitation [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands.



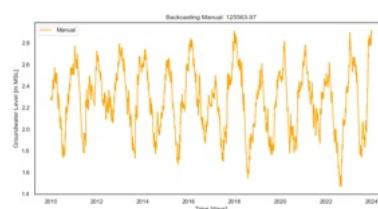
**Figure 6.35:** Potential evaporation [m/day] over a period of 2010-2024, measured by KNMI weather station 344 in Rotterdam, The Netherlands.

### Reversed forecasting

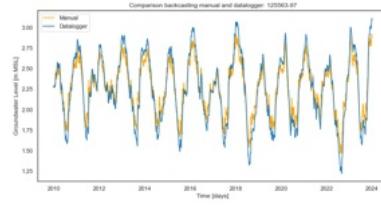
Reversed forecasting, or also known as backcasting, of the observed GWL data is a feature in the Pastas Python package. The process is applied for every unique monitoring well within the case study area. For every well, three types of visualizations are generated: 1) Backcasting for data logger data; 2) Backcasting for the manually collected data; 3) A combination figure of figures 1 and 2 combined, see figures 6.36-6.38. Monitoring well 125563-97 shows an example of the three visualized figures. The following applies for all three figures: The x-axis includes a period of 2010-2024 and the y-axis shows the simulated GWL data [m NAP].



**Figure 6.36:** Reversed forecasting data based on data logger data for monitoring well 125563-97.



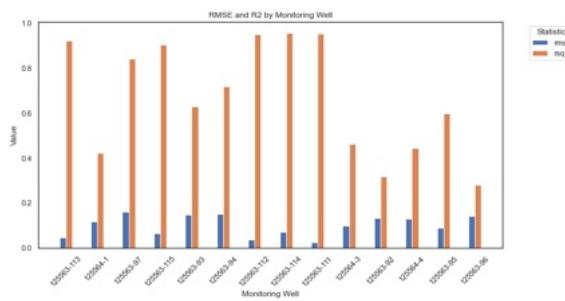
**Figure 6.37:** Reversed forecasting data based on manual measurement data for monitoring well 125563-97.



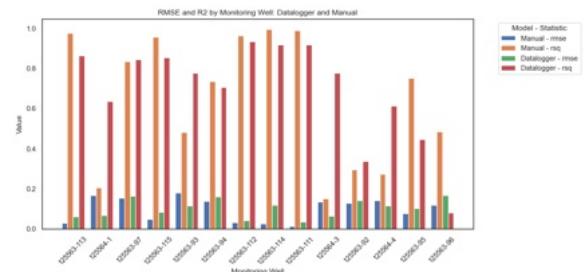
**Figure 6.38:** Reversed forecasting data based on manual measurement data and datalogger data for monitoring well 125563-97.

### Performance metrics

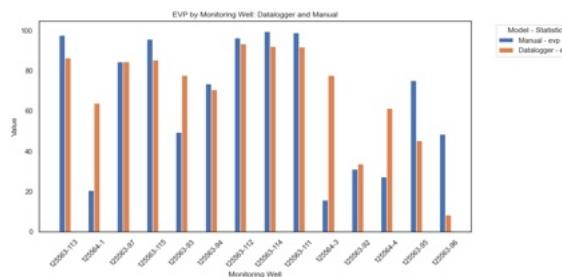
The bar plot 6.39 and 6.40 shows the two metrics RMSE and  $R^2$  for all unique monitoring wells. In figure 6.39, a first overview is given of the t-Test with a distinction between RMSE and  $R^2$  and an additional distinction between data loggers and manually collected wells. In figure 6.39, the  $R^2$  value (orange bar) of all monitoring wells seem higher than the values of the RMSE (blue bar) of the same monitoring wells. Figure 6.40 encompasses four categories: 1) manual-rmse; 2) manual-rsq; 3) datalogger-rmse; 4) datalogger-rsq. From this figure it is not clear if one of the datagroups has a better model performance, the performance of every model statistic differs between the monitoring wells. An additional bar plot, figure 6.41, describes the EVP [%] across the dataset. The EVP compares the variance of the observed data and the variance of the residual data across the 14 monitoring wells [3]. It is not clear how the values relate to each other, because there is a distinction between the datalogger and manual collection method. A low EVP might indicate that data could be missing in the dataset, the spatial pattern could be a reason. A statistical t-Test is a follow-up to identify and statistically substantiate the difference in performance between datagroups. Based on the results of the barplot, the decision was made to execute further statistical tests. A Welch's t-Test was carried out, leading to the following results. The Welch's t-Test uses an  $\alpha = 0.05$ .



**Figure 6.39:** Bar plot showing the metrics RMSE and  $R^2$  for 14 monitoring wells.

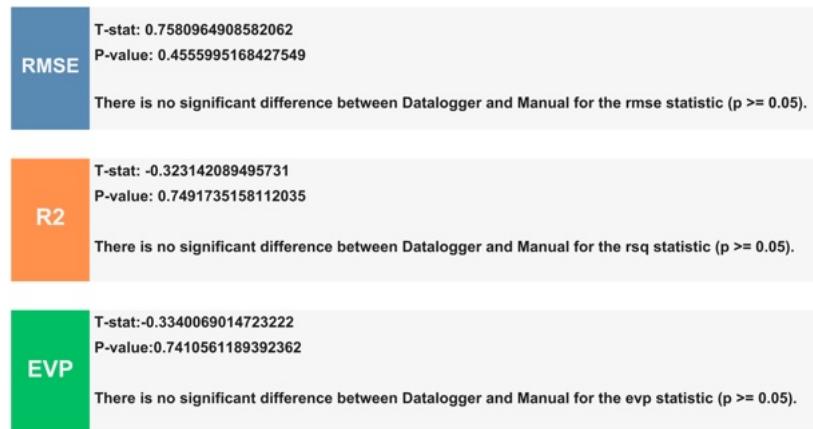


**Figure 6.40:** Bar plot showing metrics RMSE and  $R^2$  for 14 monitoring wells with a division in data logger and manual collection method.



**Figure 6.41:** Bar plot showing metric EVP for 14 monitoring wells.

Figure 6.42 displays the result of the Welch's t-Test, based on  $\alpha = 0.05$ . The results of the performance metrics are labeled as RMSE,  $R^2$ , and EVP. Each performance metric provides the results of the t-statistic and p-value of the Welch's statistical test. The RMSE (blue section), explains that the t-statistic is positive. The p-value is higher than  $\alpha = 0.05$ . A p-value lower than the set alpha normally explains that a significant difference is present between the data loggers and manual collection method. The  $R^2$  (orange section) explains that the t-statistic is negative, indicating a lower mean  $R^2$  for the data loggers than the manual collection method. The p-value is slightly higher than  $\alpha = 0.05$ , indicating no significant difference. The last metric is the EVP (green section). The t-statistic of the EVP has a negative value, suggesting that the data loggers might have a lower mean EVP than the manual collection method. The p-value is higher than the set alpha, but results in no significant difference.

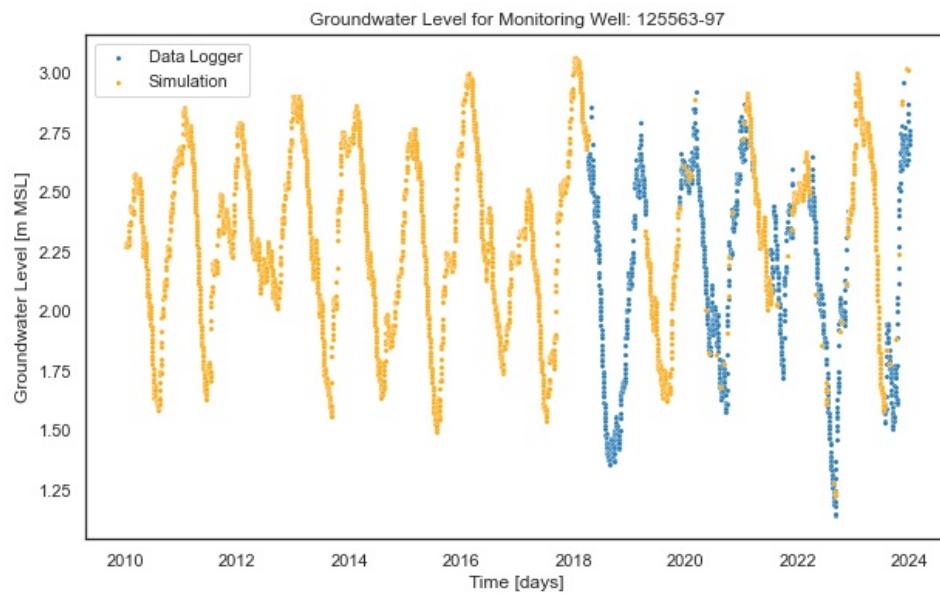


**Figure 6.42:** Overview of the performance of RMSE,  $R^2$ , EVP after the Welch's t-test and the difference between data loggers and manual collection method.

The Welch's t-Test does not reveal substantial discrepancies between the data groups. None of the p-values are below the threshold value of 0.05. Despite statistical significance, factors as data availability and reliability are decisive. Hence, the data logger group is chosen, merging observed and simulated data into one dataset and excluding the manual measurement from the research.

#### Creation of a new data frame

The scatter plot 6.43 below visualizes the observed data logger data with the blue scatter and the simulated data of the data logger is visualized by the orange scatter. On the x-axis the time period [days] is shown and on the y-axis the groundwater level [m NAP]. Figure 6.43 visualizes only monitoring well 125563-97, the remaining monitoring wells of the neighborhood are available through Github.



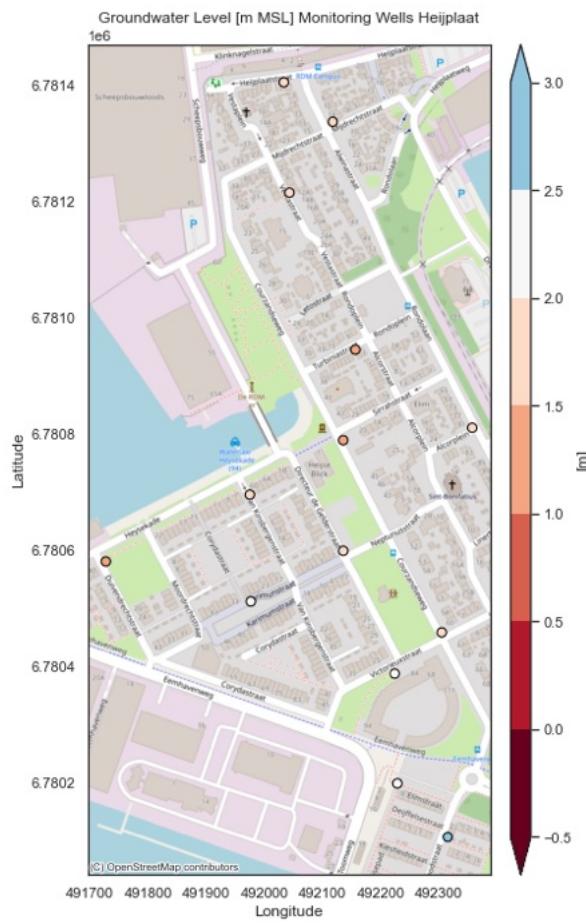
**Figure 6.43:** Scatter plot of simulated and observed data by data loggers for monitoring well 125563-97. The x-axis shows a period of 2010-2024 and the y-axis shows a range of +1.25 to 3.0 meters NAP.

### 6.2.3. QR factorization

Using QR factorization as foundation, a hierarchical list of the monitoring wells in Heijplaat can be created. The optimal reduction percentage is determined and reduction tests can be executed, resulting in an overview of eliminated monitoring wells and their capability to reconstruct future groundwater levels in the network.

#### 1D hydrograph data

Figure 6.44 is a geographical representation visualizing the spatial distribution of monitoring wells in the neighborhood Heijplaat. Within the map, a scatter plot with a color scale represents the groundwater level measurements taken at the monitoring wells. The groundwater level has a range between -0.5 m to +3.0 m NAP. Each monitoring well represents a location as well as the color of monitoring well represents the value on the color scale. At the south side of the neighborhood, the GWL is increasing from approximately 2.0 meters to 3.0 meters [NAP]. For Heijplaat the increasing GWL can not necessarily be defined through the elevation difference of the ground level, a digital terrain model of Heijplaat is displayed in figure 5.5.



**Figure 6.44:** Groundwater level [m NAP] across Heijplaat, using a color-coded system to represent the mean GWL observations for 14 monitoring wells. Based on EPSG:28992/Amersfoort RD New.

