Improved Trend Evaluation and Workflow Optimisation of Groundwater Monitoring Programmes

Alternate Approaches for Small Chlorinated-Solvent Datasets

Master's thesis in the Master's Programme Infrastructure and Environmental Engineering

MATILDA SAVOLAINEN

LEO ÅKESSON

DEPARTMENT OF ARCHITECTURE AND CIVIL ENGINEERING
DIVISION OF GEOLOGY AND GEOTECHNICS

CHALMERS UNIVERSITY OF TECHNOLOGY
Master's thesis ACEX30
Gothenburg, Sweden 2026

MASTER'S THESIS ACEX30

Improved Trend Evaluation and Workflow Optimisation of Groundwater Monitoring Programmes

Alternate Approaches for Small Chlorinated-Solvent Datasets

Master's Thesis in the Master's Programme Infrastructure and Environmental Engineering

MATILDA SAVOLAINEN
LEO ÅKESSON

Department of Architecture and Civil Engineering
Division of Geology and Geotechnics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026

Improved Trend Evaluation and Workflow Optimisation of Groundwater Monitoring Programmes
Alternate Approaches for Small Chlorinated-Solvent Datasets
Master's Thesis in the Master's Programme Infrastructure and Environmental Engineering

MATILDA SAVOLAINEN
LEO ÅKESSON

Supervisor: Lena Torin, WSP
Supervisor: Jenny Norrman, Division of Geology and Geotechnics at Chalmers University of Techonolgy
Examiner: Lars Rosen, Division of Geology and Geotechnics at Chalmers University of Techonolgy

Examensarbete ACEX30
Institutionen för Arkitektur och Samhällsbyggnadsteknik
Chalmers Tekniska Högskola, 2026

Department of Architecture and Civil Engineering
Division of Geology and Geotechnics
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone +46 31 772 1000

Improved Trend Evaluation and Workflow Optimisation of Groundwater Monitoring Programmes
Alternate Approaches for Small Chlorinated-Solvent Datasets

Master's thesis in the Master's Programme Infrastructure and Environmental Engineering

MATILDA SAVOLAINEN

LEO ÅKESSON

Department of Architecture and Civil Engineering
Division of Geology and Geotechnics
Chalmers University of Technology

ABSTRACT

Groundwater monitoring programmes (GMP) for sites contaminated with chlorinated solvents often rely on qualitative assessments that rely on the judgement of the evaluator. However, the datasets encountered in Sweden are often small, typically fewer than 20 wells, and present challenges such as skewed distributions, irregular sampling intervals, and limited spatial coverage, which could reduce the reliability of traditional trend evaluation methods. This thesis investigates alternative and complementary analytical approaches that can strengthen the understanding of chlorinated solvent behaviour and movements over time in groundwater.

The results demonstrate that small datasets require analytical approaches that go beyond qualitative interpretation. Statistical characterisation confirmed pronounced skewness, heteroskedasticity, and non-normality, supporting the use of non-parametric methods such as the Mann-Kendall test and Sen's slope. Importantly, visual trend assessments showed substantial uncertainty, with only a minority of wells classified with high confidence, underscoring the limitations of qualitative judgement alone. Site-wide plume evaluation using the Ricker Method revealed strong sensitivity to interpolation choice and well placement, while scenario testing highlighted how removing key wells can impair plume mass estimation.

Although no single software tested fully meets GMP needs, GWSDAT offers strong visual and spatial functions but limited statistical depth, ProUCL provides the most robust statistics, and AquaChem is broad but time-intensive. GWSDAT ranked highest overall, but with results being sensitive to user preferences.

Overall, the findings show that reliable interpretation of small chlorinated solvent datasets requires combining multiple analytical perspectives. A workflow integrating non-parametric trend testing, individual constituent evaluation, and plume mass investigation can substantially improve decision making regarding sampling frequency, plume stability, and long-term monitoring optimisation.

Key words: Groundwater Monitoring, Contamination, Trend Analysis, Chloroethenes, Small Datasets, Optimisation, Software.

Contents

Preface

In this study, an evaluation of methods and software tools for improving the interpretation of groundwater monitoring data at chlorinated solvents contaminated sites was conducted throughout January 2026 to June 2026. The project has been conducted at the Department of Architecture and Civil Engineering, Engineering Geology, at Chalmers University of Technology in collaboration with WSP Sweden.

The work was supervised by Jenny Norrman and examined by Lars Rosén at Chalmers University of Technology, with Lena Torin serving as supervisor from WSP Sweden.

We would like to express our sincere gratitude to Lena Torin for her expertise, encouragement, and enthusiasm throughout the work, as well as to the entire Contaminated Sites department at WSP for their support and valuable discussions. We also extend our warm thanks to Jenny Norrman for her exceptional guidance during the thesis, and to Lars Rosén for the constructive advice and academic insight we have received. Furthermore, we would like to thank Joe Ricker at WSP USA for generously providing personal guidance on his method, which has been highly valuable for the development of this thesis.

We would also like to thank the Chalmers Writing Centre for their support with thesis structure and academic writing, and Aila Särkkä, Head of Division and Full Professor in Applied Mathematics and Statistics, Department of Mathematical Sciences, Chalmers University of Technology, for her valuable help in addressing questions related to statistical trend tests.

Gothenburg, June 2026
Matilda Savolainen & Leo Åkesson

Notations

Roman upper case letters

CV	Coefficient of variation
H_0	Null hypothesis (no monotonic trend present)
H_a	Alternative hypothesis (monotonic trend present)
M_{x}	Molecular weight of X [μg/mol]
N	Number of unique observation pairs in Theil–Sen estimator
N	Dataset size (number of samples) used in RMSE calculation
Q_i	Individual Theil–Sen slope estimate
$RMSE$	Root mean square error for cross-validation
S	Mann–Kendall test statistic
S_{base}	Original weighted score for each software (MCA)
S^+	Increased weighted score after +10% weight perturbation (MCA)
S^-	Decreased weighted score after –10% weight perturbation (MCA)
W_i	Weight assigned to neighbouring node i in kriging interpolation
W_i	Weight of criterion i in the multi-criteria analysis (MCA)
Z_A	Estimated concentration at location A (kriging)
Z_i	Measured concentration at point i

Roman lower case letters

e	Residual (difference between observed and modelled value)
i	Index of summation in RMSE calculation
n	Number of neighbouring nodes included in kriging interpolation
n	Number of observations in a time series
$n_{\text{predicted}}$	Predicted value used in RMSE computation
n_{actual}	Actual measured value used in RMSE computation
x_i, x_j	Observed concentrations at times i and j

Greek letters

α	Significance level
β	Regression coefficient (slope or intercept)
μ	Mean concentration of a dataset
σ	Standard deviation of a dataset

Abbreviation

DCE	Dichloroethene
DNAPL	Dense Non-Aqueous Phase Liquid
GIS	Geographic Information System
GMP	Groundwater Monitoring Programme
GUI	Graphical User Interface
GWSDAT	Groundwater Spatiotemporal Data Analysis Tool
MCA	Multi-Criteria Analysis
ND	Non-Detect
OLS	Ordinary Least Squares Regression
PCE	Tetrachloroethene
PCE-eq	Tetrachloroethene-Equivalent
SGU	Sveriges Geologiska Undersökning
TCE	Trichloroethene
US EPA	United States Environmental Protection Agency
VC	Vinyl Chloride (Chloroethene)

1 Introduction

The number of contaminated sites in Sweden is substantial. Approximately 1300 sites are classified as risk class 1, indicating a potentially *very* large risk to human health and the environment, while around 14000 sites fall under risk class 2, assessed as having potentially large risks (the key difference being the level of 'large risk') (Naturvårdsverket, 2016). Additionally, there are about 85000 locations that are currently or are known to have conducted environmentally hazardous activities, and can therefore pose a threat to surrounding areas (Naturvårdsverket, 2025). In urban areas, some of the most concerning contaminants include hydrocarbon compounds, chlorinated solvents, and various metals.

Chlorinated organic compounds are of particular concern due to their persistence, mobility, and the potential health risks they pose to groundwater users living near contaminated areas (Pino et al., 2026). Among these, chlorinated solvents such as perchloroethylene (PCE; IUPAC: tetrachloroethene (European Chemicals Agency, n.d.-a)) and trichloroethylene (TCE; IUPAC: trichloroethene (European Chemicals Agency, n.d.-b)) have been widely used historically and remain in limited use today. Their primary applications include metal degreasing and dry-cleaning operations (Pino et al., 2026; Föreningen, 2011). In Sweden, TCE has been the most commonly used solvent in industrial settings, whereas PCE has been the dominant solvent in dry-cleaning facilities (Föreningen, 2011). These compounds are excellent solvents, chemically stable, non-flammable, and highly effective at dissolving grease, properties that made them attractive in industrial contexts but problematic from an environmental perspective (Föreningen, 2011).

Because of these characteristics, PCE and TCE frequently occur as contaminants at sites where they have been handled. Their subsurface behaviour is complex: as dense non-aqueous phase liquids (DNAPLs), in free phase form they are denser than water and can migrate downward through the soil profile until they encounter low-permeability layers. Their relatively low viscosity and high penetration ability allow them to move deeply into the subsurface, forming pools and residual saturation zones that act as long-term contaminant sources (Föreningen, 2011). The interaction between DNAPL properties and heterogeneous geological conditions makes plume development difficult to predict.

This property is especially pronounced for sites contaminated with chlorinated solvents, where remediation is complicated due to back-diffusion: contaminants stored in low-permeability zones such as clay or silt slowly diffuse back into more permeable aquifer layers once concentrations there are reduced (Ding et al., 2025). Back diffusion can sustain groundwater contamination for decades or even centuries, meaning that monitoring programmes can remain active for long periods to track plume behaviour and verify remediation progress. As a result, GMP becomes a substantial and ongoing financial burden (McCarty, 2010). Such monitoring ensures that unexpected contaminant behaviour does not pose renewed risks to human health or the environment, particularly when remediation strategies focus on containment or immobilisation rather than complete removal (Denham et al., 2020). In some cases, money spent on monitoring yields more information than is necessary to make decisions about the operation of the remedy or the progress toward closeout (U.S. Environmental Protection Agency, 2005).

Therefore, optimisation provides an opportunity to improve the cost-effectiveness of monitoring programmes by ensuring that data collection meets its objectives with an appropriate level of effort.

Groundwater monitoring programmes (GMP) optimisation methods range from qualitative to quantitative (U.S. Environmental Protection Agency, 2005). Historically, GMP have largely relied on qualitative assessments and professional judgement to interpret hydrologic conditions (U.S. Environmental Protection Agency, 2005). Today, Excel is commonly used to visualise trends in monitoring data, which is then combined with engineering experience to determine sampling frequency. While this approach is familiar and accessible, more advanced tools exist that can provide stronger mathematical support and broader analytical capabilities. Ling et al. (2005) highlights several studies aimed at optimising GMP networks. However, many of these approaches are mathematically complex and require considerable expertise, making them unlikely to be widely adopted by technical or regulatory professionals. A limitation of temporal analysis is that monitoring data often exhibit substantial short-term noise, meaning that long time series are required to detect statistically robust trends. However, temporal methods are not the only aspect of programme design that constrains interpretability. A further limitation is that spatiotemporal analytical approaches require a large number of monitoring wells to establish a reliable spatial correlation structure. When the number of wells is small (fewer than 20–30), deriving a dependable semivariogram becomes difficult, reducing the effectiveness of spatial or spatiotemporal evaluations. Together, these constraints show that both sampling frequency and network size fundamentally shape which analytical methods can be applied and how confidently trends can be interpreted.