### Sampling data

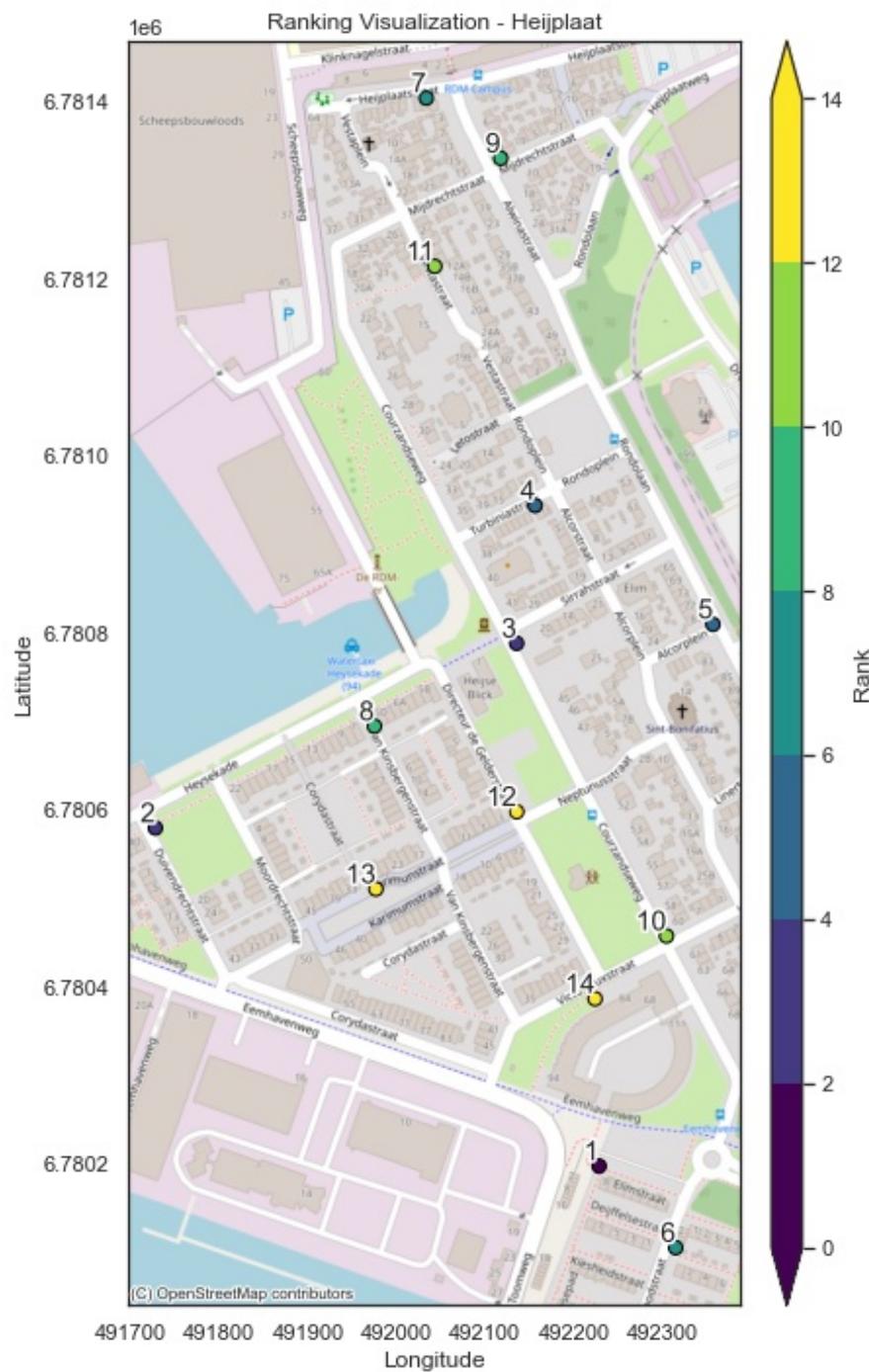
Preprocessing of the data includes transforming the data frame into an array, implementing both global and local centering methods, and splitting the dataset into an 80/20 ratio training and test set distribution. The result of the preprocessing is as follows, see table 6.3.

**Table 6.3:** Summary of sampling data.

Parameter	Value
Sampling period	2020-01-02 00:00:00 to 2024-01-01 00:00:00
Number of samples	1462
Number of features (sensors)	14
Shape of X (samples, sensors)	(1462, 14)
Min. and max. value	0.7 [m]; 3.36 [m]
Train data, Test data	1169 , 293
Train data, Test data (percentage)	80%, 20%

### Hierarchy of monitoring wells

A geographical representation of data points plotted on a x,y coordinate system is visualized in figure 6.45. The horizontal axis is labeled as longitude and the vertical axis is labeled as latitude. The data points are scattered across the study area within the neighborhood Heijplaat. Each monitoring well is color-coded according to the color bar on the right side of the plot. The color indicates the rank of each monitoring well and has a range between dark blue for the lowest values (1) to yellow for the highest values (14). The distribution of the color code does not follow a clear pattern within the neighborhood. Some clusters of higher or lower ranked monitoring wells are present. The monitoring wells with additional value are marked blue, while the monitoring wells that likely do not have additional value to the network are marked yellow. The most redundant monitoring wells appear to be located towards the southern edge of the neighborhood. At the southern side, monitoring wells are clustered and address a region of low data variability and flashiness.

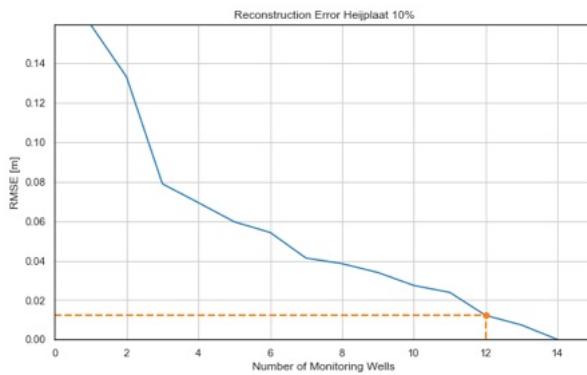


**Figure 6.45:** Visualization of the hierarchical list of 14 monitoring wells across Heijplaat. Based on EPSG:28992/Amersfoort RD New.

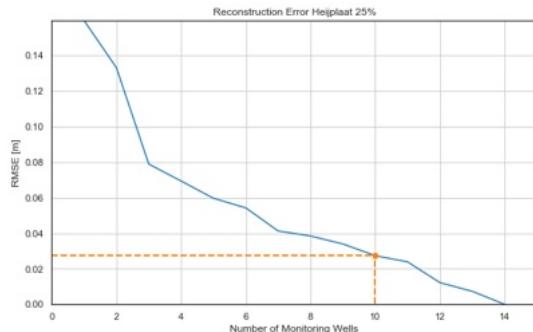
### Determining the optimal reduction rate

Figures 6.46-6.50 visualize the Reconstruction Error of Heijplaat, plotting the number of monitoring wells on the x-axis against the root mean square error (RMSE) in meters on the y-axis. The line graph shows a decreasing function as the number of monitoring wells increases and the RMSE decreases, suggesting that more monitoring wells contribute to a lower reconstruction error in the data. In the figures, the orange marked point marks the optimal number of monitoring wells where the reconstruction error reaches the threshold that is indicated by the orange line. The threshold is an acceptable level of RMSE for the reconstruction process. The original groundwater monitoring network of Heijplaat includes 14 monitoring wells.

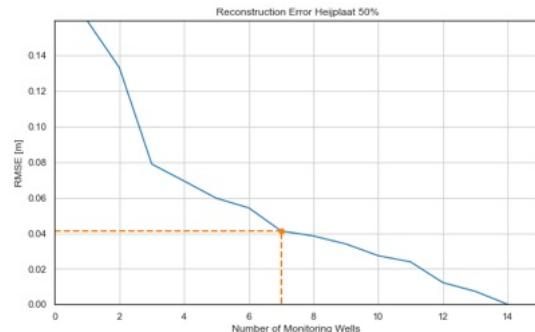
Starting with a reduction percentage of 10%, the network is reduced to 12 monitoring wells with a reconstruction error of below 0.02 meters, see figure 6.46. Figure 6.47 describes a reduction of 25% and explains that a number of 10 monitoring wells is the most optimal with a removal of 4 monitoring wells. The RMSE is calculated to determine the optimal number of monitoring wells in the neighborhood. Figure 6.48, demonstrates a function based on a reduction percentage of 50%. The RMSE decreases as more monitoring wells are added to the network. At the point of  $n=7$  on the x-axis, an orange dot is indicated. Beyond this point, the decrease in RMSE slows down, suggesting returns on error reduction after a specific number of monitoring wells is reached.



**Figure 6.46:** RMSE based on a network reduction of 10%. The optimal number of wells is 12.



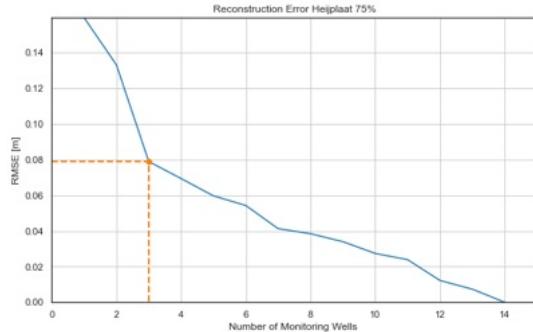
**Figure 6.47:** RMSE based on a network reduction of 25%. The optimal number of wells is 10.



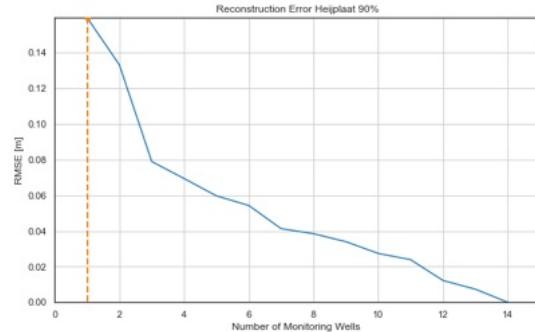
**Figure 6.48:** RMSE based on a network reduction of 50%. The optimal number of wells is 7.

Figure 6.49, displays a similar function where the RMSE decreases as the number of monitoring wells increases. The RMSE declines as the monitoring well count reaches the point above 3 monitoring wells. This is the point marked with the orange dashed line. The line marks the optimal number of monitoring wells where the reconstruction error reaches the threshold value. The threshold value is an acceptable level of RMSE for the reconstruction process. Initially, Heijplaat includes 14 monitoring wells. The results of the RMSE explains that only 3 monitoring wells remain in the network, 11 monitoring wells have to be removed from the local network if a reduction of 75% is being used in the approach. A reduction percentage of 75% explains an equilibrium state between the number of monitoring wells and the reconstruction error that is achieved.

Continuing with figure 6.50, the reconstruction error for a reduction of 90% is displayed. The figure plots the number of monitoring wells on the x-axis against the RMSE in meters on the y-axis. The line graph shows a decreasing function in the RMSE as the number of monitoring wells increases. The orange marked point, marks the optimal number of monitoring wells where the reconstruction error reaches the threshold value that is indicated by the orange line. The threshold value is an acceptable level of RMSE for the reconstruction process. Initially, Heijplaat includes 14 monitoring wells. The results of the RMSE calculation explains that a number of 1 monitoring well will remain in the network, while 13 monitoring wells will be eliminated from the network.



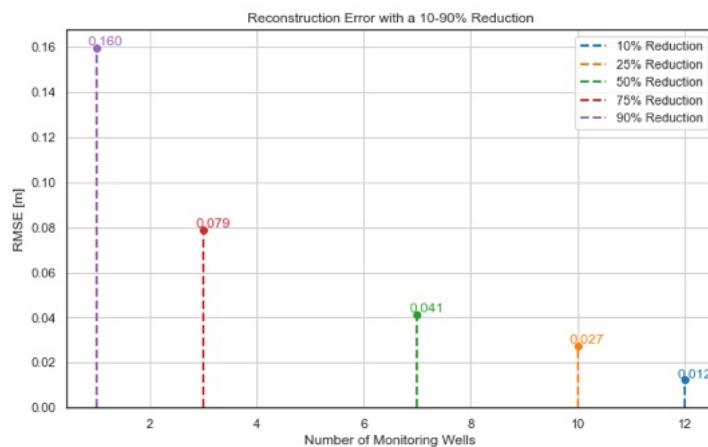
**Figure 6.49:** RMSE based on a network reduction of 75%. The optimal number of wells is 3.



**Figure 6.50:** RMSE based on a network reduction of 90%. The optimal number of wells is 1.

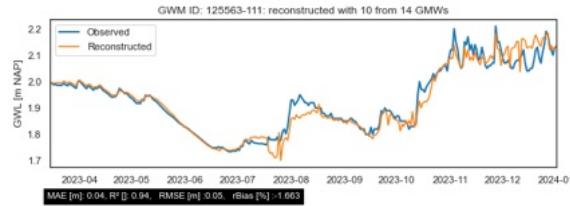
### The optimal reduction rate

The optimal reduction rate for a monitoring network consisting of 14 monitoring wells counts up to a value where the RMSE reaches a point of stabilization. Figure 6.51 visualizes all reduction percentages between a range of 10-90% with their corresponding reconstruction error. A reduction percentage of 25% implies that the optimal number of monitoring wells is 10, meaning that 4 monitoring wells might be eliminated from the network, see figure 5.48. Additionally, the RMSE counts up to 0.027 meters. As stated in the chapter 4 the reconstruction capability of the eliminated monitoring wells is tested through the creation of hydrographs. In figures 6.52-6.55, hydrographs are shown. The hydrographs include a measured (blue line) and reconstructed (orange line) function. The reconstructed functions are based on observed data and the Goodness-of-Fit of the model. Performance metrics regarding the MAE, RMSE,  $R^2$  and rBIAS determine the Goodness-of-Fit of the functions, they are available underneath the figures 6.52-6.55.

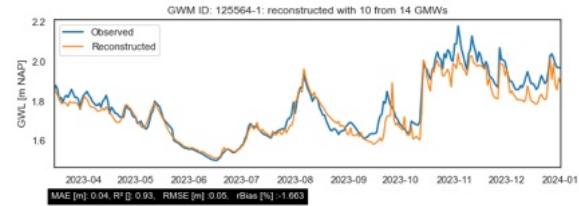


**Figure 6.51:** Overview of RMSE for a range of reduction (10-25-50-75-90%).

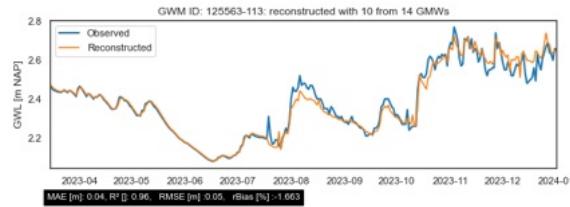
As can be seen from the figures, the plots have comparable functions; a decreasing groundwater level from April to July 2023. An increase occurs in July towards mid August 2023. During the start of Fall, a small increase in groundwater level occurs in October with strong increases in November and December 2023. In January 2024, the groundwater level is just as high as at the beginning of November 2023. Based on the performance metrics of the potentially eliminated monitoring wells, it can be seen that the MAE and RMSE are constant values throughout the network. The MAE reaches a value of 0.04 meters and the RMSE counts up to 0.05 meters as well. The coefficient of determination differs for every monitoring well, but ranges between 0.87 and 0.96. An average  $R^2$  of 0.925 [] is calculated.