This thesis work will focus on proposing methods and tools that enhance the interpretation of monitoring data and support decision-making regarding e.g. sampling frequency in GMP at contaminated sites, particularly in an occupational context. This work draws on several existing groundwater monitoring datasets provided by WSP. A typical WSP monitoring programme includes no more than 20 monitoring wells and max 30 sampling points varying sampling frequency. Such programmes are relatively limited in scope, and a key challenge will be to identify tools that can be effectively applied to these limited datasets.

1.1 Aim

The overall aim of this thesis is to suggest methods and tools that can improve the interpretation of the collected monitoring data and provide decision support regarding frequency of sampling and trend detection for groundwater monitoring programmes at contaminated sites. This will be achieved by first identifying common approaches for evaluating monitoring programme data using both graphical and statistical techniques. Subsequently, these will be tested for chlorinated solvents and compared from both a technical and a user-friendliness perspective.

Research Questions

- What are common or alternative methods for evaluating how contaminant concentrations change over time in monitoring programmes?
- How effectively can the methods support decisions on contamination behaviour for monitoring programmes of limited size?

- Which software can be of most assistance for professionals of contaminated site monitoring programmes when evaluating collected data?

1.2 Limitations

This thesis evaluates methods and not specific sites. It is not a case-study, but rather uses real monitoring data to illustrate method performance under realistic conditions. Consequently, it does not provide decisions or recommendations for how a particular monitoring programme should be managed.

The analysis is based on chlorinated-solvent datasets provided by WSP. The type of projects they handle and their standard project design could be distinct from other Swedish companies. Overall, the understanding is that these datasets have common traits with most sites in the country, and therefore should provide a satisfactory representation of generic situations.

To produce meaningful results from time series evaluations, only monitoring wells with at least 3 sampling dates were included. Also, wells with significantly deeper sampling were excluded to ensure elevation homogeneity within each location. The original contaminant-substance that has historically been used is PCE for five out of the six locations, but the sum-representative value PCE-equivalent was used for all sites for a more coherent structure.

For the spatial gridding required for the Ricker Method, only dates that sampled 3 or more wells were included since this is required for interpolation. Furthermore, these tests were only carried out using one software, which means that all belonging evaluations are affected by its computational capabilities. The number of interpolation methods and scenarios tested for the Ricker Method was limited due to time constraints; additional tests would have provided a basis for a more thorough discussion. The choice of wells for the scenarios was supported by supervisor insights, but other options could also have been justified.

The desirable functional capabilities and multi-criteria analysis input of software rely solely on feedback from WSP employees, which may not reflect broader user experiences. Similarly, the visual assessment of trend detection was conducted by the authors, and although motivated still includes a limited viewpoint.

The software discussed in this thesis focus on methods for optimising monitoring frequency of wells, with an emphasis on trend-analysis and data-review tools. Other categories of methods, such as groundwater-modelling software, risk-assessment tools, or cost-benefit analysis frameworks are outside the scope of this work. This project evaluates the statistical performance and functionality of trend-testing software. It does not include site-specific characterisation or environmental interpretation, as these are outside the defined scope.

Due to time constraints, the search for candidate software was limited to a defined search period of 3 weeks. All software identified within this period was screened, and the 3 most suitable tools were selected for testing. Price is an additional limitation: commercial software was not excluded, but budget constraints may influence the final selection.

2 Theory

This thesis discusses several chemical and mathematical concepts that require some level of understanding in order to grasp the context in which they are mentioned. The following chapter explains relevant methods and theory at a reasonably concise level.

2.1 Chlorinated Solvents

Chlorinated solvents are organic compounds commonly grouped into three categories based on their structural characteristics: chlorinated methanes, chlorinated ethanes, and chlorinated ethenes (Cwiertny & Scherer, 2010).

For this thesis, the focus is limited to chlorinated ethenes (Table 2.1). Important examples that frequently occur as groundwater contaminants include tetrachloroethene (PCE), trichloroethene (TCE), the dichloroethene isomers (collectively DCE), and the monochlorinated species vinyl chloride (VC) (Cwiertny & Scherer, 2010).

Table 2.1: Chlorinated ethenes and common abbreviations (Cwiertny & Scherer, 2010).

Compound	Common Name	Abbreviation	Formula
Tetrachloroethene	Perchloroethene	PCE	C_2Cl_4
Trichloroethene	–	TCE	C_2HCl_3
cis-1,2-Dichloroethene	cis-Dichloroethene	cis-DCE	$C_2H_2Cl_2$
trans-1,2-Dichloroethene	trans-Dichloroethene	trans-DCE	$C_2H_2Cl_2$
Chloroethene	Vinyl chloride	VC	C_2H_3Cl

Knowledge is also required of natural attenuation, which is the sum of processes leading to a decrease in pollutant concentrations at a contaminated site. Understanding the degradation pathways of chlorinated ethanes and ethenes in water is crucial in the context of implementing or not implementing remediation, and in the former case, which process to choose (Tobiszewski & Namieśnik, 2012).

2.1.1 Biodegradation

At many contaminated sites, PCE is the primary source of chlorinated ethenes, largely due to its historical use in dry cleaning and metal degreasing. Once released into the subsurface, PCE undergoes sequential reductive dechlorination, is transformed to TCE, then to cis-1,2-DCE, and finally to VC, before complete reduction to ethene. Although TCE can degrade to any of the three DCE isomers, cis-1,2-DCE is the predominant isomer formed under typical biodegradation conditions (Ricker et al., 2020).

PCE and TCE can be reduced to DCE under mildly reducing conditions, but further reduction to VC and ethene requires stronger reducing environments. The ability of chlorinated ethenes to serve as electron acceptors increases with the number of chlorine atoms, making PCE more readily reduced than TCE, DCE, or VC (Byl & Williams, 2000).

2.1.2 Guideline Values

Chlorinated ethenes are regulated through several partially overlapping guideline systems. At the European level, EU Directive 2020/2184 (Directive (EU) 2020/2184,

2020) establishes drinking-water quality requirements that are implemented nationally through Livsmedelsverket's drinking-water regulation (Livsmedelsverket, 2022). Together, these set a combined limit of 10 µg/L for PCE and TCE and a limit of 0.5 µg/L for VC. These values are mirrored as groundwater thresholds in SGU-FS 2024:1 (Sveriges geologiska undersökning (SGU), 2024) and function as comparison criteria in Swedish contaminated-site assessments.

Additional solvents have SGU threshold values that are partly based on international sources. For example, 1,2-DCE has an SGU value of 50 µg/L (cis+trans), derived from WHO and US EPA drinking-water guidelines (Sveriges geologiska undersökning, 2024). WHO guideline values differ somewhat from Swedish limits, with TCE 8 µg/L, PCE 100 µg/L, and VC 0.3 µg/L (World Health Organization, 2022).

2.1.3 Molar Normalisation

Ricker et al., 2020 observed that at a vast majority of groundwater assessment and remediation sites, chlorinated ethenes are evaluated on a mass concentration basis.

When evaluating the behaviour of chlorinated ethenes, it is not uncommon for practitioners to evaluate degradation or sourcing trends by summing the mass concentrations of each chlorinated ethenes species detected in groundwater and reporting the data as "total chlorinated ethenes". While this is a common approach in the environmental industry, it is not necessarily the most useful approach and can lead to misinterpretation of the data with unintended consequences (Ricker et al., 2020).

Ricker et al., 2020 demonstrates that the same molar concentration of PCE and its degradation products (TCE+DCE+VC) will yield different mass concentration. This highlights why molar-based evaluation is more appropriate for understanding transformation pathways.

WSP converts each compound to molar units and express the total as PCE-equivalents (PCE-eq.) present concentration trends, which provides a more chemically consistent measure of total contaminant mass. Although molar-equivalent calculations can be performed using any chlorinated ethene as the reference compound, PCE-equivalents are used in this study because PCE is the primary parent contaminant at the majority of sites used in this study. PCE-equivalents are calculated as follows:

$$PCE_{\text{Eq}} = \left(\left(\frac{PCE(\frac{\mu g}{l})}{M_{\text{PCE}}} \right) + \left(\frac{TCE(\frac{\mu g}{l})}{M_{\text{TCE}}} \right) + \left(\frac{DCE(\frac{\mu g}{l})}{M_{\text{DCE}}} \right) + \left(\frac{VC(\frac{\mu g}{l})}{M_{\text{VC}}} \right) \right) * M_{\text{PCE}} \tag{2.1}$$

Each term represents the molar concentration of the respective chlorinated ethene, obtained by dividing its mass concentration by its molecular weight (M). The sum of all molar concentrations is then multiplied by the molecular weight of PCE to express the result in PCE-mass units. While one could avoid this final multiplication and report the total molar concentration directly, regulatory guidelines, background values, and most reporting frameworks are expressed in mass concentration units, which are also more intuitive for practitioners and clients.

Analysis of PCE-equivalents provides a useful overview of the total molar burden of the plume. However, a stable PCE-equivalent does not necessarily indicate that no degradation is occurring. For example, decreases in PCE may be offset by increases in daughter products, resulting in a constant total molar concentration. Therefore, it remains essential to evaluate the individual chlorinated ethenes to correctly interpret degradation processes and plume evolution.

2.2 Statistical Methods for Groundwater Monitoring Data

The objective of any long-term groundwater contamination monitoring is to identify trends in concentrations, often with the help of statistical methods. Frequency of sampling can be determined with the help of analysing those trends, and so the choice of statistical test is crucial for the assessment to be accurate (U.S. Environmental Protection Agency, 2005). Furthermore, Grath et al. (2001) points out important instructions from the EU Water Framework Directive regarding handling of groundwater:

> *"Member States shall use data from both surveillance and operational monitoring in the identification of long term anthropogenically induced upward trends in pollutant concentrations and the reversal of such trends."* (Grath et al., 2001)

and

> *"Reversal of a trend shall be demonstrated statistically and the level of confidence associated with the identification stated."* (Grath et al., 2001).

2.2.1 Descriptive Statistics

Basic descriptive statistics, including the mean (μ) and standard deviation (σ), are used to characterise concentration levels and variability in each time series (U.S. Environmental Protection Agency, 2009). The mean represents the average concentration over time, while the standard deviation describes the spread of observations around the mean. These measures form the basis for derived metrics such as the coefficient of variation and are used to support interpretation of distributional characteristics and potential heteroskedasticity, i.e. whether the variability in the data changes with concentration level.

2.2.1.1 Coefficient of Variation

The coefficient of variation (CV) is a dimensionless measure of relative variability defined as the ratio between the standard deviation and the mean. It may be expressed either as a fraction or as a percentage and is commonly used to describe the level of uncertainty and dispersion in a dataset (U.S. Environmental Protection Agency, 2009). Lower CV values indicate that observations are closely clustered around the mean, whereas higher values reflect greater relative variability.

In groundwater contamination data, concentrations often vary substantially in both space and time, which can make variability difficult to compare across datasets with different magnitudes. For positively valued and approximately normally distributed data, the CV is expected to be relatively small. As a rule of thumb, values above approximately 0.5 are often interpreted as indicating high variability and potential deviation from normality.

Because the CV depends directly on the mean, it becomes less stable when mean values are close to zero. It is therefore best interpreted as a descriptive measure of relative spread rather than a formal test of statistical properties or trend reliability.

2.2.1.2 Skewness

Skewness is a sample statistic that describes the asymmetry of a distribution (U.S. Environmental Protection Agency, 2009). A value close to zero indicates a symmetric distribution, which is consistent with the assumption of normality, while positive and negative values indicate right- and left-skewness, respectively.

In groundwater monitoring applications, skewness is often used to assess whether data deviate from symmetry prior to applying statistical methods that assume normality. Absolute skewness values below approximately one are commonly interpreted as indicating approximate symmetry, whereas larger values suggest increasing departure from normality.