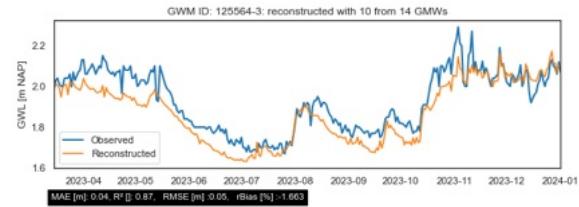
**Figure 6.52:** Eliminated monitoring well: 125563-111.



**Figure 6.53:** Eliminated monitoring well: 125564-1.



**Figure 6.54:** Eliminated monitoring well: 125563-113.



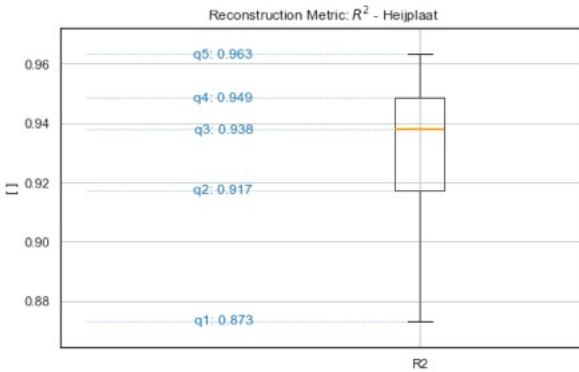
**Figure 6.55:** Eliminated monitoring well: 125564-3.

The measured and reconstructed groundwater levels are tested with the Welch's t-Test to determine whether there is a case of a significant difference between the measured and reconstructed groundwater levels. The results are displayed in table 6.4. For Heijplaat, 3 out of 4 eliminated monitoring wells experience a significant difference between the measured and reconstructed groundwater level data. This means that the p-values of the monitoring wells are lower than  $\alpha = 0.05$ .

**Table 6.4:** Overview of statistical results of the eliminated monitoring wells with a network reduction of 25%. The overview explains the p-value of the monitoring wells if they experience a significant difference between the observed and reconstructed data.

Monitoring Well	Result t-Test
125563-111	No significant difference (p-value = 0.134)
125563-113	Significant difference (p-value = 0.000)
125564-1	Significant difference (p-value = 0.000)
125564-3	Significant difference (p-value = 0.000)

Continuing with a box plot of the coefficient of determination, the  $R^2$ . Box plot 6.56 visualizes the distribution of the  $R^2$  across the dataset after a reduction of 25% took place. On the y-axis, the score in meters is shown, which ranges between 0.873 and 0.963. The orange line explains a median value of 0.938 [ ]. The whiskers at the bottom and top of the box do not have equal lengths. The bottom whisker ranges from 0.873 to 0.917, while the whisker at the top ranges from 0.949 to 0.963 [ ].



**Figure 6.56:** Performance metric  $R^2$ . The orange line indicates the median value. The y-axis ranges from 0.928 to 0.987 [ ].

#### Mean Absolute Error of Reduction

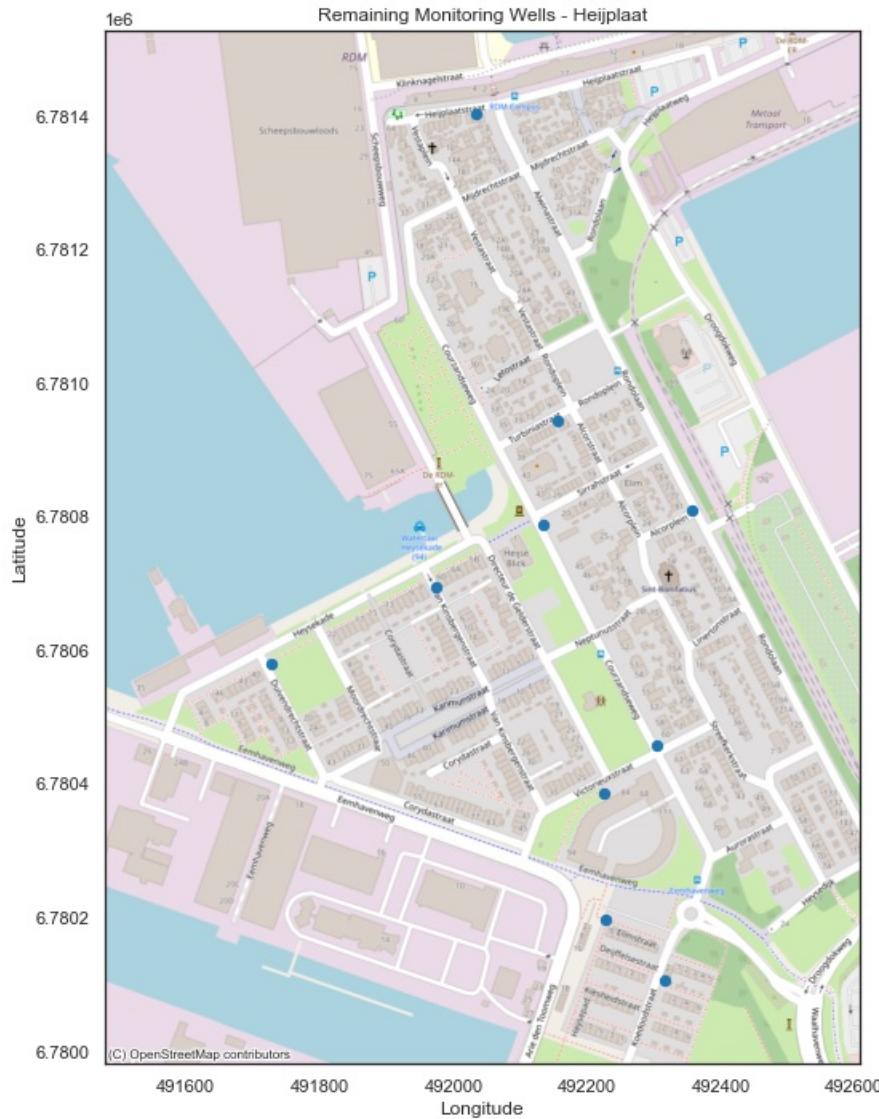
Based on a reduction percentage of 25%, reconstruction hydrographs could be plotted in figures 6.52-6.55. The extent to which the reconstructed groundwater level data corresponds to the actual observed data can be determined by the mean absolute error [meters]. The mean absolute error is calculated for every eliminated monitoring well. A color rank visualizes the level of the calculated mean absolute error in figure 6.57. Blue indicates a low MAE, while yellow indicates a high MAE. The visualization enables the identification of monitoring wells with higher or lower predictive accuracy, which could influence decision-making regarding the placement of future monitoring wells or the scope of maintenance efforts on existing wells. The MAE measures indirectly the performance of the monitoring well regarding data reconstruction. A high MAE in a network of lower MAE values can suggest a vulnerable monitoring well for external environmental factors. Monitoring wells with a high MAE can not be reproduced easily, because the prediction can be unreliable. The monitoring well that is marked with a yellow color indicates a higher MAE value compared to the other monitoring wells.



**Figure 6.57:** Mean Absolute Error [meters] across Heijplaat, using a color-coded system to represent the calculated MAE for the eliminated monitoring wells. Based on EPSG:28992/Amersfoort RD New.

### Remaining Network

After the network is reduced, 75% of the monitoring wells will remain in the GWMN after reduction takes place. A geographical representation of the monitoring wells is available in figure 6.58. The blue sites indicate the location of the remaining monitoring wells in the network. Originally, the neighborhood Heijplaat had a density of one monitoring well/2.78 hectares. After the reduction-optimization process, the network changes to one monitoring well/3.90 hectares.



**Figure 6.58:** Map depiction of the remaining monitoring wells in Heijplaat after following a 25% reduction of the network.  
Based on EPSG:28992/Amersfoort RD New.

# 7

## Discussion

The discussion critically examines the findings on optimizing the groundwater monitoring network of Rotterdam, based on groundwater level data from the period of 2020-2024. The discussion aims to give insight into the implications and suggested directions for future investigation and application of the model into the field.

### Observed data

The substantiation for the time period of the observed data, ranging from 2010-2024, to a more narrow scope of 2020-2024 potentially affects the precision and reliability of the results of the stress model. More specifically, a wider time frame may provide a better view of the groundwater dynamics and anomalies, referring to the outliers recorded in the observed manual measurements between 2014 and 2018. Nevertheless, an extended time period might introduce variability, which does not improve model performance. A shorter time frame, however, might not capture the entire range of hydrological events that are necessary to create a robust model. On the other hand, a wider time frame can positively influence the results of the model, but taking into account the effects of external factors such as spatial planning or inefficient sewage systems, could influence variability of the data. The observed data is collected by field service employees of the field measurement service of City of Rotterdam through manually collected measurements. The measurements take place at the end of the month, so sampling every month. The manual collection method could be the reason as to why monitoring wells with the manually collected data have more data inconsistencies. Besides that, the transition to different field service employees in the year 2014 and 2018 possibly resulted in differences, and maybe even errors, in the approach of data collection. From 2019 on, groundwater data is collected solely by City of Rotterdam.

### Data preparation

Since a combination of data loggers and a manual collection method is applied for multiple neighborhoods within the municipality, data gaps are identified. Time series modeling appeared to be an effective approach for bridging data gaps with dependable data. A query is, however, if the measuring frequency of the City of Rotterdam should be adjusted, to hourly measurements for example, in order to produce more frequent data. This way, data gaps do not have to be bridged with the use of Pastas time series modeling. Pastas time series modeling is based on simulation data of precipitation [m/day] and potential evaporation [m/day] data, in order to calculate the recharge rate [m/day]. For this research study, the data is based on weather station 344 in Rotterdam, The Netherlands. Weather station 344 is the station that is commonly used to retrieve data for City of Rotterdam. However, as the crow flies, both Rozenburg and Heijplaat have a distance of approximately 15 kilometers from the weather station. Therefore, the measured data at weather station 344 can differ between the in reality observed and measured precipitation and evaporation in the neighborhoods Rozenburg and Heijplaat. However, the use of Pastas time series modeling presents a limitation because of its focus on evaporation and precipitation parameters. Hydrological parameters such as hydraulic conductivity and drainage are not considered in the model. This raises questions regarding data gaps: 1) How should data gaps be addressed upon identification?; 2) How does inclusion of simulated data, to fill data gaps, influence the

model results? Addressing these queries in future research should be a priority to optimize the model and its results, see chapter 9.

### **QR factorization**

The process of QR factorization and the additional reduction of monitoring wells introduces the relationship between network optimization and the RMSE performance metric. The findings of the model suggest that an optimized network is marked by both a reduced number of monitoring wells and a lower RMSE for the network, indicating that the level of optimization is correlated to the quality of data derived from the chosen monitoring locations rather than the quantity of data. Therefore, achieving a high level of optimization necessitates a careful selection of monitoring wells that ensures data quality for the monitoring network. Data quality means, in this context, the possibility to lower the RMSE without influencing the data quantity. Strategic selection of monitoring wells through QR factorization enables the network to maintain and improve the model performance with fewer monitoring wells.

### **Sensitivity and performance metrics**

It is necessary to question whether the RMSE and MAE are the only sensitive parameters that reflect model accuracy. The RMSE describes a measure of the magnitude of errors and it remains unaffected by changes in the variability of the input of the observed groundwater level data. As stated by Ohmer, a greater RMSE, in comparison with a MAE, indicates a larger variation of errors and highlights areas of variability in the performance of the reconstruction of the eliminated monitoring wells [35]. It has been observed that monitoring wells with a lower variability in data are likely to add less informational relevance to the network and are, thus, more likely to be eliminated from the network. Besides, the added value of centering the input data is questionable. Centering is an often used method for preprocessing data, when the reference point is the mean of the dataset. The result of centralization encompasses an analysis of class-centered data that examines the variance within sample groups, but ignores variation in data within the sample groups. Applying global and local centering, followed by decentralizing the data in the reconstruction phase, influences the outcomes of the RMSE, MAE,  $R^2$ , and rBIAS of the monitoring wells. Essentially, applying a centralization approach ensures that the mean of the observed data is as near as zero as achievable.

### **Implications of reduction rates**

In the research methodology, reduction percentages ranging from 10-90% are tested. When the reduction percentage reaches 25%, monitoring wells with the lowest information content are identified and eliminated from the network. With a reduction percentage of 25%, the network is polished, the prediction accuracy is slightly influenced, and there is the least amount of network destruction. Therefore, the remaining reduction percentages of 10, 50, 75, and 90% are not considered in the elimination process. The elimination process is based on 25% for both study areas Rozenburg and Heijplaat. It is hypothesized that the results according to a reduction of 25% have a lower loss of prediction accuracy for the RMSE, reconstruction accuracy and performance metrics (RMSE, MAE,  $R^2$ , rBIAS) of the eliminated monitoring wells in comparison with the other stated reduction scenarios. A low prediction accuracy, meaning low performance metrics, could be the result of previous processes like the type of data collection and modeling implications of Pastas time series modeling.

### **Model robustness and variability**

The RMSE remains unchanged when increasing or decreasing the variability of the input data. This suggests that the model performance is consistent and the robustness to changes in variability is relatively high. Determining the robustness is necessary to examine whether the model is able to remain consistent when experiencing a variability of hydrological conditions and input scenarios. Accordingly, the elimination of monitoring wells and the choice of study period requires an equilibrium regarding data quality and quantity. Both the RMSE and MAE are performance metrics that evaluate model performance, because they ensure perspective in model sensitivity and performance. From the model results, it has become clear that the RMSE is not sensitive to changes in variability input. At the same time, the direct influence of the MAE on the results marks the complexity of GWMNO.