Although useful for a quick assessment, skewness should be interpreted with caution, as it does not uniquely identify non-normality, symmetric non-normal distributions can also yield skewness values near zero.

2.2.2 Data Pre-processing

After receiving sampling data from labs, data can often include non-detects that need to be assessed and handled. Non-detects (ND's) are measured values that are below the detection limit of the testing/laboratory equipment and can therefore not be quantified. Although this can mean contamination in that time and place is not of concern, it is important to include these findings in data analysis. There are some common methods in statistical procedures for incorporation of non-detects, the most common of which simple substitution. This approach replaces the unknown ND's with a numerical value based on what the detection limit is. Depending on the situation the substitution is commonly set to zero or a fraction of the reporting limit, often 50% (U.S. Environmental Protection Agency, 2009). Two alternatives are Kaplan-Meier and Regression on Order Statistics (ROS), which are more calculation-based techniques found more commonly in literature than in practice (U.S. Environmental Protection Agency, 2009).

In any monitoring programme there are bound to be measurements that do not fit the broad pattern of the other samplings. This can either be a result of e.g. a hydrogeological extreme event that caused a higher reading than expected, or it can be a fault of any step along the sampling, testing, or data handling procedures (U.S. Environmental Protection Agency, 2009). In any case, it is relevant to investigate site characteristics to identify the nature of any outliers. Common choices of outlier identification tests include Dixon's test and Rosner's test.

2.2.3 Trend Analysis and Statistical Methods

Trend detection methods in hydrologic time series are commonly divided into parametric and non-parametric approaches. Parametric trend tests rely on stronger assumptions, typically requiring independent observations and normally distributed residuals. Non-parametric methods, by contrast, make minimal distributional assumptions and only require independence of observations. The Mann–Kendall test and Sen's slope estimator are widely used non-parametric techniques for detecting monotonic trends, i.e. long-term changes that proceed consistently in one direction.

2.2.3.1 Hypothesis Testing

A central component of statistical inference is hypothesis testing, which provides a formal framework for deciding whether observed changes in groundwater quality reflect real trends or can be explained by natural variability. In trend analysis, the null hypothesis (H_0) states that concentrations show no monotonic change over time, while the alternative hypothesis (H_a) states that an increasing or decreasing trend is present (U.S. Environmental Protection Agency, 2009).

Trend tests (e.g., regression-based or nonparametric tests) compute a test statistic and compare it with its sampling distribution under H_0. If the resulting p-value is below a chosen significance level (e.g. $\alpha = 0.05$) H_0 is rejected and the trend is classified as statistically significant. This provides a transparent and comparable decision rule across different software tools.

Two types of errors may occur. A Type I error (false positive) means concluding that a trend exists when it does not, potentially triggering unnecessary investigations or inappropriate reductions in monitoring. A Type II error (false negative) means failing to detect a real trend, which can delay remediation or lead to inefficient sampling strategies.

2.2.3.2 Ordinary Least Squares Regression

Ordinary Least Squares (OLS) (Zdaniuk, 2024; Wikipedia contributors, 2024) regression is one of the most widely used statistical methods for identifying relationships in data. The core idea behind OLS is to find the line that best represents the observed data. Since real-world data rarely align perfectly along a straight line, the model measures how far each observation lies from the predicted trend. These differences are referred to as residuals, or errors. OLS selects the trend line that minimises the overall size of these errors, giving greater weight to larger deviations by squaring them. This is the origin of the term "least squares."

The method relies on several important assumptions. Most notably, it assumes that the relationship between variables is approximately linear and that the residuals are distributed normally. It also assumes that observations are independent and that the variation of the residuals remains relatively constant across the data.

In practice, the model estimates coefficients for each predictor variable. These coefficients describe both the direction and strength of the relationship with the outcome variable. Positive values indicate that the outcome tends to increase alongside the predictor, while negative values suggest the opposite.

2.2.3.3 Local Linear Regression

Local linear regression, unlike OLS regression, is a non-parametric method that, instead of viewing an entire dataset as linear, estimates linear models in subsets of data and then connects them. This creates a curve, therefore capturing local complexities and allowing for non-linearity across subsets (Fan, 1993; Jones et al., 2026).

2.2.3.4 Mann-Kendall Trend Test

The Mann-Kendall trend test (Mann, 1945; Kendall and Gibbons, 1962) is a non-parametric method used to detect monotonic trends in time series data without requiring assumptions about the underlying data distribution. Rather than fitting a regression line, the test evaluates the relative ordering of observations over time by comparing each data point with all subsequent observations.

The Mann-Kendall test statistic S is calculated as

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}(x_j - x_i) \tag{2.2}$$

where n is the number of data points, x_i and x_j are the observations at time series i and $j(j > i)$ respectively, and $sgn(x_j - x_i)$ is the sign function as:

$$\operatorname{sgn}(x_j - x_i) = \begin{cases} +1, & \text{if } x_j - x_i > 0, \\ 0, & \text{if } x_j - x_i = 0, \\ -1, & \text{if } x_j - x_i < 0. \end{cases} \tag{2.3}$$

The statistic S represents the difference between the number of increasing and decreasing observation pairs. A positive value of S indicates an upward trend, while a negative value indicates a downward trend.

For larger sample sizes, the Mann–Kendall statistic is approximately normally distributed and can therefore be converted into a standardised test statistic used to calculate the p-value. For smaller datasets, the exact distribution of the statistic is used instead of the normal approximation. The statistical significance of the trend is assessed using the associated p-value.

Although the Mann-Kendall test can determine whether a statistically significant monotonic trend exists, it does not estimate the magnitude of the trend. Trend magnitude is commonly estimated using the Theil–Sen estimator, a non-parametric method that calculates the median slope between all pairs of observations. The Theil–Sen estimator is frequently used together with the Mann–Kendall test as a robust alternative to ordinary least-squares regression (U.S. Environmental Protection Agency, 2009).

2.2.3.5 Mann-Kendall Trend Test (Seasonal)

The seasonal Mann-Kendall test is an extension of the Mann-Kendall trend test designed for time series that exhibit recurring seasonal patterns (U.S. Environmental Protection Agency, 2009). It is used to detect the presence of a monotonic trend while accounting for systematic seasonal variation, such as recurring annual fluctuations in groundwater concentration data. The method operates by dividing the data into seasonal subsets (e.g. measurements from the same month or sampling period across different years) and applying the Mann-Kendall test within each subset. The resulting test statistics are then combined to obtain an overall measure of trend across all seasons.

2.2.3.6 Theil-Sen estimator

The Theil-Sen estimator (also called Sen's slope estimator) is a non-parametric method used to estimate the magnitude of a monotonic trend in time series data (Theil, 1950; Sen, 1968). The method is robust against outliers and does not require the residuals to be normally distributed.

For each pair of observations (x_j, x_k) with $j > k$, an individual slope is calculated as:

$$Q_i = \frac{x_j - x_k}{j - k} \quad \text{for} \quad i = 1, ..., N, \tag{2.4}$$

where Q_i is the slope between two observations and N is the total number of unique observation pairs. If there is only one observation per time period, the total number of pairs is given by

$$N = \frac{n(n-1)}{2}. \tag{2.5}$$

where n is the number of observations.

The Theil-Sen slope estimate is obtained as the median of all calculated slopes. A positive slope indicates an increasing trend, while a negative slope indicates a decreasing trend. The magnitude of the slope represents the rate of change over time.

Confidence intervals for the slope estimate can be calculated based on the variance of the Mann–Kendall statistic described in Section 2.2.3.4. If the confidence interval includes zero, the estimated trend is not statistically significant at the chosen significance level.

2.3 Interpolation

Interpolation is the process of estimating values between points in a dataset via mathematical modelling. It is sometimes also referred to as gridding, since it involves creation of a continues surface, or grid, from scattered points. Since this is handled through computational equations, there are various methods with differing levels of complexity (Golden Software, 2025). The three techniques used in this thesis work are called kriging, natural neighbour, and nearest neighbour.

2.3.1 Kriging

One of the most used interpolation methods is kriging, due to its performance for a wide range of dataset characteristics. It can consistently generate adequate estimations even if data is on the more limited side, which often makes it fitting for environmental purposes (Golden Software, 2025).

Kriging bases estimation of unknown grid nodes on close-by known data points by also accounting for distance. The calculation for a single point is therefore a result of weighing all nodes used in the calculation, with weighing totaling 1 which makes sure clustered points of data do not get over represented in the final estimated value. In its most basic form, the equation for this method is as follows:

$$Z_A = \sum_{i=1}^{n} W_i Z_i \tag{2.6}$$

Where Z_A is the approximated value at a given point A, n is number of included neighbouring nodes, and Z_i is the value of a node at point i with the assigned weight W_i.

The kriging interpolation algorithm is tied to what is called a variogram. To produce the weights used in kriging, a variogram must be appropriately fitted to the data so that the spatial correlation between datapoints is accurately represented. A variogram, in general, measures how similarity decreases with distance, based on the idea that points closer together tend to be more alike than points farther apart (Golden Software, 2024c).

2.3.2 Nearest & Natural Neighbour

A much more straightforward option is the nearest neighbour method, which simply assigns an empty node with the same value of the closest data point. This makes it most fitting for filling gaps between consistently spaced points and when gaps are limited (Golden Software, 2025).

Natural neighbour is a gridding method that also estimates node values based on close-by data points, but groups multiple points together into subsets. Like kriging, weight is assigned to these subsets which is good for estimation, but the method is generally unfitting for datasets with large areas without data (Golden Software, 2025).

2.3.3 Cross Validation

Cross validation is a quality test for interpolation methods that produces a quantified value indicating just how well suited a gridding technique is for a given dataset. It involves removing one data observation at a time and letting the chosen interpolation method predict the value at that location based on the remaining point's measurements. That estimation for a point *n* is subtracted with the real value, which yields the error for that point. This is reproduced for a dataset up until the last point, always returning the previously removed measurement to the set before moving to the next (Golden Software, 2024a). A measurement often calculated is the root mean square error (RMSE) which highlights the spread of and severity of errors. Since it squares the error it penalises large differences, which is many times a preferred indicator (Sharma et al., 2022). The equation is simply:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(n_{\text{predicted}} - n_{\text{actual}})^2}{N}} \tag{2.7}$$

For a dataset of size N. It needs to be put into context and has no decided cutoff for when the RMSE indicates poor performance, but in general a low value relative to the data range is positive.

3 Methods

This chapter describes how the study was conducted, including the methodologies used for the literature review, data processing, and method evaluation. The practical workflow is divided into two main components. The first concern is identifying relevant analytical methods through information search, data pre-treatment, and assessing the suitability of each method. The second component focuses on evaluating software tools in relation to today's methods.

3.1 Strategic Literature Search

To better understand the various approaches for evaluating monitoring programmes, a strategic literature search was conducted. Academic sources were identified using several bibliographic search engines, including Google Scholar, Scopus, and the Chalmers University Library. Additional resources, such as government reports, will also be essential.

An initial set of key references was used to establish a foundational understanding of statistical methods for groundwater monitoring data, including the Unified Guidance U.S. Environmental Protection Agency, 2009 and documentation for two commonly referenced optimisation tools: GWSDAT and MAROS. These sources served as core literature from which central concepts, terminology, and methodological approaches were extracted.

Building on these initial sources, a snowballing strategy was applied, in which references cited in this key papers were reviewed to identify additional relevant literature and deepen the investigation of specific tools and methods. This process also helped define the set of search terms needed for a more systematic search.

Using these keywords which were derived both from the research question and from the snowballing phase, targeted searches were conducted, initially limited to peer-reviewed articles published since 2000. Together, these steps formed the basis for a structured literature search plan, presented in Table 3.1.

The search began with a broad search string to identify general literature evaluation of GMP related to chlorinated solvents:

> *("Groundwater" AND ("Monitoring" OR "Monitoring Network" OR "Monitoring Programme" OR "Monitoring Program") AND ("Analysis" OR "Evaluation" OR "Assessment" OR "Optimization" OR "Performance") AND ("Chlorinated Solvents"))*

With a publication-year filter (post-2000), this search returned 212 results in Scopus. Screening of abstracts revealed a notable absence of temporal analysis. From the core literature, it was already clear that temporal evaluation is a key component in assessing contaminated GMP. Specific temporal terms were therefore incorporated into the search string:

> *("Groundwater" AND ("Monitoring" OR "Monitoring Network" OR "Monitoring Programme" OR "Monitoring Program") AND ("Analysis" OR "Evaluation" OR "Assessment" OR "Optimization" OR "Performance") AND ("Contamination" OR "Pollution" or "Chlorinated*

Table 3.1: Literature search plan for approaches for evaluating monitoring programmes.

Keywords	Monitoring programme – groundwater monitoring, water quality monitoring, contamination monitoring, long-term monitoring, monitoring network
	Groundwater – groundwater quality, contaminated sites, chlorinated solvents
	Evaluation – assessment, performance assessment, monitoring optimisation, sampling optimisation, sampling frequency, network optimisation, redundancy analysis
	Statistical methods – trend analysis (linear regression, log-linear regression, Mann–Kendall test, Seasonal Mann–Kendall, Theil–Sen slope), non-parametric trend tests, time-series analysis
	Spatial and spatiotemporal analysis – Ricker method, kriging, spatial interpolation, spatiotemporal modelling, penalised splines, non-parametric smoothing, plume metrics (mass, area, centre of mass)
Bibliographic source	Google Scholar
	Scopus
	Chalmers Library
	Website of national authorities
Criteria	Research papers published in peer-reviewed English-language journals
	Grey literature, either English- or Swedish-language
	Produced post-2000

Solvents") AND ("Time Series" OR "Temporal" OR "Long-term" OR "Trend"))

This refined search string returned 64 results in Scopus, suggesting that the evaluation of GMP for chlorinated solvents particularly with respect to temporal analysis, may not be a highly researched topic.

Since an overall understanding of the tools and methods had already been established, the search strategy was refined to focus on more detailed and method-specific studies. (1) Spatiotemporal and (2) trend detection terms were therefore incorporated into the search strings:

(1) *("Groundwater" AND "Chlorinated Solvents") AND (("Spatiotemporal" OR "Spatio-temporal" OR "Space-time")*

(2) *("Groundwater" AND "Chlorinated Solvents" AND ("Trend Analysis" OR "Mann-Kendall" OR "MK test" OR "Seasonal Mann-Kendall" OR "Sen's slope" OR "Theil-Sen" OR "Kendall tau" OR "Non-parametric trend" OR "Time Series Analysis" OR "Trend Detection"))*

The search string containing spatiotemporal and trend-detection terms returned 31 and 7 results in Scopus, respectively, further indicating that the evaluation of groundwater monitoring programmes for chlorinated solvents remains under-researched. Consequently, much of the relevant literature has been obtained through snowballing and grey-literature sources, such as technical guidance documents from the US EPA and consultancy reports produced by WSP.

In order to determine if the literature were worth further instigation the title, keywords and abstract were evaluated. After the first inspection an overall of 13 research papers and 4 grey literatures were reviewed. The research paper was used to answer specific questions. The grey literature was useful to understanding applied practices and regulatory context. To ensure relevance, the literature selection prioritised recent publications, with 7 sources from 2020 onward, 5 between 2010-2020, and 5 for before 2010.

This literature review was conducted as a strategic rather than a systematic review, meaning that it does not aim to capture all existing publications on evaluating monitoring programmes. Although multiple databases and search strategies were used, relevant studies may have been missed due to database coverage, keyword selection, publication language, and access limitations.

3.2 Data & Study Material

All data used for analysis throughout this work was provided by the WSP. More specifically, all datasets were from GMP investigating chlorinated solvents in different locations all over Sweden. The datasets include parameters such as PCE, TCE, DCE, VC, and PCE_{eq}. The monitoring networks varied in size and design: numbers of wells per site, sampling frequency, monitoring duration, and the presence of inactive or substantial changes in sampling frequency over time, differed from project to project. The number of sampling events per well ranged from 3 to 29.

Any sensitive information that could link a project to a client company, e.g. location or well names, were altered as to ensure anonymity. Coordinate values were not reported, but the relative geometry of the coordinate system (i.e., the spatial ratios between wells) was preserved. Constituent concentrations, dates and were kept the same, however.

Table 3.3 provides an overview of the datasets used in the evaluation, with a more detailed table presented in Appendix A.1.

Table 3.2: Summary of the most important characteristics of the datasets.

Location	Number of wells	Number of samples per well	Monitoring period
A	4	11–17	12 year
B	8	4–13	13 year
C	13	5–29	29 years
D	6	3–7	6 year
E	16	3–17	23 year
F	12	4–18	6 year

In total, 6 monitoring networks and 59 wells were included, representing a broad spectrum of real-world groundwater conditions. This variability ensures that the evaluation reflects typical challenges encountered in groundwater trend analysis all over Sweden.

As few geological parameters as possible was used to maintain current workflow simplicity. However, the aquifer thickness and effective porosity has been used for the

Ricker Method calculations (Section 3.4.2). Since Location A was not included, none of its parameters are presented below. Aquifer thickness and soil type were provided by WSP, and effective porosity is estimated from the soil type for all locations except C, where it was provided directly.

Table 3.3: Summary of geological parameters for the datasets.

Location	Aquifer thickness [m]	Estimated effective porosity [%]	Soil type
B	5	20	gravelly/fine sand
C	2	12.5	-
D	3	25	mainly silty sand
E	0.5	10	clayey, silty till
F	3.75	5	silty clay

Locations A and F have undergone remediation prior to their first sampling dates, and the substances historically used at the sites are PCE for Location A and TCE for Location F. Location A has also been remediated, but in the middle of the monitoring range at the Other than that none of the locations have gotten remediated and instead have PCE as their origin-substance.

3.3 Data Processing and Pre-treatment

To produce meaningful results from time series evaluations, only monitoring wells with at least 3 sampling dates were included. Also, wells with significantly deeper sampling were excluded to ensure elevation homogeneity within each location. For the spatial gridding required for the Ricker Method, only dates that sampled 3 or more wells were included since this is required for interpolation.

To ensure consistency in the evaluation of temporal trends, all analyses were performed using PCE_{eq}. This metric is the current standard at WSP and has an additional advantage of removing the influence of non-detect (ND) handling in the software comparison, since none of the PCE_{eq} time series contain ND.

For monitoring programmes where PCE_{eq} values were not provided, they were calculated by the authors following WSP's standard procedure (See Section 2.1.3). The calculation incorporates PCE, TCE, DCE and VC, where non-detects are substituted with ND/2 in accordance with WSP's established practice for trend evaluation. This ensured that all datasets were harmonised prior to software implementation.

All data used in this study consist of evaluation datasets provided by WSP. Since these datasets have already undergone quality control, no additional outlier treatment was included in the data processing workflow.

3.3.1 Statistical Characterisation of PCE_{eq} Time Series

To support both visual and statistical trend evaluation, basic descriptive statistics were calculated for each well, including the number of sampling events, mean concentration, standard deviation, coefficient of variation (CV), and skewness (Appendix A.2). These statistics were calculated exclusively for the PCE_{eq} time series, as that is the basis for all trend evaluations in the study.

Skewness was included to assess deviations from normality to choose appropriated

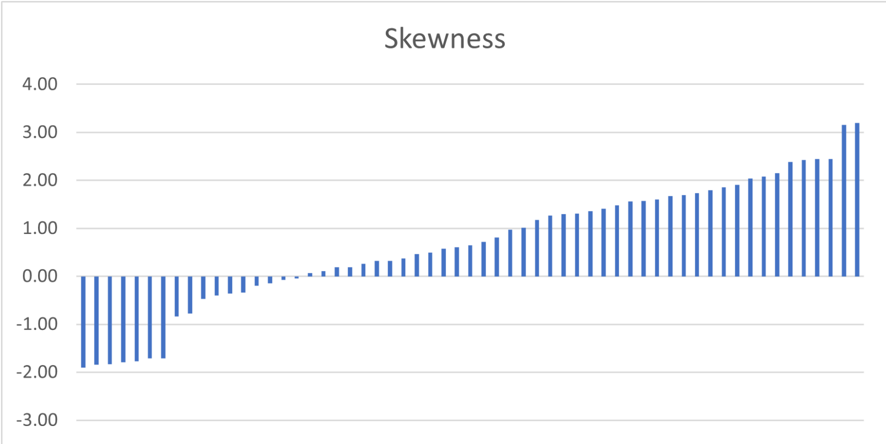


Figure 3.1: Skewness for all wells in all locations (unitless).

trend evaluation, as groundwater contaminant data often exhibit asymmetric distributions with occasional extreme values. Figure 3.1 presents the skewness values for all 59 time series, ordered from lowest to highest, to illustrate the distributional characteristics relevant. A few wells displayed strongly negative skewness, most likely reflecting generally elevated concentrations with a single unusually low measurement, potentially a non-detect. In contrast, the majority of wells showed clear positive skewness, indicating non-normal distributions with occasional high concentration values. Several wells exhibited skewness values above 2, indicating extremely right-skewed distributions. Such distributions violate the normality assumption required for ordinary least squares (OLS) regression, as the resulting residuals become heavy-tailed and asymmetric. Under these conditions, OLS-based trend tests yield unreliable p-values and confidence intervals.

A scatterplot of log-transformed standard deviation versus log-transformed mean concentration was used to assess heteroskedasticity, revealing that wells with higher mean concentrations generally exhibited higher variance. This variance structure is important for visual trend interpretation, as increasing variability can make trends appear weaker or more irregular (Figure 3.2).

Because the data were skewed, heteroskedasticity and because sampling intervals were irregular, the Mann–Kendall test was selected over linear regression for quantitative trend evaluation (Section 3.4.1). The same skewness patterns also motivate the use of log-transformation in the spatial analyses, including the kriging comparison in the Ricker method evaluation (Section 3.4.2).

3.3.2 Software Specific Data Requirements

In contrast to the statistical characterisation, the software pre-processing involved all individual chlorinated ethenes (PCE, TCE, DCE and VC) as well as PCE_{eq}. Each software tested required different data-import formats. Although all programs accepted the same file type, their sheet layouts, column structures, and metadata requirements differed, necessitating additional pre-processing before import. Furthermore, the datasets provided by WSP originated from multiple projects and consultants, resulting in non-standardised formats, inconsistent naming conventions, and varying levels of completeness. Because of this heterogeneity, no universal conversion template could be applied.