Additionally, both results of Rozenburg and Heijplaat mark a significant difference between the observed and reconstructed groundwater level data of the eliminated monitoring wells. Even though the performance metrics of the eliminated monitoring wells mirror a good fit of the reconstructed ground-

water level data with the observed groundwater level data by City of Rotterdam. An explanation for the difference between the results of the performance metrics and Welch's t-Test could be the objective of the assessment, meaning that the Welch's t-Test indicates the probability of observed data if the hypothesis is true. Nonetheless, the performance metrics can address a good fit of the reconstructed and observed data, a t-Test can address if there are statistically significant differences between the two data sets. The combination of assessing both model performance and a statistical test provides insight in model accuracy that could be beneficial for further model development and application.

### **Implementation of the model in the field**

Before the research study started, the objective was set to create a generic model to be able to test the level of optimization for the entire municipal area of Rotterdam on neighborhood scale. Two case study areas were selected: Rozenburg and Heijplaat, to test and train the model with a low number of monitoring wells. A query is, however, if the methodology and model would be robust enough to handle external factors in urban areas, that correlate to the location of wells, that would possibly have an influence on groundwater levels locally. For instance: The design of the filter of the well in correlation with clay or sand, the proximity to water bodies, and construction activities that might affect surface runoff or impact the sewage system. The reduction-optimization approach relies on the proposition that the physical system remains unchanged in its present state in order to operate effectively.

# 8

## Conclusion

The exploration of the groundwater monitoring network optimization within City of Rotterdam is the central point of this research study. The study addresses the following research question:

*"To what extent is the groundwater monitoring network of City of Rotterdam optimal?"*

Sub questions that were introduced in chapter 1 *"Introduction"* are answered.

1. What constitutes an optimal groundwater monitoring network, particularly in terms of the ideal reduction in monitoring wells, to maintain effective monitoring while maximizing efficiency?

Theoretically, an optimal groundwater monitoring network is defined as a network that achieves the maximum possible information content at the lowest feasible expenses, which aligns with the Pareto Optimum principle. This means that the lowest number of monitoring wells ensures the highest quality of data through limiting construction measures regarding removal and reconstruction. An optimal GWMN with high quality data does not necessarily mean that high quantities data need to be available from the monitoring wells in the area. A network can be optimized by improving the density and measuring frequency of the monitoring wells. In practice, the placement of the monitoring wells should be optimal to ensure that they cover critical areas, vulnerable to the distance of a water body, construction works that could influence surface runoff and sewage system, the filter construction in relation to clay or sand. Next to that, the network should be able to retrieve high quality data rather than high quantities data. Every monitoring well should be equipped with certified sensors to ensure accurate data. Ultimately, a GWMN reaches its optimal state when it gathers crucial data through efficient placement of monitoring wells with data loggers, ensuring saving of expenses among others. The approach focuses on quality over quantity of data to maximize the network's impact and effectiveness in managing groundwater resources. For edification: Recommendations regarding including cost analysis in the model are described in the chapter 9.

2. How is the hydrogeological system characterized within the municipal boundaries of Rotterdam?

The hydrogeological system of City of Rotterdam can be addressed as a fairly complex heterogeneous formation. The deeper Holocene layer is characterized by formations that are marked by local gully erosion and cuts with layers that differ in structure, composition, density and permeability. Due to the influence of the North Sea and adjacent rivers in The Netherlands, there is a significant spatial variation in the Holocene layer. The Holocene layer is followed by a phreatic formation that is called the Anthropocene layer. This layer is permeable and the thickness varies between centimeters to 30 meters at the western side of the municipal area, reaching to the dunes of Hoek van Holland. In the urban areas, the top layer is strongly affected by anthropocentric influences. Like the Holocene layer, the Anthropocene layer varies in composition, but mainly consists of sand.

Zooming into the case studies, in Rozenburg the Holocene formations consist of a 21 meter thick layer, which consists of sandy-clay, medium-fine sand, peat, and coarse sand. The presence of these

sediments suggest a varied deposition environment that are influenced by fluvial and marine processes. The diverse grain size and the presence of an organic peat layer could influence the infiltration rate and so the local groundwater recharge and storage capacity. For the second study area, Heijplaat is located at the same side of the "Nieuwe Waterweg", but the lithology of the area is different than Rozenburg. Unlike Rozenburg, the subsurface of Heijplaat begins with a 4 meter thick anthropocentric layer that is followed by a complex, Holocene layer that consists of sandy-clay, medium-fine sand, clay, peat, and little coarse sand. An under-layer of peat indicates wet conditions in the past, which could influence the groundwater dynamics with layers of varying hydraulic properties. In Heijplaat, the peat landscape has been drained, which altered the landscape and resulted in reshaping groundwater recharge, flow conditions, and the storage ability in the area.

Given these facts, understanding the hydrogeological conditions of the study areas is crucial for accurately assessing groundwater dynamics. The knowledge is essential for effective groundwater management and, more specifically, setting up a framework regarding the evaluation of the level of optimization of the GWMN of Rotterdam.

### 3. Does a correlation exist between network density and the degree of optimization?

A dense network of monitoring wells does not directly indicate a high level of optimization. The research methodology that is applied for this study incorporates a reduction-optimization approach, in which a high density of monitoring wells suggests potential for optimization. Accordingly, the network density may decrease as a result of applying the reduction-optimization method, signifying improved efficiency. Ultimately, a correlation does exist between network density and the degree of optimization when looking at the information content and data quality of the monitoring wells, as a higher network density tends to offer more coverage but mostly a lower degree of optimization. Thus, while a denser network provides broader coverage of the network, optimization is achieved through strategic reduction and so enhancement of data quality, indicating a nuanced correlation between GWMN density and the efficiency of the reduction-optimization approach.

### 4. How does increasing the measuring frequency and density increase the effectiveness of the groundwater monitoring network?

A GWMN with a high quantity of data does not necessarily equate to high qualities of data. Frequent monitoring can contribute to data quality by providing more consistent and reliable datasets that are less susceptible to anomalies or data gaps. Obstacles are often shortages in data and the occurrence of data gaps. In this research study, the Python package Pastas was used, because it provides a time series modeling technique to ensure that data gaps could be filled with simulated values. Increased measuring frequency and density can reduce the reliance on simulated values, leading to a dataset that is more representative of the actual conditions of the study area. Dense and more frequent data enhances the ability to monitor changes in trends to detect potential threats or failures and execute measures promptly. Overall, enhancing the frequency and density of groundwater monitoring does not only improve data reliability and accuracy, but also improves the capacity of the network for threat detection and immediate response, elevating the effectiveness of groundwater management within City of Rotterdam.

## 5. What is the optimal monitoring reduction rate to achieve the most optimal state of the GWMN?

To begin with, the optimal reduction percentage differs across study areas, it is not a one-size-fits-all solution. The ideal reduction percentage, suitable for a specific neighborhood, varies depending on the characteristics of the study area. In this research study, a reduction percentage of 25% is chosen for both study areas Rozenburg and Heijplaat for the reduction-optimization approach. A substantiation for this decision is that a minimum reduction percentage of 10% is suitable for a reduction-extension optimization approach, as stated by Ohmer [35] previously. When only executing a reduction-optimization approach, 25% reduction polishes the network and the prediction accuracy of the reconstructed groundwater data is not influenced. Therefore, the remaining reduction percentages that were discussed in chapter 4 are not considered in the elimination process. The approach removes a number of monitoring wells, based on the reduction percentage, regarding their mutual informational relevance. Irrelevant monitoring wells are eliminated based on their informational relevance. Their groundwater data is simulated with a hydrograph reconstruction tool. The data of the eliminated network integrity is maintained through the hydrograph reconstruction tool, despite the reduction-optimization approach. Consequently, a reduction-optimization approach underpinned with a 25% reduction, ensures the GWMN can function effectively and is able to maintain a high quality of collected data and analysis capabilities.

The above mentioned and answered sub questions act as a tool to answer the main research question: "*To what extent is the groundwater monitoring network of City of Rotterdam optimal?*"

The research study encompasses two neighborhoods within City of Rotterdam to test the generated model. Therefore, it is a challenge to assess the degree of optimization of the monitoring network for the entire municipal area. The model results indicate changes in monitoring well distribution with an overall increase in the ratio of monitoring wells to surface area across both study areas, see table 8.1. The overview of the change in the ratio of monitoring well to surface area for both study areas represents the first step towards GWMNO. The methodology that is outlined in this research study offers a framework to evaluate the level of optimization. Following the framework, the approach focuses on maximizing the information content while minimizing expenses for City of Rotterdam, aligning with the Pareto Optimum principle, labelled as a reduction-optimization approach. The change in the ratio of monitoring wells to surface area, changes the efficiency of the network. By reducing the amount of monitoring wells, the expenses of the overall network are lowered as well as other parameters as frequency and intensity of required maintenance proceedings. In this research study, the model is developed and applied to Rozenburg and Heijplaat. Yet it is possible to apply the model to the remaining neighborhoods of City of Rotterdam. Interested parties have to take into account that the model is not a 'one-size-fits-all' solution, as well as a measure for the optimal density of monitoring wells in urban areas.

**Table 8.1:** Overview of the ratio monitoring well to surface area across study areas Rozenburg and Heijplaat.

Area	Prior	Posterior
Rozenburg	one well/11.17 ha	one well/15.42 ha
Heijplaat	one well/2.78 ha	one well/3.90 ha

The results in the table above represent the first step towards GWMNO. The methodology outlined in this research study offers a framework to evaluate the level of optimization across the entire municipal area. Following the framework, further enhancement of the municipal area within the GWMN could be pursued through an additional extension-optimization approach. This future step towards optimization is discussed in chapter 7 and chapter 9.

# 9

## Recommendations

Based on the results of the research study, the main suggestion for this study is to execute additional research regarding a reduction-extension approach.

### **1D reduction and 2D extension methodology**

Instead of researching 1D hydrograph data only, it is possible to engage more into groundwater level contours to ensure reconstruction and extension are included. Through this approach, an indication of the predicted expenses can be depicted, which could persuade municipalities or interested parties to optimize their local groundwater monitoring network. The gradual cost function is applicable in scenarios where weighting increases with the proximity to an infrastructural object. With the 2D approach, the weighting depends upon the system, basis, and cost function and should be adjusted to suit the research area. Additionally, Ohmer states that the reconstruction error for the reduction-extension optimization approach is slightly lower than the results of the reduction-optimization approach [35].

### **Municipal area of Rotterdam**

The boundaries of the municipal area of Rotterdam reach from Hoek van Holland to Kralingen and the industrial areas. Gemeente Rotterdam desires to construct additional monitoring wells in public spaces throughout the industrial areas. Use of a reduction-extension model could be useful to examine if they should be removed, replaced or extended by adding monitoring wells to the network. And, to determine if the network reaches its optimal state.

### **Optimizing the model**

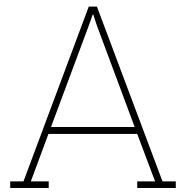
The optimization-reduction approach could be refined by expanding the hydrological perspective of the model. In the current study, the hydrological parameters evaporation and precipitation are included in Pastas time series modeling. However, this gives the time series a limited, hydrological perspective, as mentioned in 7. The time series model could be extended in the future with parameters as hydraulic conductivity or drainage for example. Another research limitation regarding the model includes the unknown mathematical calculations that are present in the libraries of the used Python packages, for example PySensors. The calculations that are present in the libraries of the PySensors package defines the mathematical substantiation for the reconstruction of the groundwater level data. Therefore, it is recommended to determine which formulas and calculations have been carried out to achieve the results from the model. Potentially, these formulas could be used to forecast and reconstruct groundwater level data in hydrographs.