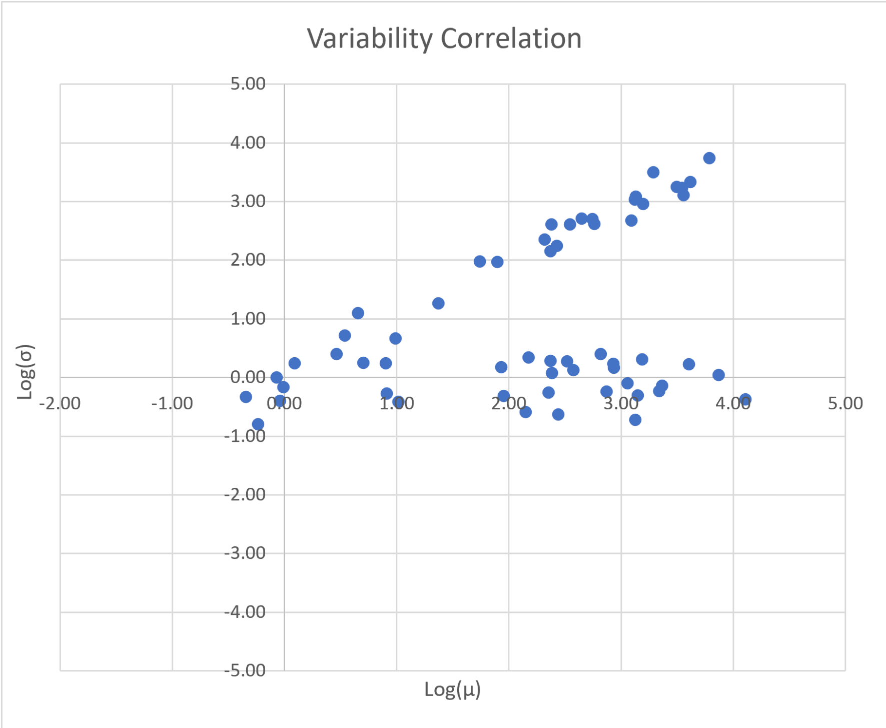


Figure 3.2: Standard deviation (σ) plotted against mean (μ) concentration for all wells, log-transformed.

3.4 Evaluation of Analytical Methods for Groundwater Programme Optimisation

After the most used time-related trend analysis methods were found in the literature search (Section 3.1) and the provided datasets were processed, comparisons could be made. By evaluating the differences between the identified approaches, their advantages and limitations were highlighted to support a discussion on the validity of current practices. The two techniques that were tested was regular time-series plots (using individual constituents or a sum of constituents expressed as PCE-equivalent) and site-wide plume variation, specifically following the steps of the Ricker Method, explained in Section 3.4.2.

3.4.1 Time series Analysis

Evaluation of the time series was carried out using both qualitative and quantitative approaches. The qualitative assessment was based on visual interpretation of each well's temporal pattern, while the quantitative assessment applied the Mann–Kendall trend test to the same datasets.

To examine the visual assessment, two assessors (Person A and Person B) independently classified the trend for all wells using individual PCE_{eq} time-series graphs. Figure 3.3 shows an example of a PCE_{eq} graph on a logarithmic scale, presenting the concentration trends for all wells at one location (Appendix B.1).

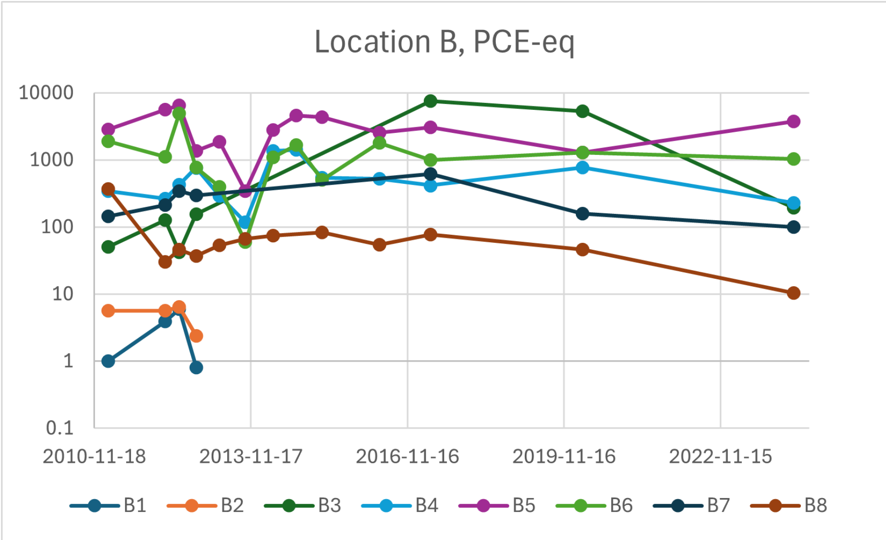


Figure 3.3: PCE-eq plotted for all wells for Location B, logarithmic scale.

The trends were identified as Decreasing, Stable, or Increasing, each accompanied by a certainty rating (“very sure”, “somewhat sure”, “somewhat unsure”, “very unsure”). Classifications were based on the assessors’ understanding of temporal variation, noise, and outliers. A proper evaluation of all time series therefore required a clear understanding of the underlying data structure, as described in Section 3.3.1.

To synthesise the two independent assessments, a set of transparent decision rules was applied:

- If both Person A and Person B selected the same trend and both were "very sure", the trend was classified as Decreasing, Stable, or Increasing with high confidence.
- If either Person A or Person B was "very sure" and the other "somewhat sure", the trend was classified as “probably decreasing/stable/increasing”, depending on the shared direction.
- In all other cases the trend was identified as non-existing.

The two independent assessments were then compared to identify similarities and differences in both trend classification and certainty levels. The full comparison tables and certainty distributions are provided in Appendix B.2

In addition to the visual assessment, a quantitative trend evaluation was performed using the Mann–Kendall test. Due to its statistical nature and its established use in environmental data analysis, the ProUCL software was selected to compute the relevant parameters for all 59 temporal datasets.

A temporal trend analysis based on the Mann–Kendall approach was applied to each well. The method uses the Mann–Kendall S statistic, the confidence factor, and the coefficient of variation to classify trends into six categories: decreasing, probably decreasing, stable, no trend, probably increasing, and increasing (Newell et al., 2006). The classification criteria used in this study are summarised in Table 3.4.

The resulting trend classifications were later compared with the visual assessments to evaluate the degree of agreement between subjective interpretation and statistical testing, and to identify systematic differences between the two approaches.

Table 3.4: Trend classifications based on Mann-Kendall S statistic, the confidence factor, and the coefficient of variation (CV) (Newell et al., 2006).

Trend class	S statistic	Confidence factor	CV
Increasing	> 0	$\geq 95\%$	–
Probably increasing	> 0	$\geq 90\%$	–
Stable	$= 0$	$< 90\%$	≤ 1
Probably decreasing	< 0	$\geq 90\%$	–
Decreasing	< 0	$\geq 95\%$	–
No trend		$< 90\%$	> 1

3.4.2 The Ricker Method

The Ricker Method® is an open-source statistical tool designed to assess the stability of an entire contamination plume (Golden Software, n.d). Rather than evaluating changes well by well, the Ricker Method leverages statistical analyses of multiple sampling events across all monitoring wells.

To test this method, the following steps were carried out for each location and date; (1) gridding of concentration data to a continuous surface, (2) definition of the plume boundary using a cutoff concentration, (3) calculation of the planar area and concentration volume from the gridded surface, and (4) computation of of the plume mass using aquifer properties. This follows the methodology in the provided guide from Golden Software (2023).

Steps 1 through 3 were carried out with the Surfer software, a contouring and grid-estimation programme from Golden Software which is often utilised for Ricker Method calculations. Several sites were examined with anonymised well names but with real coordinate systems and sampling concentrations over time in order to properly judge if this could be a complementary tool for GMP optimisation for small scale projects. This method can be conducted for individual constituents and for total plume expressed as e.g. PCE- or TCE-equivalent, but for simplification in this thesis work the PCE-equivalent values for a given site were converted to μmol/l before implementation in the software, in accordance with the theory explained in Section 2.1.3.

There are multiple interpolation methods available, for this test kriging was chosen for all sites, although the suitability could be argued and will therefore be compared (see Section 3.4.2.1). However, even though kriging relies on modelling an appropriate variogram, the default linear variogram was chosen for all datasets, since this was the recommendation for surface level testing of kriging (Golden Software, 2024c; Golden Software, 2024b). As per recommendation in the guide, the option to log-transformation the data through the software was done, with outputs remaining the same as the upload file units. Furthermore, variograms for kriging was chosen to not be investigated, and therefore Surfers default linear variogram was used. For step 2, the plume cutoff concentration was set to 0.06 μmol/l to represent a PCE concentration of 10 micrograms per liter (Section 2.1.2). Since the concentrations used are a sum of multiple constituents with varying guideline values, this cutoff was chosen due to the other chemicals having more strict regulations. Step 3 was simply derived from the software, which could then be used for step 4.

The 4th step of the workflow is based on the calculation of the plume mass provided by

Ricker (2008):

$$\begin{aligned}\text{Plume mass (mol)} = {} & \text{Planar area } (\mathrm{m}^2) \cdot \text{Average concentration } (\mu\text{mol/l}) \\ & \cdot \text{Aquifer thickness (m)} \cdot \text{Effective porosity} \\ & \cdot 1000\,\mathrm{l/m}^3 \cdot \frac{1\,\mathrm{mol}}{10^6\,\mu\mathrm{mol}}\end{aligned} \tag{3.1}$$

The average plume concentration is calculated as the ratio between the concentration volume and the planar area, both derived from step 3:

$$\text{Average plume concentration} = \frac{\text{Concentration volume}}{\text{Planar area}} \tag{3.2}$$

In (3.2), the concentration volume represents the integrated concentration across the plume footprint, derived from the gridded concentration surface. To obtain the average concentration of the entire plume, the cutoff concentration was added to the interpolated average. For more detailed explanations regarding the Ricker Method, read Ricker (2008).

3.4.2.1 Tests in the Surfer Software

Alternative interpolation methods were applied to the Ricker Method for all dates for Location F to observe differences in results. The two other techniques chosen was nearest neighbour for its simplicity and natural neighbour due to its recurrence in literature regarding environmental contamination. Additionally, these gridding methods was also applied to the Ricker Method for three dates from each location, specifically the date with the most amount of wells sampled, the date with the fewest, and a date with a number in between the prior two. Location D was excluded from this test, however, due to its consistency in number of wells sampled.

The other tests conducted were targeted at evaluating what effect well-removal has on krigings ability to estimate plume variations. This was done through the following theoretical scenarios, all targeted at Location F (Figure 3.4):

- *F10 Added*: F10 was sampled for the first three dates of the GMP (which indicated low contamination) but was then never tested again. Before removal this well was located on the perimeter of the sites grid, which could therefore assist in improved gridding estimations. It was added for all subsequent dates with a constant theoretical value equal to the average of the three real measurements (0.854 μmol/l PCE-equivalent).
- *F3 & F10 Added*: F3 and F10 are located in the same general direction from the contamination source (F2). F3 is also a perimeter well and was sampled for longer than F10 (until the end of 2021, but with two gaps) but was ultimately removed from the programme as well, the difference being that F3 exhibited higher concentrations. For this scenario the purpose was to support conclusions regarding if keeping one well in the outskirts of a plume can suffice compared to multiple wells in the same direction. F10 was added in the same way as scenario 'F10 Added'. F3 was added similarly for all missing dates but gap-dates were assigned the midpoint value of the previous and following samplings (105.7 and

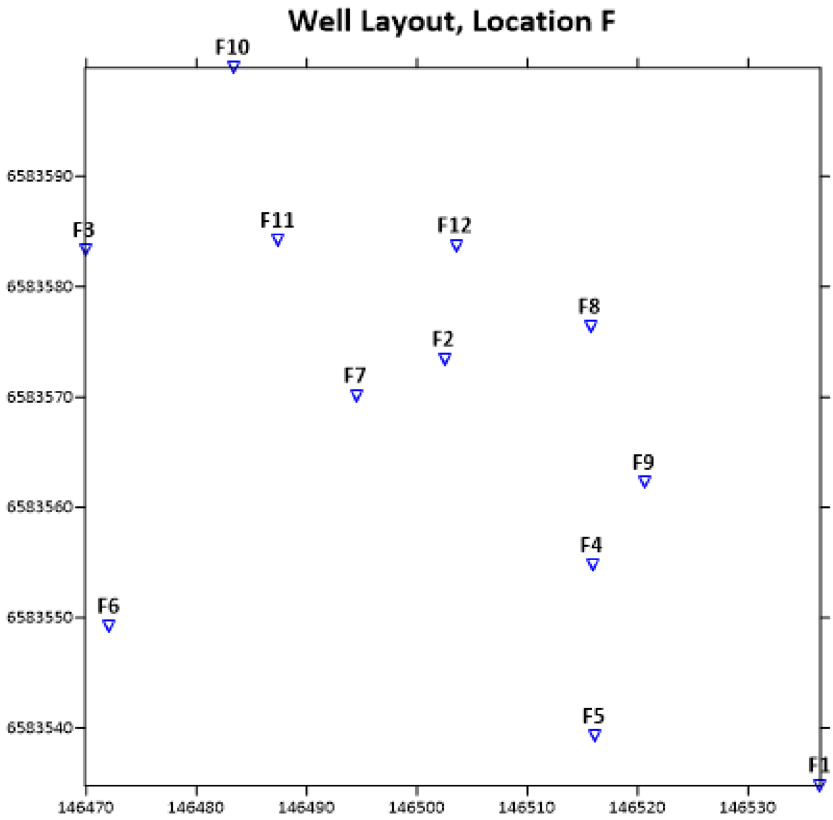


Figure 3.4: Anonymised map of well placements for Location F.

115.4 μmol/l PCE-equivalent), and all dates after 2021 were assigned the average value of the last four real measurements (175.3 μmol/l PCE-equivalent).

- *F7 Removed*: F7 is the well closest to the contamination source well (F2), and was tested from the second date in 2019 until the start of 2021. Although quite high measurements were observed, in relation to the source they were much smaller and decreasing. Before September 2023 (when F6 was added) there are no other wells in that direction. This scenario was therefore created to see how severely the plume estimations are affected if F7 was never a part of the monitoring programme. F7 measurements were simply removed from all sampling dates.

3.5 Software Evaluation

As per the third research question, promising software that can assist in the workflow of GMP were identified and evaluated. The selection of software to evaluate in depth and the procedure of that evaluation is explained in the following chapter.

3.5.1 Survey of Career Professionals Input

A survey in the form of a questionnaire was sent out to career professionals at supervisor firm WSP. The purpose of the survey was to gather insights into which software areas practitioners consider important in groundwater pollution control programs, and also, to determine which of these areas are valued the highest. Other than gaining insights into needs and requirements pertaining to the workflow of experienced personnel, a goal was to compare those insights with literature findings. Since an objective of this section was to improve upon current groundwater monitoring program evaluation, it was essential to identify any potential divide between research within this field and current prioritisation of industry professionals. The result of the survey are presented in Appendix D.1.

3.5.2 Identification of Software

Identification and selection will be the initial steps toward answering the third research question. This will mainly be approached through systematic searches using general web search engines. Some software were identified throughout the literature search. The identification-stage will only observe software and tools for their front-face value. Criteria such as groundwater focus, trend analysis capabilities, and previous usage for contaminant evaluation will later be apart of the screening-stage to determine the most suited options for data review.

Another important aspect is cost. There exist free software, one-payment per installation plans, and monthly payment services. Although price will not be an excluding factor in identifying and testing software in this work, it is important to account for during later assessment stages due to data-handling security and accessibility parameters.

3.5.3 Selection of Software Tools

Certain features were deemed as essential from literature if a software was to be further considered. First, the most critical requirements concerned the software's statistical capabilities, particularly those related to time-dependent analyses. Given the monitoring-oriented context of this work, this included both analytical and visualisation-oriented functions, with time-series representation and temporal trend detection being the two most central statistical features. Second, a set of criteria related to the graphical user interface (GUI) was included. Because an important aim of this thesis is to ensure usability for personnel without experience in terminal-based tools, the software needed to meet basic standards of accessibility, clarity, and user-friendliness. Third, essential technical requirements were defined to ensure that the software is compatible with contemporary operating systems, actively maintained, and reasonably accessible.

A total of 48 software were identified through an initial market scan, and were evaluated based on homepages, user manuals, and/or reviews of users. This made it so that prospects were disqualified continually when they were found to not meet essential criteria requirements.

Of these, 12 met the two or more initial criteria. They were investigated further and categorised as either 'promising' or 'potentially promising' based on inclusion of important statistical parameters and overall impression. Following a secondary screening, 3 tools were selected for detailed testing and evaluation according to the procedure described in Section 3.5.4. In Appendix D.2 all the identified software are presented.

3.5.4 Testing of Selected Software

The testing consisted of implementing the PCE_{eq} concentrations from 6 groundwater monitoring locations followed by identifying which criteria were fulfilled, and in some cases in what way/to which degree. Each tool was evaluated across three complementary areas: (i) functional capabilities, (ii) quantitative outputs, and (iii) subjective scoring using a MCA (Section 3.5.5).

3.5.4.1 Functional Capabilities and Characteristics

A set of criteria, defined as specific, measurable indicators of how well an option performed in relation to a given objective was identified based on both the authors' assessment and responses from the survey (Section 3.5.1).

The parameters criteria evaluation focuses checking if features exist in each software. The goal is to assess whether the tools differ in existing features and methods. For each software, the following parameters are extracted and compared:

- **Handling of ND's** – see Section 2.2.2
- **Identification of outliers** – see Section 2.2.2
- **Identification of significant or non-significant trend** – does the software provide a clear indication of trend significance results?
- **Statistical measurements** – e.g. OLS, Mann-Kendall.
- **Is input-data presented in a table** – this simple feature helps reduce unnecessary back-and-forth between programmes.
- **File formats supported**
- **GIS map import**
- **Well removal decision support** – are there any indications regarding specific groups of data points that do not add value to the results?
- **Unit conversion** – can the software assist in any form in standard unit conversions?
- **2-D interpolation** – is this spatial feature included?
- **Can highlight reference value exceedance** – e.g. users can enter a reference value and all exceeding data points are marked as red.
- **Mass-flux calculations** – is it in some way included?
- **Targeted at chlorinated solvents** – was the software created with chlorinated solvents in mind?
- **Price** – although not a function, price is an important characteristic that plays a role in choice of software.

3.5.4.2 Quantitative Outputs

The quantitative evaluation focuses on comparing the numerical trend-analysis outputs produced by each software for the same datasets. The goal is to assess whether the tools generate consistent statistical results and whether any systematic differences appear across methods or datasets, since user manuals often only provide surface-level explanations for their computational calculations. For each dataset and software, the following parameters are extracted and compared:

- **Mann-Kendall S-value** – raw test statistic indicating monotonic trend direction.
- **Mann-Kendall p-values** (tabulated and approximate) – significance of the monotonic trend.
- **Sen's slope** – non-parametric estimate of trend magnitude.

- **Confidence interval for Sen's slope** – uncertainty range around the trend estimate.

These parameters were compared across software to identify systematic differences in trend magnitude, significance assessment, and overall analytical behaviour.

3.5.4.3 Subjective Criteria

These are parameters that cannot be compared entirely objectively, and therefore must be considered with caution during the results. The purpose for including this category of criteria is largely connected to user-friendliness, since many of the following parameters are directly tied to the experience of the user. Notably, the criteria "Plots" is the only one with subcriteria due to its broad range of meaning. The aspects scored for this category are:

- **Plots**

  - **Features**: What features can be included in the plots (e.g. titles, axis-units, and dates)?

  - **Easy to produce**: How easy is it to include the available features?

  - **Quality**: How is the visual quality and clarity of the plots, with all wanted features included?

  - **Add/remove series**: How easy is it to add/remove constituents or entire well measurements to a graph?

  - **Multiple plots**: Can multiple plots be viewed at the same time, and how easy is it to switch between plots?

- **Intuitiveness** – How easy is the software to get into for the first time?
- **Initial time investment** – Does a user have to spend a lot of time to get each new project to produce results?
- **Data modifiability** – Is it possible to add or change data after a project has already been implemented, and if so how easily?
- **Viewability of multiple results** – Can e.g. multiple plots be produced at the same time, and does it allow for easy swapping between those results?
- **Transparency of methods** – Is it clear what statistical and mathematical methods were used to yield the results?

3.5.5 Multi-criteria Analysis

To properly evaluate software based on their most important features, these features first needed to be identified. This thesis therefore relied on input from a small group (5) of industry professionals from WSP working with groundwater contamination monitoring programmes to ensure that relevant needs and requirements were captured. To contextualise the relative importance of the various aspects discussed in previous chapters, a MCA was also conducted, specifically using a simple additive weighting method (Dean, 2022).

$$S = \sum(score_i * weight_i) \tag{3.3}$$

A set of main- and subcriteria (Section 3.5.5), defined as specific and measurable indicators of how well an option performs in relation to a given objective, was identified based on both the authors' assessment and responses from the survey (see Section 3.5.1). Each criterion was assigned a corresponding weight representing its relative importance compared to the other criteria under consideration; in other words, criteria linked to highly important objectives received higher weights.

To derive the final criterion weights, five respondents independently assigned importance values to each criterion (Appendix D.3). To ensure comparability and to make the weights usable in the MCA, the individual values were aggregated and normalised so that the resulting final weights sum to 1 and represent the weights used in the analysis (Table 3.5) .

Table 3.5: Total weight of each criteria based on respondents rankings.

Main criteria	Weights
Initial time investment	0.12
Data modifiability	0.19
Plots/Graphs	0.20
Additional statistical measures	0.07
Table view	0.17
Comparison values / Reference values	0.11
Intuitiveness	0.13

The sub-criteria are weighted only within their own main criterion. This means that their percentages always add up to 100% within the group, regardless of the total weight of the main criterion (Table 3.6).

Table 3.6: Total weight of each plot subcriteria based on respondents rankings.

Subcriteria	Weights
Plots – Visual parameters	0.04
Plots – Simplicity of inclusion	0.03
Plots – Quality	0.04
Plots – Adding/removing data series	0.05
Plots – Multiple plots	0.04

Once each software option had been demonstrated and tested in a dedicated session with industry professionals, the respondents assigned a performance score to each criterion. A performance score was defined as a dimensionless number on a predefined scale (–5 to +5) that reflected how well the software compared to their current workflow, which is primarily based in Excel. Higher-performing options received higher scores, while lower-performing options received lower scores.

Tables 3.7 shows the total normalised score for each criterion. These values constitute the input to the MCA and were later combined with the criterion weights to generate the final performance results. Appendix D.4 presents the raw evaluation scores assigned to each software by the respondents.

Table 3.7: Scoring of all main- and subcriteria for each software based on evaluators opinions.

Main- and subcriteria	GWSDAT	ProUCL	AquaChem
Initial time investment	-0.33	0.67	-2.67
Data modifiability	0.00	-0.67	-0.33
Plots – Visual parameters	1.67	-1.00	0.83
Plots – Simplicity of inclusion	1.17	1.50	0.50
Plots – Quality	2.00	-1.83	0.67
Plots – Adding/removing data series	1.67	0.17	0.33
Plots – Multiple plots	0.00	-0.67	1.17
Additional statistical measures	-0.17	2.83	1.83
Table view	-0.67	-0.67	-1.33
Comparison values / Reference values	2.17	0.83	0.67
Intuitiveness	1.33	0.83	-1.67

3.5.5.1 Sensitivity Analysis

To evaluate the robustness of the MCA, a sensitivity analysis was performed. Three complementary approaches were used to examine how changes in weights and scoring assumptions influence the final ranking of the software alternatives.