# References

- [1] Stichting Infrastructuur Kwaliteitsborging Bodembeheer [SIKB]. *Plaatsen van handboringen en peilbuizen, maken van boorbeschrijvingen, nemen van grondmonsters en waterpassen. Protocol 2001*. Dec. 2013.
- [2] "Sensitivity analysis of Welch's T-Test". In: ed. by Ahad and Syed-Yahaya. Vol. 1605. 2014. DOI: 10.1063/1.4887707. URL: [https://www.researchgate.net/publication/278099739\\_Sensitivity\\_Analysis\\_of\\_Welch's\\_t-Test](https://www.researchgate.net/publication/278099739_Sensitivity_Analysis_of_Welch's_t-Test).
- [3] von Asmuth et al. *Handleiding Tijdreeksanalyse*. 978.90.5773.939.2. 2021. URL: <https://www.stowa.nl/sites/default/files/assets/PUBLICATIES/Publicaties%202021/STOWA%202021-32%20Handleiding%20tijdreeksanalyse%283%29.pdf>.
- [4] Ayvaz and Elçi. "Identification of the optimum groundwater quality monitoring network using a genetic algorithm based optimization approach". In: *Journal of Hydrology* 563.1078-1091 (July 2018). DOI: 10.1016/j.jhydrol.2018.06.006. URL: [https://www.researchgate.net/publication/325546266\\_Identification\\_of\\_the\\_optimum\\_groundwater\\_quality\\_monitoring\\_network\\_using\\_a\\_genetic\\_algorithm\\_based\\_optimization\\_approach](https://www.researchgate.net/publication/325546266_Identification_of_the_optimum_groundwater_quality_monitoring_network_using_a_genetic_algorithm_based_optimization_approach).
- [5] van Bakel et al. *Verstedelijking en verdroging*. NOV rapport 4. Mar. 1995. URL: <https://www.debakelstroom.nl/wp-content/uploads/lijst-van-publicaties-van-Jan-van-Bakel-jan-2018.pdf>.
- [6] Centraal Bureau voor de Statistiek. *Oppervlakte 2020*. 2020. URL: [https://onderzoek010.nl/jive?cat\\_open\\_code=cgdhegjtjcidphg2](https://onderzoek010.nl/jive?cat_open_code=cgdhegjtjcidphg2).
- [7] Collenteur et al. *Pastas: open source software for the analysis of groundwater time series*. Comp. software. July 2019. URL: <https://ngwa.onlinelibrary.wiley.com/doi/10.1111/gwat.12925>.
- [8] van Damme-Jongsten. "De teloorgang van natuurmoment De Beer. Een eersteklas landschap". In: (2008). URL: <https://natuurtijdschriften.nl/pub/642185/Natura2008105001006.pdf>.
- [9] Douglas. *Urban Hydrology*. 2nd ed. 2020, pp. -. URL: <https://www.taylorfrancis.com/chapters/edit/10.4324/9780429506758-16/urban-hydrology-ian-douglas?context=ubx>.
- [10] Dufour. *Hydrologische situatie en processen in West-Nederland*. 1998. URL: <https://repository.tno.nl/islandora/object/uuid%3Ad6b7add5-7b50-4949-8a78-ef42b8fc549f>.
- [11] Etikala, Madhav, and Gowd Somagouni. *Chapter 1 - Urban Water Systems: an overview*. Vol. 6. 2022, pp. 1–19. DOI: 10.1016/B978-0-323-91838-1.00016-6. URL: <https://www.sciencedirect.com/science/article/abs/pii/B9780323918381000166>.
- [12] European Commission. *Annex II to Directive 2006/118/EC of the European Parliament and the Council on the Protection of Groundwater Against Pollution and Deterioration*. OJL182. 2014.
- [13] European Environment Agency. *Europe's groundwater — a key resource under pressure*. 2022. URL: <https://www.eea.europa.eu/publications/europe-s-groundwater>.
- [14] van Ganswijk, van Claessen, and Veenbaas. *Een leidraad voor de hydrologische systeembeschrijving van natuurgebieden met enkele voorbeelden*. 1988.
- [15] Gemeente Rotterdam. *Bestemmingsplan Nieuwe dorp Heijplaat. Archeologie*. 2013. URL: [https://www.planviewer.nl/imro/files/NL.IMRO.0599.BP1011HeijNwDorp-oh01/t\\_NL.IMRO.0599.BP1011HeijNwDorp-oh01\\_4.7.html](https://www.planviewer.nl/imro/files/NL.IMRO.0599.BP1011HeijNwDorp-oh01/t_NL.IMRO.0599.BP1011HeijNwDorp-oh01_4.7.html).
- [16] Gemeente Rotterdam. *Grondwater*. URL: <https://www.rotterdam.nl/grondwater>.
- [17] Gemeente Rotterdam. *Het stedelijk watersysteem en de verantwoordelijkheden*. Oct. 2020. URL: <https://www.rotterdam.nl/grp>.

- [18] Gemeente Rotterdam. *Van buis naar buitenruimte. Gemeentelijk Rioleringsplan Rotterdam 2021-2025*. Oct. 2020. URL: [https://rotterdam.notubiz.nl/document/9422425/1/s20bb016815\\_1\\_39408\\_tds](https://rotterdam.notubiz.nl/document/9422425/1/s20bb016815_1_39408_tds).
- [19] Geul. *Actualisatie Visie achtergrondmeetnet en projectmeetnetten*. IB-2019-0114. Rotterdam, nl, June 22, 2022.
- [20] Harp. *Algorithms in Harp-DAAL*. URL: <https://dsc-spidal.github.io/harp/docs/harpdaal/algorithms/>.
- [21] Hendriks et al. *Integrale Grondwaterstudie Nederland. Module 1: Landelijke analyse*. Feb. 2023.
- [22] Hermawanto. *Genetic Algorithm for Solving Simple Mathematical Equality Problem*. Aug. 2013. URL: [https://www.researchgate.net/publication/255994405\\_Genetic\\_Algorithm\\_for\\_Solving\\_Simple\\_Mathematical\\_Equality\\_Problem](https://www.researchgate.net/publication/255994405_Genetic_Algorithm_for_Solving_Simple_Mathematical_Equality_Problem).
- [23] Hoogvliet et al. *Klimaatverandering en grondwaterbeheer stedelijk gebied*. 2020. URL: <https://www.stowa.nl/deltafacts/ruimtelijke-adaptatie/stedelijk-waterbeheer/klimaatverandering-en-grondwaterbeheer>.
- [24] Hosseini and Kerachian. *A data fusion-based methodology for optimal redesign of groundwater monitoring networks*. 2017.
- [25] Janssen, Schelberg, and van Rhijn. *Scan stadshavens Rotterdam (Bodem en RO)*. 2004-0201. Jan. 4, 2005. URL: <https://www.bodemplus.nl/onderwerpen/bodem-ondergrond/bodembeleid/praktijkvoorbeelden/bodem-ruimtelijke/scan-stadshavens/>.
- [26] Jones, Shi, and Bradley. *Schematic graph showing the difference between unconfined and confined aquifers*. 2013. URL: <https://www.twdb.texas.gov/groundwater/docs/GAMruns/Task13-031.pdf>.
- [27] Alisa Jung et al. “Detecting Bias in Monte Carlo Renderers using Welch’s t-test”. In: *Journal of Computer Graphics Techniques* 9.2 (2020). URL: <https://cg.ivd.kit.edu/publications/2020/welch/Jung2020Bias-lowres.pdf>.
- [28] Koomen and Nieuwenhuizen. *Klimaatbestendig Nederland: Systeemanalyse*. 2011.
- [29] Koshikawa et al. “Solving System of Nonlinear Equations with the Genetic Algorithm and Newton’s Method”. In: International Congress on Mathematical Software. URL: <https://av.tib.eu/media/47899>.
- [30] Manohar et al. “Data-Driven sparse sensor placement for reconstruction: demonstrating the benefits of exploiting known patterns”. In: *IEEE Journals Magazine | IEEE Xplore* (2018), pp. 63–86. DOI: 10.1109/MCS.2018.2810460. URL: <https://ieeexplore.ieee.org/document/8361090>.
- [31] Massop and van der Gaast. *Optimalisatie grondwatermeetnet Waterschap De Maaskant. Meetnet regio Oost en beschrijving grondwaterstanden periode 1990-2002*. 864. 2003.
- [32] Milieudienst Rijnmond. *Milieudienst Rijnmond*. Dec. 21, 2023. URL: <https://www.dcmr.nl/actueel/dossiers/dossier-heijplaat>.
- [33] Mirzaie-Nodoushan, Bozorg-Haddad, and Loáiciga. “Optimal design of groundwater-level monitoring networks”. In: (2017). URL: [https://www.researchgate.net/profile/Hugo-Loaiciga/publication/319063024\\_Optimal\\_design\\_of\\_groundwater\\_level\\_monitoring\\_networks/links/5aecab9baca2727bc004ecac/Optimal-design-of-groundwater-level-monitoring-networks.pdf](https://www.researchgate.net/profile/Hugo-Loaiciga/publication/319063024_Optimal_design_of_groundwater_level_monitoring_networks/links/5aecab9baca2727bc004ecac/Optimal-design-of-groundwater-level-monitoring-networks.pdf).
- [34] Ohmer, Liesch, and Wunsch. *Spatiotemporal optimization of groundwater monitoring networks using data-driven sparse sensing methods*. 15. 2022, pp. 4033–4053. URL: <https://hess.copernicus.org/articles/26/4033/2022/>.
- [35] Marc Ohmer, Tanja Liesch, and Nico Goldscheider. “On the Optimal Spatial Design for Groundwater Level Monitoring Networks”. In: *Water Resources Research* 55.11 (Nov. 1, 2019), pp. 9454–9473. DOI: 10.1029/2019wr025728. URL: <https://doi.org/10.1029/2019wr025728>.
- [36] Port of Rotterdam. *Rozenburg*. URL: <https://www.portofrotterdam.com/nl/bouwen-aan-de-haven/lopende-projecten/rozenburg>.

- [37] Port of Rotterdam, Projectbureau Stadshavens Rotterdam, and Jacco Huijsen Fotografie. *Stad-ShavenS RotteRdam*. 2008. URL: <https://www.portofrotterdam.com/sites/default/files/2021-05/rondje-heijplaat-benenwagen.pdf>.
- [38] Ritzema et al. *Meten en interpreteren van grondwaterstanden. Analyse van methodieken en nauwkeurigheid*. 2345. 2012.
- [39] Verkeer en Leefomgeving] RWS WVL [Rijkswaterstaat Water. *Handreiking juridische helderheid grondwaterbeheer*. 2012. URL: <https://www.helpdeskwater.nl/onderwerpen/wetgeving-beleid/handboek-water/thema-s/grondwater/handreiking/>.
- [40] Sebregts. *Tuindorp Heijplaat vs. Tuindorp Oostzaan. Stedelijke idealen in de havens van Rotterdam en Amsterdam*. 2019.
- [41] Siddique et al. *Hydrogeological aspects in an urban groundwater scenario*. 2017. URL: [https://www.researchgate.net/publication/337858013\\_Hydrogeological\\_aspects\\_of\\_PAQM\\_in\\_Urban\\_Groundwater\\_scenario\\_in\\_SE\\_Bengaluru-ACWADAM\\_Aug\\_2017](https://www.researchgate.net/publication/337858013_Hydrogeological_aspects_of_PAQM_in_Urban_Groundwater_scenario_in_SE_Bengaluru-ACWADAM_Aug_2017).
- [42] de Silva et al. "PySensors: A Python Package for Sparse Sensor Placement". In: *The Journal for Open Source Software* 6.58 (2021). URL: <https://arxiv.org/pdf/2102.13476.pdf>.
- [43] Elizabeth C. Smith and Stephen K. Swallow. *Lindahl pricing for public goods and experimental auctions for the environment*. Jan. 1, 2013, pp. 45–51. DOI: 10.1016/b978-0-12-375067-9.00107-8. URL: <https://doi.org/10.1016/b978-0-12-375067-9.00107-8>.
- [44] Statistics Easily. *When The P-Value is Less Than 0.05: Understanding Statistical Significance*. Feb. 12, 2024. URL: <https://statisticseasily.com/when-p-value-is-less-than-005/#:~:text=By%20setting%20the%20threshold%20at,convention%20in%20statistical%20hypothesis%20testing>.
- [45] STOWA. *Methodiek voor de evaluatie en optimalisatie van routine waterkwaliteitsmeetnetten. Deel III: Stappenplan voor meetnetoptimalisatie*. 17. 1998.
- [46] TNO Geologische Dienst Nederland. *Het voorkomen van water in de grond*. In: vol. 16. TNO, 1996, p. 46.
- [47] TNO Geologische Dienst Nederland. *Ondergrondgegevens. Data en informatie van de Nederlandse ondergrond*. 2023. URL: <https://www.dinoloket.nl/ondergrondgegevens>.
- [48] University of Regina. *Computing the QR factorization. Applied Linear Algebra*. Math 415. URL: [https://uregina.ca/~franklam/Math415/Math415\\_QR\\_Factorization.pdf](https://uregina.ca/~franklam/Math415/Math415_QR_Factorization.pdf).
- [49] Valstar et al. *Grondwatermodel Rotterdamse havengebied: Achtergrondinformatie*. 11201981-002. Mar. 1, 2019.
- [50] van de Ven, Visser, and Biron. *Ontwatering in stedelijk gebied*. 2007.
- [51] van de Visch and Bos. *Een waterveilige Rotterdamse haven*. BF4776. Feb. 2, 2022. URL: [https://www.portofrotterdam.com/sites/default/files/2022-06/adaptatiestrategie-voor-een-veilige-haven\\_0.pdf](https://www.portofrotterdam.com/sites/default/files/2022-06/adaptatiestrategie-voor-een-veilige-haven_0.pdf).
- [52] de Vries. *Stedelijk grondwaterbeheer. Het optimaliseren van het grondwatermeetnet van de gemeente Ede*. 23. Apr. 1992. URL: <https://library.wur.nl/WebQuery/wurpubs/fulltext/217870>.
- [53] Witte, Aggenbach, and Runhaar. *Grondwater voor natuur: deel II*. 607300003. 2008.
- [54] Wu. "Optimal design of a groundwater monitoring network in Daqing, China". In: *Environmental geology* 45.4 (Feb. 1, 2004), pp. 527–535. DOI: 10.1007/s00254-003-0907-x. URL: <https://doi.org/10.1007/s00254-003-0907-x>.
- [55] van 't Zelfde. "De waarde van een gemeentelijk grondwatermeetnet". In: *H2O Tijdschrift* 44.17 (2011), pp. 8–9. URL: <https://edepot.wur.nl/339546>.



# Pastas Time Series Modeling

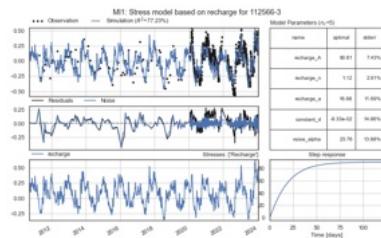
## A.1. Results of Pastas time series modeling

For stress model 1, the depicted figures explain a model summary based on the parameter *recharge rate* for the case study areas Rozenburg and Heijplaat. The figure consists of 4 subplots and a summarizing table [7]. The first plot visualizes the observed and simulated groundwater level data [m NAP] with a coefficient of determination [ $R^2$ ] that compares the simulated data with the observed data. The second plot visualizes the residual and noise series of the model. The residuals can indicate if a certain stress is missing from the model, in a situation where a clear trend of the residuals is visible. The residuals are auto-correlated with extended periods, where the modeled heads are higher or lower than the observed heads, while the noise shows little auto-correlation. The third plot, below, plots the estimated recharge over the project period [m/day]. The fourth plot, on the bottom right side of the figure, shows the estimated step response that explains the increase in recharge rate to a final level of recharge rate [mm/day]. The table provides results of the fit report that is returned by the solve method and shows information on the model settings, the fit statistics, and the estimated parameters. The solve function has a number of default options that can be specified with keyword arguments. The fit report encompasses an overview of the fitting procedure, the optimal values obtained by the fitting routine, and basic statistics.

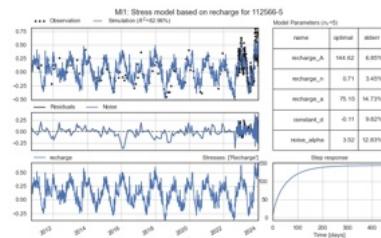
The model encloses five parameters: 1)  $A$ ; 2)  $n$ ; 3)  $a$ ; 4)  $d$ ; 5)  $\alpha$ . The first three parameters are used as a *Gamma* function as the response function for the recharge, parameter  $d$  which indicates the constant base level, and parameter  $\alpha$  of the noise model that indicates if there is a significant difference between the auto-correlation in the noise. Overall, the model simulation shows a good fit with the observed groundwater levels in the calibration period if the model is supported by a low RMSE [cm] and a high Pearson  $R^2$  value.

For stress model 1, the recharge rate was estimated as  $P$  minus  $E_a$  [7]. To improve the model,  $E_a$  as a factor is estimated, times  $E_{pot}$ . Stress model 2 encompasses the precipitation and evaporation series with an additional parameter for the evaporation factor  $f$ . Stress model 2 gives an improved representation and fit of the model [7]. Meaning that the RMSE is lowered with a higher explained variance  $R^2$ . However, according to the Akaike Information Criterion, incorporating the evaporation factor  $f$  into the model does not significantly enhance the performance of the model. The depicted figures for stress model 2 are visualized from figure A.44 on.

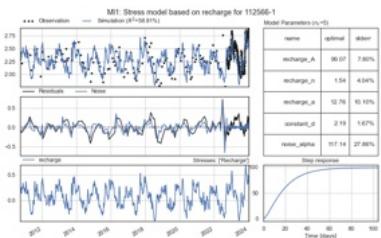
### A.1.1. Stress Model 1: Rozenburg



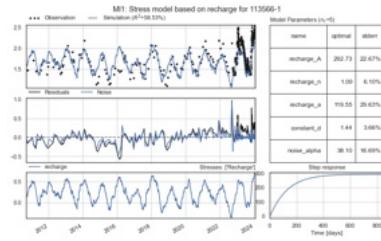
**Figure A.1:** Monitoring well: 112566-3.



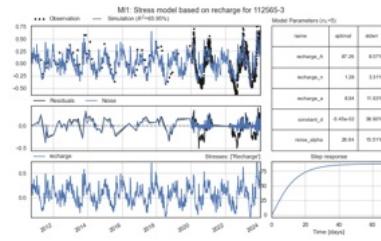
**Figure A.2:** Monitoring well: 112566-5.



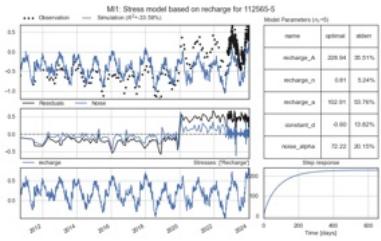
**Figure A.3:** Monitoring well: 112566-1.



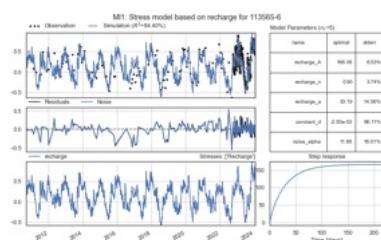
**Figure A.4:** Monitoring well: 113566-1.



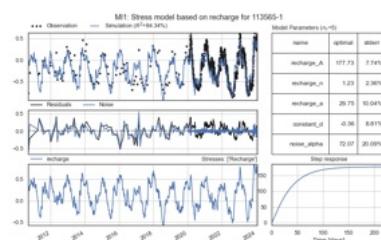
**Figure A.5:** Monitoring well: 112565-3.



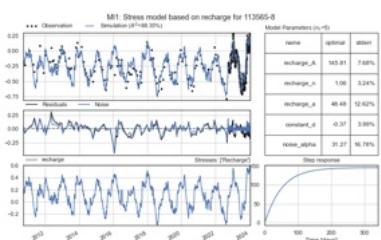
**Figure A.6:** Monitoring well: 112565-5.



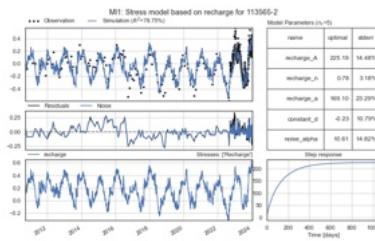
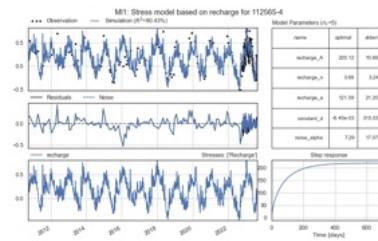
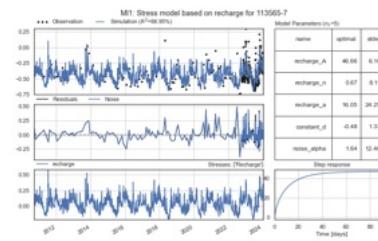
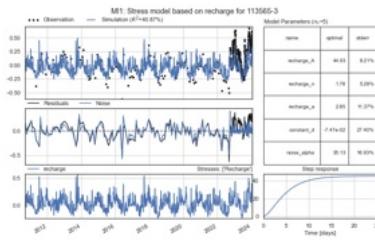
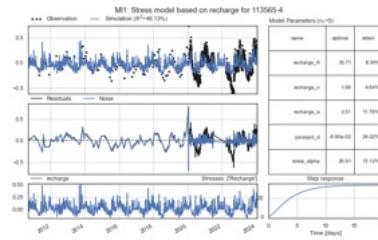
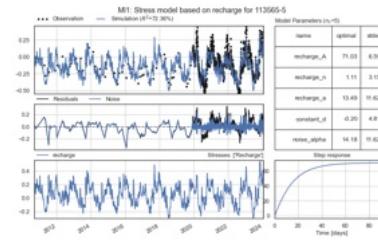
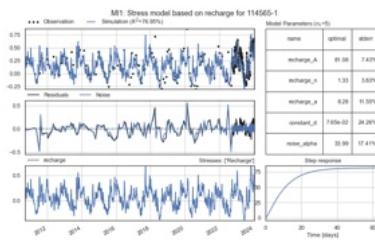
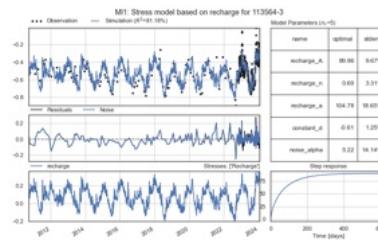
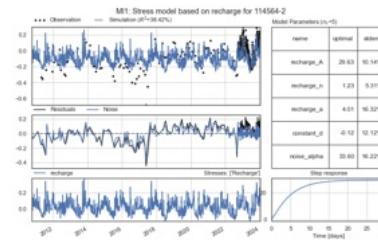
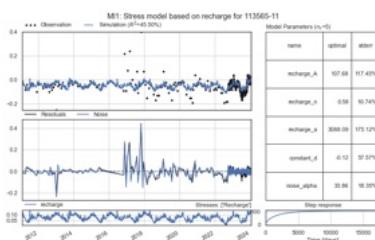
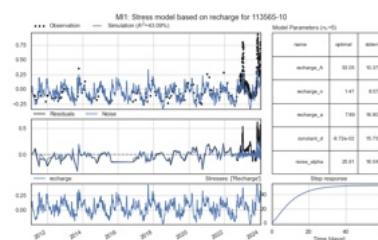
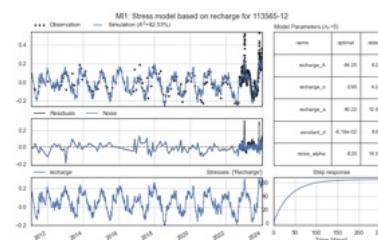
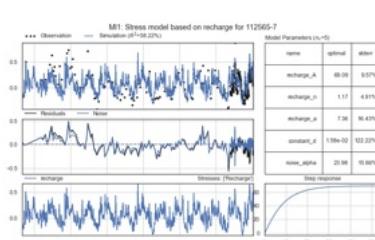
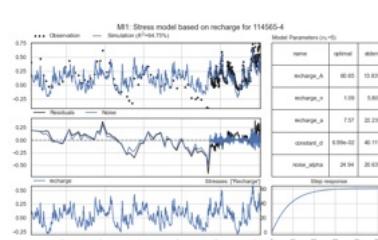
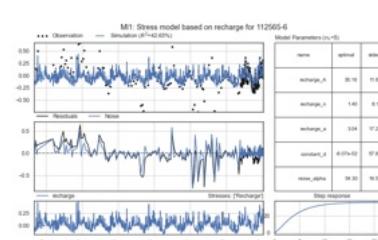
**Figure A.7:** Monitoring well: 113565-6.



**Figure A.8:** Monitoring well: 113565-1.



**Figure A.9:** Monitoring well: 112565-8.

**Figure A.10:** Monitoring well: 113565-2.**Figure A.11:** Monitoring well: 112565-4.**Figure A.12:** Monitoring well: 113565-7.**Figure A.13:** Monitoring well: 113565-3.**Figure A.14:** Monitoring well: 113565-4.**Figure A.15:** Monitoring well: 113565-5.**Figure A.16:** Monitoring well: 114565-1.**Figure A.17:** Monitoring well: 113564-3.**Figure A.18:** Monitoring well: 114564-2.**Figure A.19:** Monitoring well: 113565-11.**Figure A.20:** Monitoring well: 113565-10.**Figure A.21:** Monitoring well: 113565-12.**Figure A.22:** Monitoring well: 112565-7.**Figure A.23:** Monitoring well: 114565-4.**Figure A.24:** Monitoring well: 112565-6.

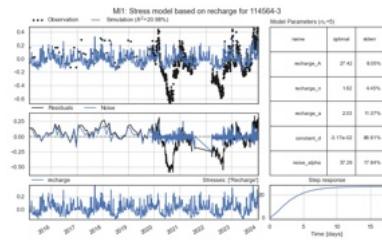


Figure A.25: Monitoring well: 1145654-3.

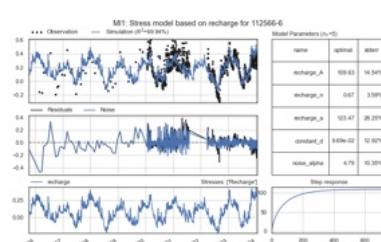


Figure A.26: Monitoring well: 112566-6.

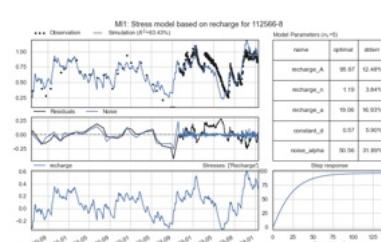


Figure A.27: Monitoring well: 112566-8.

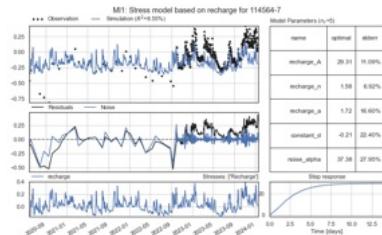


Figure A.28: Monitoring well: 114564-7.

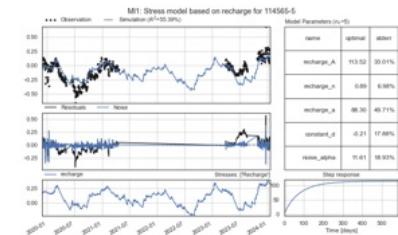


Figure A.29: Monitoring well: 114565-5.

### A.1.2. Stress Model 1: Heijplaat

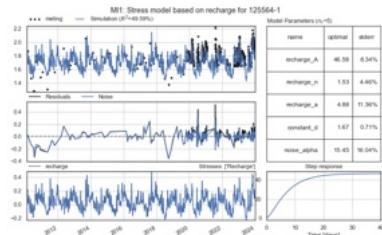


Figure A.30: Monitoring well: 125564-1.

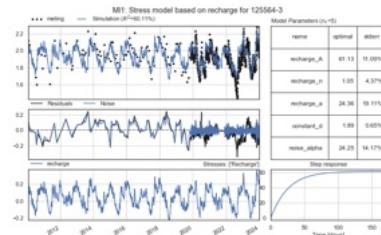


Figure A.31: Monitoring well: 125564-3.

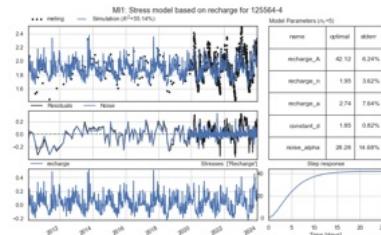
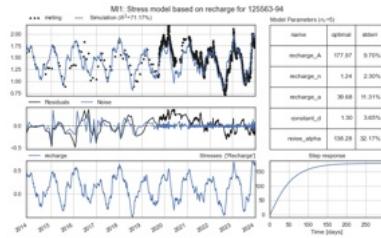
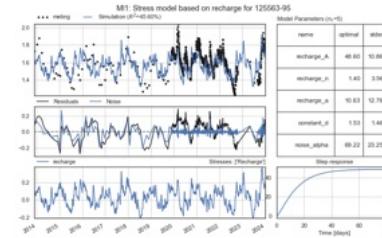


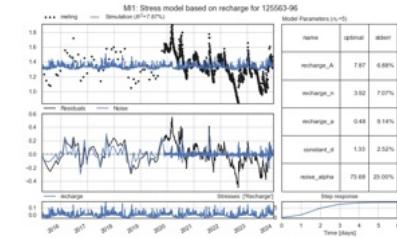
Figure A.32: Monitoring well: 125564-4.