First, a proportional weight variation was conducted using a tornado-style approach. This method is widely used in MCA and involves increasing or decreasing one criterion weight by a fixed percentage while adjusting the remaining weights proportionally so that the total weight remains equal to 1. For each criterion, the weight was varied by $+/-\ 10\%$, and the resulting change in the total performance score was calculated using:

$$S^{+} = \frac{S_{\text{base}} + 0.1\, W_i\, \text{Score}_i}{1 + 0.1\, W_i} \tag{3.4}$$

$$S^{-} = \frac{S_{\text{base}} - 0.1\, W_i\, \text{Score}_i}{1 - 0.1\, W_i} \tag{3.5}$$

where S_{base} is the original weighted score for each software, W_i is the weight of criterion i, and $Score_i$ is the corresponding performance score. This analysis highlights which criteria exert the strongest influence on the final ranking and whether small, systematic changes in the weighting scheme alter the preferred alternative.

Second, an individual-participant sensitivity test was performed to assess how dependent the results are on who participated in the scoring. Since the MCA is based on a small group, three individual scoring profiles were selected: the participant with the highest overall scoring tendency, the participant with the lowest, and one randomly chosen participant. For each case, the MCA was recalculated using Equation 3.3 together with the original weights (Tables 3.5 and 3.6). This test provides insight into whether the ranking is stable across different evaluators or sensitive to individual judgement.

Finally, three scenario tests were carried out using absolute weight shifts. These represent more drastic changes than the proportional tornado analysis and explore "what-if" situations where certain criteria suddenly become substantially more or less important. Based on practical experience from software testing, the criterion 'Easy to produce plots' was increased from 3% to 13% to reflect a scenario where plot generation is considered a major priority. In contrast, the criteria 'Intuitiveness' and 'Time investment' were each reduced by 10 percentage points. These two aspects were difficult to demonstrate during the workshop with industry professionals, and the scenario simulates how the results would change if their influence were less dominant. Together, these scenario tests reveal whether any software alternative becomes clearly favored or loses its advantage under more extreme but plausible weighting assumptions.

3.6 AI-use

Artificial intelligence tools such as ChatGPT and Microsoft Copilot were used during the thesis work as assistance in structure, sentence flow, and overall comprehension of information. AI was not used to generate information or entire texts; rather, it was used to improve drafts that were then thoroughly examined before usage. The authors take full responsibility for the content of this thesis publication.

4 Results & Discussion

The findings are divided up into three main categories: the literature study findings, results from evaluating various approaches, and the results from the software search. This chapter results for these areas with relevant discussions regarding their meaning and reasonability to provide a more cohesive overall picture. Final thoughts on how these categories tie together is brought up later in the conclusion (Section 5).

4.1 Approaches to Contamination Evaluation in GMP

The literature shows that different analytical methods highlights different aspects of chlorinated solvents plume behaviour. Temporal trends describe concentration changes at individual wells, whereas plume-wide mass calculations provide insight into degradation ratio and stability of the whole site. PCE-equivalent and daughter-product analyses further highlight degradation ratio that may be obscured in weight-based datasets. Because each method captures a different dimension of plume evolution, applying them in combination is expected to yield a more complete and reliable interpretation of plume stability at the study site.

4.1.1 Baseline Method: Individual Well Time-Series Evaluation

Across the reviewed material, individual well time-series emerged as the most consistently applied and standard method for evaluating groundwater monitoring data. Approaches for GMP optimisation range from qualitative to quantitative (U.S. Environmental Protection Agency, 2005). While qualitative evaluations rely heavily on expert judgement to determine whether the monitoring network and sampling frequency are appropriate for site conditions, quantitative long-term monitoring optimisation techniques use numerical and statistical analyses to evaluate temporal patterns, such as trends, periodic fluctuations, or long-term mean concentrations, and to recommend adjustments to sampling frequency. Within these quantitative approaches, trend-detection methods represent the most widely referenced tools in both research and practice.

For example, Aziz et al. (2003) apply Mann-Kendall and linear regression analyses to estimate concentration trends for each well, and the Unified Guidance (U.S. Environmental Protection Agency, 2009) similarly highlights these methods as standard approaches for detecting monotonic or linear trends. The Mann–Kendall test and simple linear regression is referenced in a large number of research papers (McLean et al., 2019; Ricker, 2008; McHugh et al., 2025; Newell et al., 2006), guidelines (U.S. Environmental Protection Agency, 2009) and practical tools (Jones et al., 2026; Aziz et al., 2003). In applied contexts, trend analysis is common but far from universal; T. M. McGuire et al. (2004) report that “a trend analysis, such as Mann-Kendall statistics, was used at 24 percent of the sites”. Several studies also classify plume behaviour using Mann-Kendall-based categories such as increasing, stable, or decreasing, often combined with measures of variability such as the coefficient of variation (Aziz et al., 2003; Newell et al., 2006). By incorporating these quantitative tools, monitoring programs can move beyond subjective assessments and simple spreadsheet trend plots toward more defensible, data-driven decision making.

In addition to monotonic trend tests, there are methods addressing non-monotonic temporal behaviour in groundwater concentration data. Local linear regression, as described by Jones et al. (2026), provides a flexible smoothing approach capable of revealing curved, episodic, or cyclic patterns that simple linear or Mann–Kendall tests may overlook. Similarly, the seasonal Mann-Kendall test extends the standard Mann-Kendall framework by accounting for recurring seasonal structure in datasets with regular sampling intervals (U.S. Environmental Protection Agency, 2009). Although not typically considered baseline methods, these approaches highlight that groundwater time series often deviate from monotonic assumptions, underscoring the limitations of relying solely on traditional trend-detection techniques.

The literature highlights that statistical methods differ in the types of data structures they are intended to handle, making it important to select an approach that aligns with the form of the available data. For example, if the objective is to compare concentrations before and after a remediation action, the data form two groups, and a two-sample test is appropriate (Lin et al., 2025; U.S. Environmental Protection Agency, 2009). In contrast, if the goal is to evaluate whether concentrations change gradually over time, the dataset constitutes a continuous time series, and a non-parametric trend test such as the Mann–Kendall test is more suitable. Correctly identifying the data type is therefore essential for choosing an appropriate temporal evaluation method (U.S. Environmental Protection Agency, 2009).

However, the limitations of temporal analysis are repeatedly emphasised. Monitoring data often exhibit substantial short-term noise, and long time series are required to detect statistically robust trends. Mchugh et al. (2015) found that more than 7 years of quarterly monitoring data was required to identify a statistically significant concentration trend in most wells, illustrating the challenge of relying solely on temporal methods for programme evaluation. These constraints have contributed to the development of a wide range of alternative approaches for assessing GMP.

4.1.2 Complementary Approaches Beyond Temporal Trend Testing

Beyond temporal trend evaluation, the literature describes several alternative approaches. Examples of such non-temporal approaches include high-resolution spatial sampling (e.g., dense well spacing near plume margins to reduce spatial noise) and high-volume sampling (e.g., pumping larger volumes to obtain more representative concentrations and reduce random variability) (McHugh et al., 2025). Additional approaches identified in the search include tracer-based transport studies (e.g., bromide tracer tests), spatial statistical methods such as Bayesian modelling, background-value comparison, and cluster analysis. The literature also highlights more advanced spatiotemporal approaches, including numerical models such as MODFLOW 6 (Benavides Höglund et al., 2023) , which are used to represent plume heterogeneity and spatial patterns that cannot be captured through well-by-well temporal analysis alone. While these methods provide valuable insights for network optimisation and source identification, several fall outside the temporal focus of this report and will therefore not be evaluated further. Others require extensive site-specific information to produce reliable results, making them unsuitable for the more general assessment undertaken here.

As mentioned, several studies highlight that monitoring data often exhibit substantial spatial and temporal variability, requiring analytical approaches that go beyond single-well measurements. Building on this, many studies focus on spatiotemporal modelling and plume-based metrics as a way to extend single-well trend tests without requiring large amounts of new data (Jones et al., 2026; Jones et al., 2014; Ricker, 2008;

Aziz et al., 2003; J. T. McGuire et al., 2000). A foundational contribution is Ricker (2008), who outlines practical methods for estimating plume area, mass, and centre of mass using grid-based calculations rather than complex numerical models (Jones et al., 2014). Such plume metrics are widely used to assess plume stability, shrinkage, or expansion. Although more advanced spatiotemporal approaches remain relatively uncommon (McLean et al., 2019), techniques such as penalised-spline smoothing demonstrate how spatial and temporal patterns can be modelled jointly to produce more coherent representations of plume behaviour and reduce noise (Jones et al., 2014). Spatial interpolation methods, including kriging (J. T. McGuire et al., 2000), further illustrate how spatial modelling can reveal plume geometry and variability that are not apparent from point measurements alone. Other plume-scale metrics, such as contaminant mass discharge (Bøllingtoft, 2025; Bøllingtoft et al., 2026), have also been proposed for evaluating source-zone behaviour and remediation performance. However, these approaches require detailed hydraulic and mass-flux information and therefore fall outside the temporal focus of this report.

Regardless of whether groundwater behaviour is evaluated through plume-scale metrics such as contaminant mass or through well-specific concentration time series, a fundamental requirement remains the same: determining whether contamination is increasing, stable, or decreasing over time. Trend tests therefore play a central role in any assessment framework. Even when spatial or mass-flux based indicators provide valuable information about plume extent, mass removal, or source-zone behaviour, they cannot on their own establish the temporal direction of change. As a result, trend evaluation remains an essential component of interpreting monitoring data, independent of the broader analytical approach used.

4.1.3 Contaminant Metrics and Reporting Units

Apart from differences in analytical methods, the literature also varies in terms of which contaminants are evaluated and how concentrations are expressed. In current practice at WSP, concentrations are often assessed using PCE-equivalents (PCE_{eq}). This parameter is not widely used in the research literature; however, it appears in a small number of Danish studies. For example, Bøllingtoft et al. (2026) and Cremeans et al. (2018) report all concentrations as the sum of chlorinated ethenes expressed in PCE-equivalents (PCE_{eq}), allowing the mass contribution of individual compounds to be standardised and enabling a consistent representation of total chlorinated-solvent mass across monitoring rounds.

Newell et al. (2006), on the other hand, presents temporal trends for individual chlorinated solvent constituents, analysing each compound separately rather than aggregating them. Commonly referenced software tools, including GWSDAT, MAROS, and the Mann–Kendall Toolkit, evaluate temporal groundwater monitoring data using the imported concentration values and do not include functionality for calculating PCE-equivalents.

Across the reviewed literature, different reporting units were used to evaluate chlorinated solvent behaviour, and these choices influence how plume stability is interpreted. Most studies report concentrations in weight-based units (e.g. μg/l), reflecting regulatory standards. However, Ricker et al. (2020) highlight that weight-based concentrations can obscure degradation pathways because molecular weights decrease along the dechlorination chain. As shown by Ricker et al. (2020), summing weight-based concentrations may suggest attenuation even when the total number of molecules remains constant. Molar-based units preserve mass across transformation steps and therefore provide a clearer representation of degradation and commingling processes.

4.2 Results of Approaches

The most reoccurring methods in field of research was found to be time-series evaluation of individual wells (Section 4.1.1). Historically GMP time-series have relied on qualitative, visually based assessments, where trends are interpreted directly from time-series plots of well concentrations and supplemented by professional judgement (U.S. Environmental Protection Agency, 2005). This approach remains common in professional practice despite the availability of quantitative optimisation methods, as it is familiar, accessible, and does not require complex modelling. However, as noted in the literature, qualitative evaluations are highly dependent on individual expertise and may therefore lead to inconsistent interpretations (U.S. Environmental Protection Agency, 2005).