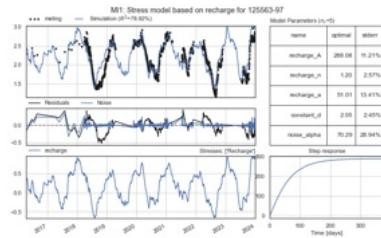
**Figure A.33:** Monitoring well:  
125563-94.



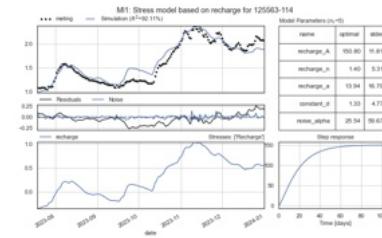
**Figure A.34:** Monitoring well:  
125563-95.



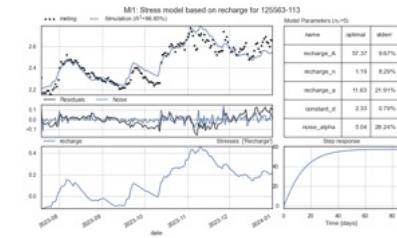
**Figure A.35:** Monitoring well:  
125563-96.



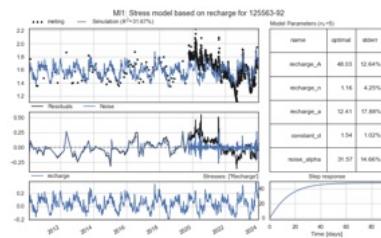
**Figure A.36:** Monitoring well:  
125563-97.



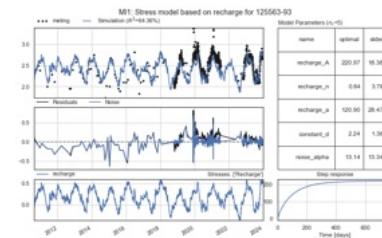
**Figure A.37:** Verander: Monitoring well:  
125564-114.



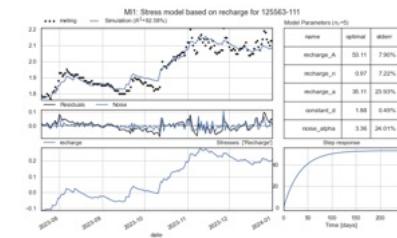
**Figure A.38:** Monitoring well:  
125563-113.



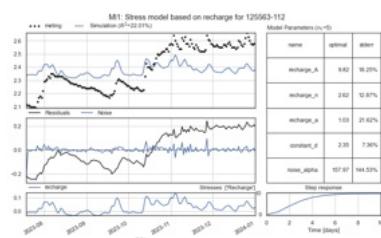
**Figure A.39:** Monitoring well:  
125563-92.



**Figure A.40:** Monitoring well:  
125563-93.



**Figure A.41:** Monitoring well:  
125563-111.



**Figure A.42:** Monitoring well:  
125563-112.



**Figure A.43:** Monitoring well:  
125563-115.

### A.1.3. Stress Model 2: Rozenburg

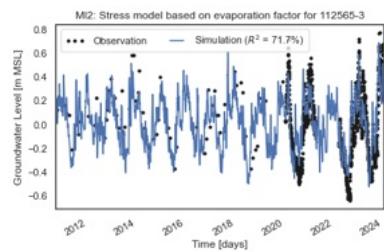


Figure A.44: Monitoring well: 112565-3.

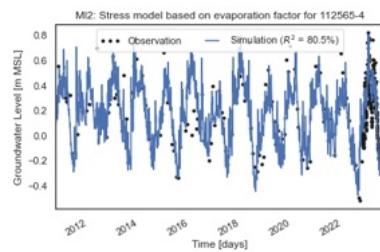


Figure A.45: Monitoring well: 112565-4.

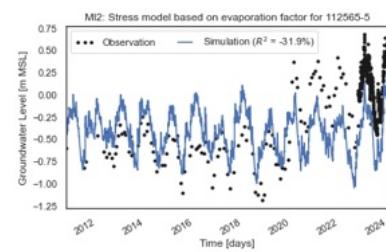


Figure A.46: Monitoring well: 112565-5.

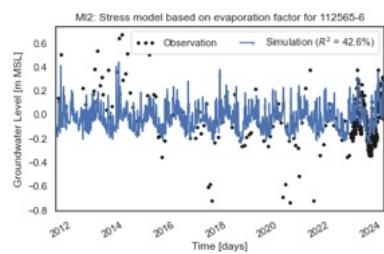


Figure A.47: Monitoring well: 112565-6.

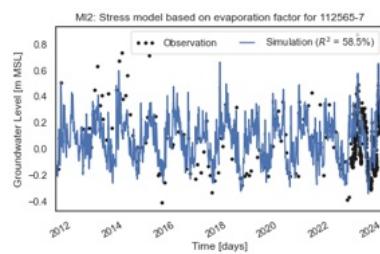


Figure A.48: Monitoring well: 112565-7

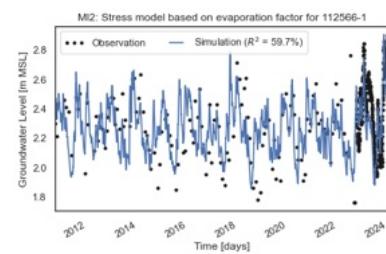


Figure A.49: Monitoring well: 112566-1.

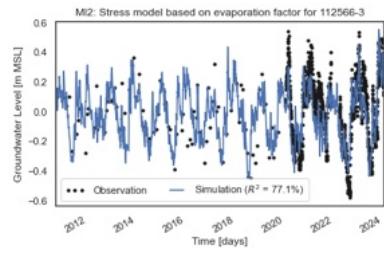


Figure A.50: Monitoring well: 112566-3.

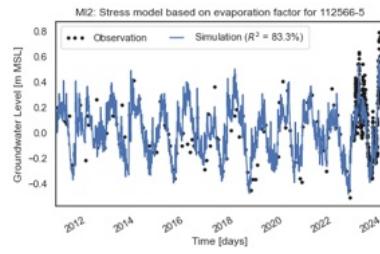


Figure A.51: Monitoring well: 112566-5.

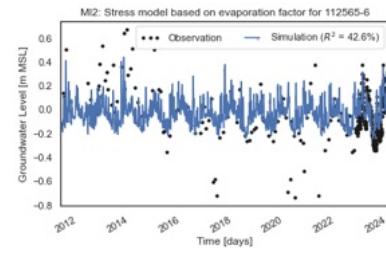
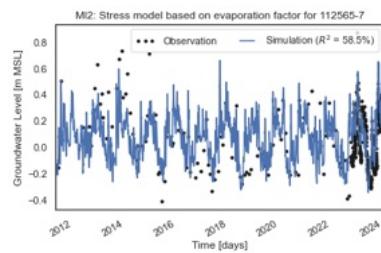
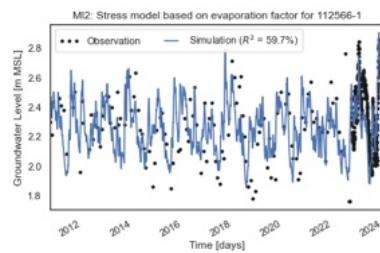
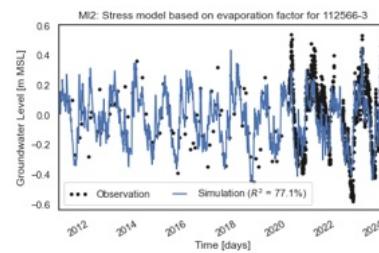
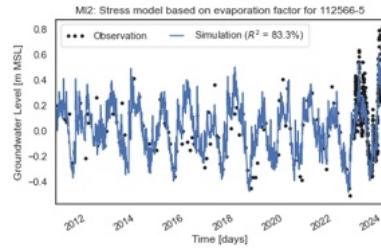
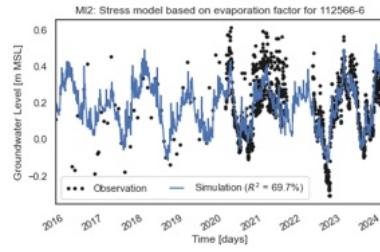
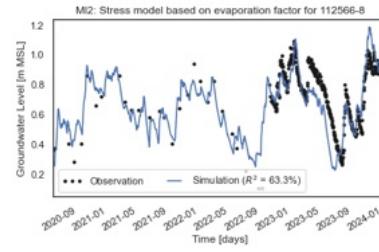
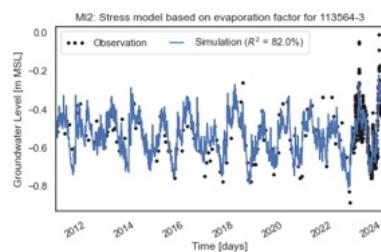
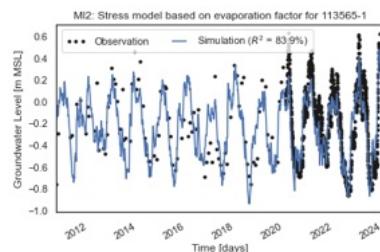
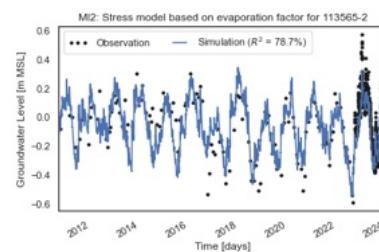
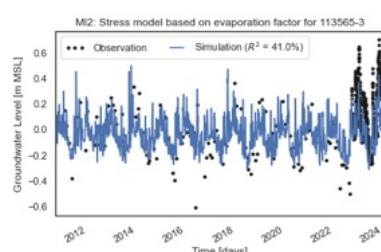
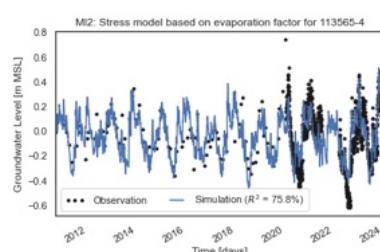
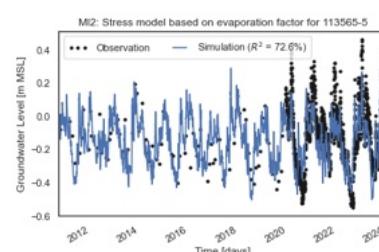
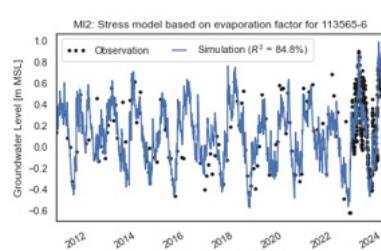
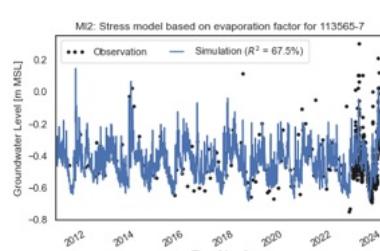
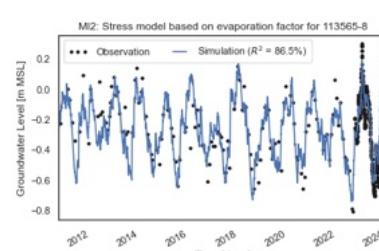
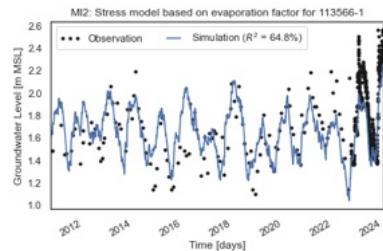
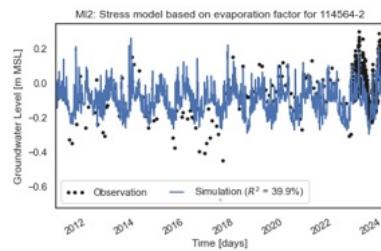
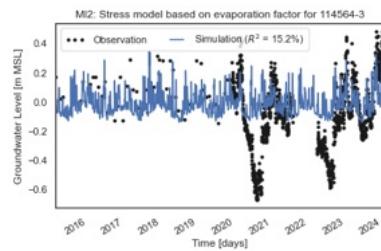
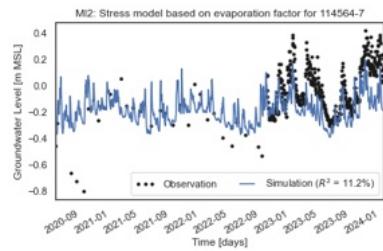
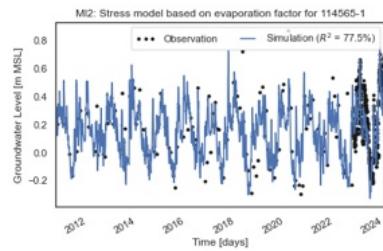
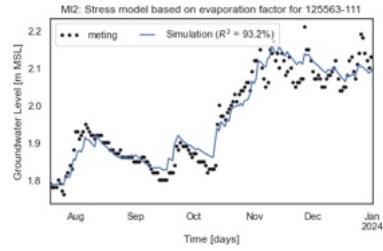
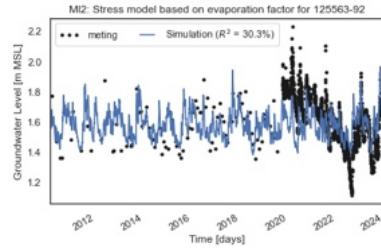
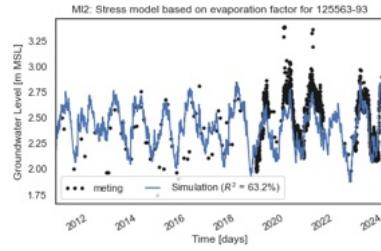


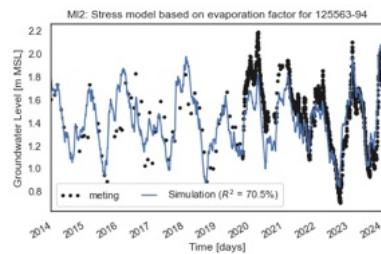
Figure A.52: Monitoring well: 112565-6.

**Figure A.53:** Monitoring well: 112565-7.**Figure A.54:** Monitoring well: 112566-1.**Figure A.55:** Monitoring well: 112566-3.**Figure A.56:** Monitoring well: 112566-5.**Figure A.57:** Monitoring well: 112566-6.**Figure A.58:** Monitoring well: 112566-8.**Figure A.59:** Monitoring well: 113564-3.**Figure A.60:** Monitoring well: 113565-1.**Figure A.61:** Monitoring well: 113565-2.**Figure A.62:** Monitoring well: 113565-3.**Figure A.63:** Monitoring well: 113565-4.**Figure A.64:** Monitoring well: 113565-5.**Figure A.65:** Monitoring well: 113565-6.**Figure A.66:** Monitoring well: 113565-7.**Figure A.67:** Monitoring well: 113565-8.

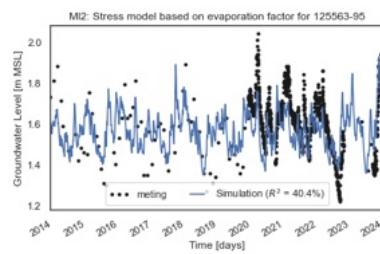
**Figure A.68:** Monitoring well: 113566-1.**Figure A.69:** Monitoring well: 114564-2.**Figure A.70:** Monitoring well: 114564-3.**Figure A.71:** Monitoring well: 114564-7.**Figure A.72:** Monitoring well: 114565-1.

#### A.1.4. Stress Model 2: Heijplaat

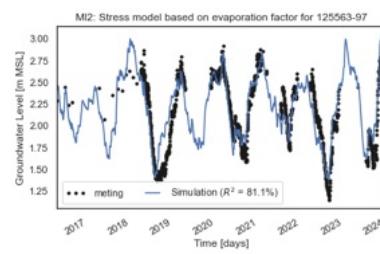
**Figure A.73:** Monitoring well: 125563-111.**Figure A.74:** Monitoring well: 125563-92.**Figure A.75:** Monitoring well: 125563-93.



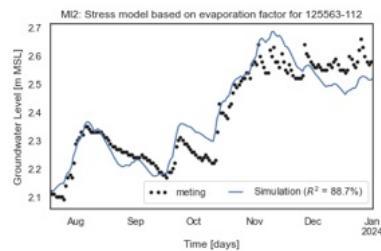
**Figure A.76:** Monitoring well: 125563-94.



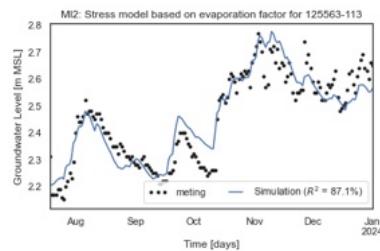
**Figure A.77:** Monitoring well: 125563-95.



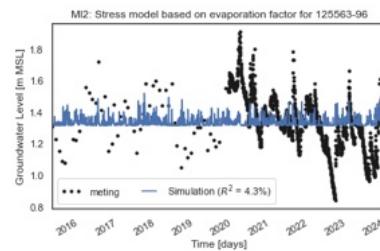
**Figure A.78:** Monitoring well: 125563-97.



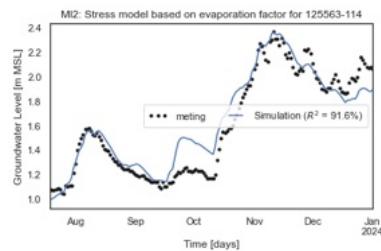
**Figure A.79:** Monitoring well: 125563-112.



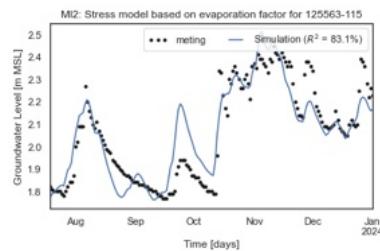
**Figure A.80:** Monitoring well: 125563-113.



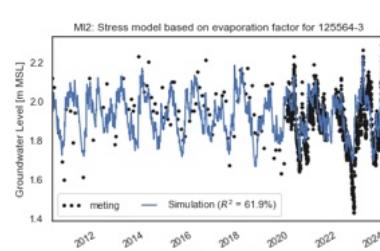
**Figure A.81:** Monitoring well: 125563-96.



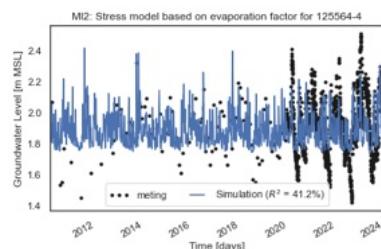
**Figure A.82:** Monitoring well: 125563-114.



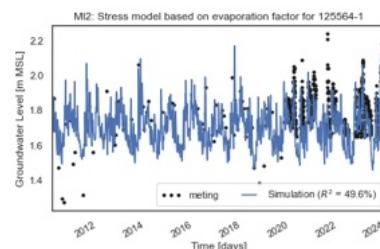
**Figure A.83:** Monitoring well: 125563-115.



**Figure A.84:** Monitoring well: 125564-3.



**Figure A.85:** Monitoring well: 125564-4.



**Figure A.86:** Monitoring well: 125564-1.

# B

## Mathematical Background

*Appendix B displays the mathematical background of the used Python packages and statistics.*

### B.1. Python Packages

#### B.1.1. Pastas

Pastas is a Python package made to analyze a sequence of data points that are collected over specific intervals of time [7]. Data gaps are present in the dataset, resulting in potential problems later on in the development of the model. A first step is to execute a system analysis, where possible hydrological variables are determined. Continuing with the model construction, it is determined how variables are transferred to fluctuations. In the end, the model has to be approved to see which variables have actual influence on model performance.

The time series model can be described in the most general way:

$$h(t) = \sum_{m=1}^M h_m(t) + d + r(t)$$

**Figure B.1:** Formula for a basic time series model [3].

$H$  displays the groundwater level,  $h_m(t)$  is the contribution of influence  $m$ ,  $d$  is the base level of the model, and  $r$  are the model residues. The total influence that contributes to the groundwater level fluctuations is  $M$  [3].

#### B.1.2. PySensors

PySensors is a Python package for scalable optimization of monitoring well placement from observed data. The package provides a tool for sparse sensor placement optimization approach that ensures a data-driven dimensionality reduction [30]. The developed method ensures near optimal placement for decision-making tasks, like the placement of monitoring wells. An advantage is that the method can be adapted to be suitable for different optimization algorithms and objective functions. Generally, the PySensors package is applied for reconstruction measures, to recover a high dimensional signal from a limited number of measurements. The foundation of the package computes data by using powerful dimensionality reduction techniques as the principal component analysis (PCA) and random projections. PySensors can be used for classification purposes by implementing the Sparse Sensor Placement Optimization for Classification algorithm (SSPOC). This algorithm provides a compressed sensing optimization that makes it possible to allow one to optimize sensor placement for classification

accuracy, but can also identify the sparsest set of sensors that reconstructs a ‘discriminating plane’ in a feature subspace. The implementation of the SSPOC algorithm is a general algorithm that can be used simultaneously with a linear classifier (an algorithm that separates two types of objects by a line or hyperplane) [30]. The PySensors package can also implement the Sparse Sensor Placement Optimization for Reconstruction method for recovering high-dimensional signals ( $x$ ) from linear sensor measurements in the form:

$$y = Cx$$

The optimal measurements of  $x$  are identified and given by operator  $C$ , which describes the components of  $x$  to observe. The SSPOR method has its purpose to find the best subset of available measurements from which the full signal can be received in the estimation problem. According to De

$$\mathbf{C}^* = \arg \min_{\mathbf{C} \in \mathbb{R}^{p \times N}} \|\mathbf{x} - \Phi(\mathbf{C}\Phi)^\dagger \mathbf{y}\|_2^2,$$

**Figure B.2:** Formula of the Sparse Sensor Placement Optimization for Reconstruction algorithm [30].

Silva, PySensors provides methods to enable straightforward exploration of the impacts of critical hyper parameters, such as the number of sensors or basic modes [42]. In this case, those are the number of monitoring wells.

## B.2. Statistics

### B.2.1. Welch's t-Test

In a Welch's t-Test, two data groups can be compared within the dataset without assuming equal data variances of both data groups. A t-Test is carried out to criticize whether a dataset with a combination of measurement types as data logger and manual collection methods are recommended to be used in the model. A Welch's t-Test is a suitable approach, because the denominator of the formula provides the possibility that the two data groups have unequal variances. The t-Test is determined by the mean of the two data groups, the variances of both data groups, and the sample size of the two data groups. Since a difference is detected between the data quantity of the data loggers and the manual collection method, a Welch's t-Test is a suitable measure to determine if there is actually a significant difference between the two data groups. The t-Test is based on the following formula [27], see figure B.3. Where  $t$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where  $\bar{x}_1, \bar{x}_2$  = mean of sample  $i$ ,  
 $s_1^2, s_2^2$  = variance of sample  $i$ ,  
 $n_1, n_2$  = size of sample  $i$ , and  $i = 1, 2$ .

**Figure B.3:** Formula to calculate  $t$  in the Welch's t-test [27].

is a combination of the properties that are combined into the t-statistic, a random variable.  $X1$  and  $X2$ , which can be explained as the mean of the sample  $i$ , in this case the mean of the groundwater level of the data loggers and the manual collection method.  $U1$  and  $U2$  can be explained as the population means, the theoretical means, which are assumed from the samples that are taken. The denominator can be described as the root of variance of the data loggers divided by the sample size of the data loggers plus the variance of the mean collection method divided by the sample size of the mean collection method [27]. In this context an  $\alpha = 0.05$  is chosen, where a  $P - value < 0.05$  indicates evidence against the null hypothesis and suggests if there is a statistically significant relationship. The alpha

threshold is arbitrary, meaning that a higher or lower alpha than 0.05 possibly depends on the context. An alpha of 0.05 manages the risk of rejecting the null hypothesis, while maintaining sufficient power to detect accurate results. The result of the t-Test, the p-value, generally does not provide much information about the magnitude or importance of the observed effect. If a p-value is lower than 0.05, the result suggests that the observed data does not provide sufficient evidence against the null hypothesis and indicates a statistically significant effect between the researched variables [44].

### B.2.2. QR factorization

QR Factorization is a mathematical approach that breaks down matrix  $A$  into the product of two matrices:  $Q$  and  $R$ . Matrix  $A$  is an  $m * n$  matrix with linearly independent columns. The Gram-Schmidt orthogonalization process results in producing  $Q$ , with orthonormal columns, and  $R$ , an upper triangular matrix with positive entries on the diagonal.  $Q$  transforms the original dataset into a new dataset with independent columns. This action is essential for evaluating the unique contribution of each monitoring well. The triangular matrix of  $R$  serves to rank monitoring wells based on their informational relevance [48]. Figure B.4 displays the decomposition of matrix  $A$  into matrix  $Q$  and  $R$  [20].

The diagram illustrates the QR factorization of a matrix  $A$  into an orthogonal matrix  $Q$  and an upper triangular matrix  $R$ . On the left, matrix  $A$  is shown as a vertical column of three vectors  $a_1, a_2, a_3$ . To its right is the equation  $=$ . To the right of the equation is matrix  $Q$ , which consists of three orthogonal unit vectors  $e_1, e_2, e_3$ . Below  $Q$  is a brace indicating it is an orthogonal unit vector. To the right of  $Q$  is matrix  $R$ , which is an upper triangular matrix with entries  $e_1^T \cdot a_1, e_1^T \cdot a_2, e_1^T \cdot a_3$  in the first row;  $0, e_2^T \cdot a_2, e_2^T \cdot a_3$  in the second row; and  $0, 0, e_3^T \cdot a_3$  in the third row. Below  $R$  is a brace indicating it is an upper diagonal matrix.

**Figure B.4:** QR factorization or decomposition of the matrix  $A$  into an orthogonal matrix  $Q$  and triangular matrix  $R$  [20].

### B.2.3. Genetic Algorithm

A genetic algorithm (GA) is based on decision variables that are encoded as binary strings, genes, for a given location - in this case the municipal area of Rotterdam. Combinations of the genes generate chromosomes which correspond to possible monitoring networks. Random chromosomes evolve to network generations. In each generation, the fitness of the “population” is evaluated. Chromosomes are selected based on their fitness on the current population and modified using genetic operators. The approach is based on ‘survival-of-the-fittest’ [22]. The network of monitoring wells corresponds to ‘chromosomes’ and each monitoring well in the network is represented by a binary bit that determines if the well will be selected for the network or not (1 = yes, 0 = no). Every unique monitoring well is evaluated by the fitness of the network by an inverse distance and interpolation approach. GA are stochastic optimization tools that work on Darwinian models and are capable of solving near-optimal solutions for multi variable functions without the usual mathematical requirements [4].

$$f(r) = \sum_{i=1}^n \sum_{j=i+1}^n \{(\bar{r}_{a_i, a_j} - \hat{r}_{a_i, a_j})^2 + (\bar{r}_{a_i, b_j} - \hat{r}_{a_i, b_j})^2 + (\bar{r}_{b_i, a_j} - \hat{r}_{b_i, a_j})^2 + (\bar{r}_{b_i, b_j} - \hat{r}_{b_i, b_j})^2\}$$

**Figure B.5:** Formula to determine the genetic algorithm in the optimization model [29].