In contrast, quantitative long-term monitoring optimisation methods use statistical analyses to evaluate temporal patterns such as concentration trends. These approaches commonly rely on statistical trend tests such as the Mann–Kendall test or linear regression (Section 4.1.1). Observations from the dataset used in this study revealed pronounced skewness, heteroskedasticity, and irregular sampling intervals (Section 3.3.1), characteristics that violate key assumptions underlying ordinary least squares (OLS) regression. For this reason, the Mann–Kendall test was selected over linear regression for quantitative trend evaluation. By incorporating such quantitative tools, monitoring programs can move beyond subjective assessments and simple spreadsheet trend plots toward more defensible, data-driven decision-making.

As identified in Section 4.1.2, Ricker (2008) has created a method that aligns with the goal of the literature search. That is, finding techniques that build on evaluation of temporal trends and expand the perspective to site-wide contaminant observations. The choice to investigate this method specifically was tied to the characteristics of the method: it requires only a small amount of hydro-geological measurements which maintains a none-complex workflow.

4.2.1 Well-based Analysis

To examine differences between quantitative and qualitative trend assessments, two assessors (Person A and Person B) independently classified the trend for each well using individual PCE_{eq} time-series graphs, based on their interpretation of temporal variation, noise, and outliers (Appendix B.1). The results of this comparison are presented in Table 4.1. As shown in the table, 44 wells were assigned the same trend by both Person A and Person B. Notably, no well was ever classified as increasing by one person and decreasing by the other. In all cases of disagreement, one assessor classified the well as stable. This suggests that uncertainty is concentrated around distinguishing “stable” from “weakly increasing/decreasing”, rather than between the two directional trends themselves.

To illustrate the confidence behind the assessments, three certainty distributions (for decreasing, stable, and increasing trends) are shown in Figure 4.1. These graphs include only the wells where both Person A and Person B identified the same trend direction,

Table 4.1: Findings from visual assessment of individual well PCE_{eq} trends for all wells in all locations.

Number of wells	Person A: Decreasing	Person A: Stable	Person A: Increasing	**Total (Person B)**
Person B: Decreasing	16	8	0	24
Person B: Stable	2	20	1	23
Person B: Increasing	0	4	8	12
Total (Person A)	18	32	9	59

allowing the comparison to focus on differences in certainty rather than differences in trend classification.

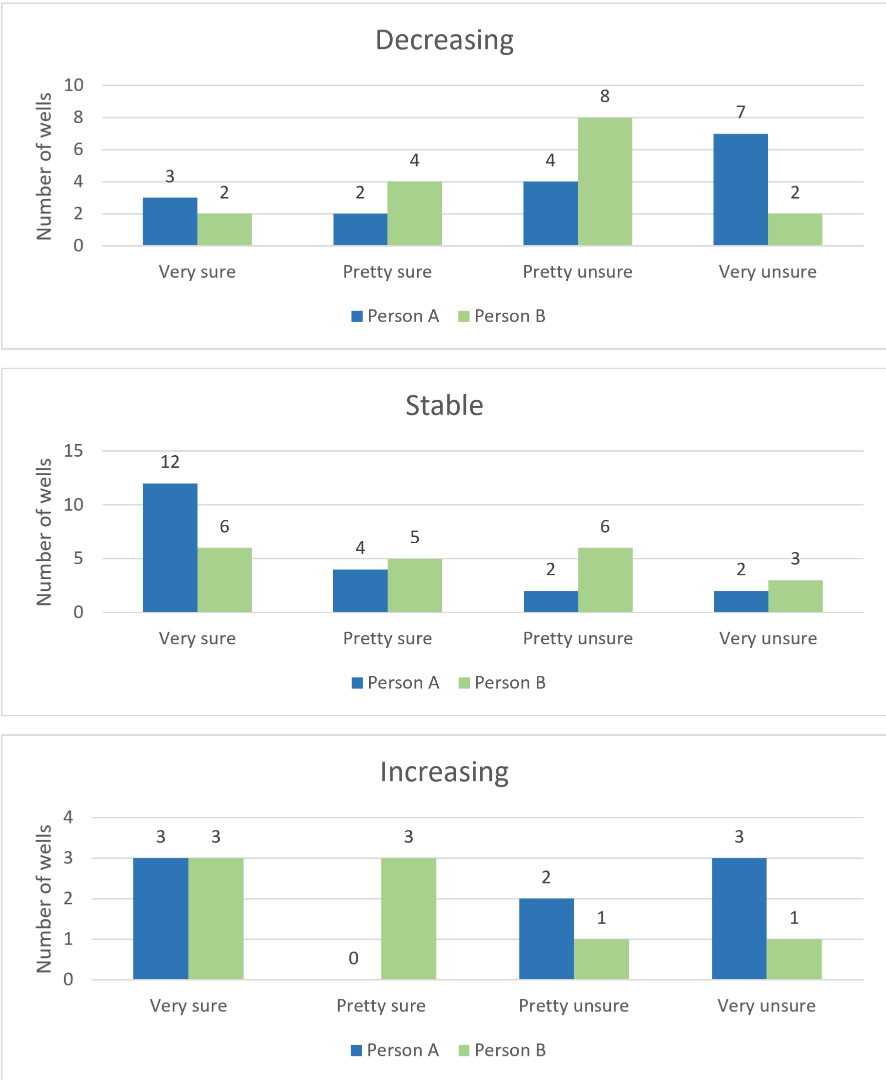


Figure 4.1: Assessors certainty distributions for visual trend classifications for decreasing, stable, and increasing trends. Includes only when Person A and Person B identified the same trend direction.

The distribution graphs show clear differences in how certain the assessors felt about their visual trend classifications. Trends identified as stable exhibit the highest overall uncertainty, with large share of wells falling into the “unsure” and “very unsure” categories. In contrast, wells classified as decreasing were more frequently associated with high confidence, indicating that downward trends were easier to identify visually. For increasing trends, the certainty distribution is more balanced, with uncertainty levels spread relatively evenly across the categories.

Even though Person A and Person B agreed on the trend for these wells, their certainty levels often differed. Across all wells with matching trend classifications, the assessors indicated different certainty levels 75% of the time for decreasing and stable trends, and 37.5% of the time for increasing trends. The single most uncertain category was "very unsure – stable", while the most confident category was "very sure – decreasing", highlighting that agreement on the trend does not necessarily imply agreement on the strength or clarity of the underlying signal.

There are five wells where both assessors were very sure of the trend (Table 4.2). In addition to that there are 7 wells were one has said very sure and the other somewhat sure (Table 4.3). These results indicate that although a large share of wells (almost 75%) were assigned the same trend direction, only a small proportion, 20% of all wells, or 27% of the wells with matching trends, were classified with any meaningful level of certainty ("somewhat sure" or "very sure"). This highlights that subjective visual interpretation introduces substantial uncertainty, even when the final trend classification appears consistent.

Table 4.2: Wells where both Person A and Person B were very sure of the trend classification.

Location	Well ID	Trend
A	A1	Decreasing
A	A2	Decreasing
C	C4	Stable
D	D4	Increasing
F	F11	Decreasing

Table 4.3: Wells where either Person A or Person B were very sure and the other somewhat sure of the trend classification.

Location	Well ID	Trend
B	B6	Probably Stable
C	C5	Probably Decreasing
C	C13	Probably Increasing
D	D1	Probably Decreasing
D	D6	Probably Stable
E	E15	Probably Decreasing
F	F7	Probably Decreasing

Subjective certainty is directly relevant for GMP decision-making, because misclassifying a well as stable when it is actually increasing may delay mitigation actions, while misclassifying a decreasing trend as stable may underestimate remediation progress or natural degradation processes. Understanding how confident assessors are in their visual interpretations is therefore essential for evaluating the reliability of this commonly used approach.

As a whole, these findings demonstrate the limitations of qualitative, visually based assessments and underscore the need for objective, reproducible trend-testing methods. The following section applies automated statistical trend tests to the same dataset to evaluate how their results compare with the subjective visual interpretations.

Using ProUCL, a Mann–Kendall based trend analysis was performed on all wells, assigning trend classes based on S-statistic direction, confidence factor, and CV (Appendix A.2). Across the datasets, 3 wells showed identical classifications, 6 wells showed agreement in trend direction, and 9 wells demonstrated a trend for only one of the Mann-Kendall or the visual assessment (Table 4.4). Red highlights in the table are potentially dangerous disagreements where assessor judgement could result in environmentally harmful decisions in the GMP.

Table 4.4: Comparison of trend significance from the Mann-Kendall test and assessors visual assessments for the wells where either method assigned high certainty of trend. Red colour highlights a potentially dangerous disagreement for environmental view, and '-' indicates 'no trend' (see Section 3.4.1).

Well ID	Mann-Kendall trend	Visual assessment
A1	Decreasing	Decreasing
A2	Decreasing	Decreasing
B1	Stable	-
B3	Probably increasing	-
B6	-	Stable
C5	Probably decreasing	Decreasing
C4	-	Stable
C12	Increasing	-
C13	Increasing	Probably Increasing
D1	Decreasing	Probably Decreasing
D4	Probably increasing	Increasing
D6	-	Stable
E15	Decreasing	Probably Decreasing
F1	Decreasing	-
F2	Decreasing	-
F7	Decreasing	Probably Decreasing
F9	Probably decreasing	-
F11	Decreasing	Decreasing

Scenarios that warrant particular caution from an environmental point of view are those where the visual assessment completely missed an increasing trend (B3 and C12), as these cases may represent emerging risks that qualitative inspection alone fails to detect. These situations illustrate how visual misclassifications can lead to delayed responses at wells that may already require remediation or further investigation.

B6 provides a contrasting example, in which assessors expressed a high confidence that the trend was stable, even though the statistical evaluation indicates that the dataset is too variable to support such a conclusion. Such overconfidence could lead to decisions to maintain or even reduce the monitoring programme, by keeping the same number of wells or lowering sampling frequency, when the underlying data may instead warrant increased attention.

In contrast, cases like B1 or F1, where the Mann-Kendall test classified the well as stable or decreasing, while the visual assessment did not assign a trend, illustrate a different type of impact. Here, visual uncertainty could support reducing sampling frequency or removing the well from the monitoring programme, offering potential cost savings without compromising environmental protection. These contrasting examples highlight that mismatches do not have uniform implications: some require precaution to avoid environmental harm, while others reveal opportunities for more efficient and economically optimised monitoring.

These discrepancies can also be understood in terms of hypothesis-testing (Section 2.2.3.1). Statistical trend tests only classify a trend when there is strong evidence to reject the null hypothesis of no monotonic change. Visual assessments, by contrast, have no formal decision rule and rely on subjective judgement. As a result, some mismatches occur because the Mann–Kendall test is conservative and avoids false positives (Type I errors), as seen in B6, C4 and D6. Other mismatches reflect potential false negatives (Type II errors) in the visual assessment, where emerging increases remain unnoticed, as in e.g. B3 and F1. Interpreting the disagreements in this way helps clarify whether they arise from statistical caution, subjective overconfidence, or limited visual sensitivity to weak trends. It should also be noted that Mann–Kendall may fail to detect weak or non-monotonic trends, particularly in short or highly variable time series.

4.2.1.1 Comparing Constituents Individually and as a Sum

Inspection of PCE_{eq} compared with the individual constituents shows that, in most cases, the curves follow a similar pattern, an example can be seen in Figure 4.2. This is often the case at sites without remediation, where the contaminants historically used at the site still dominate the contamination profile, causing the PCE_{eq} curve to follow a similar shape. However, a drawback of considering only PCE_{eq} is the limited understanding it provides of the behaviour of the individual constituents.

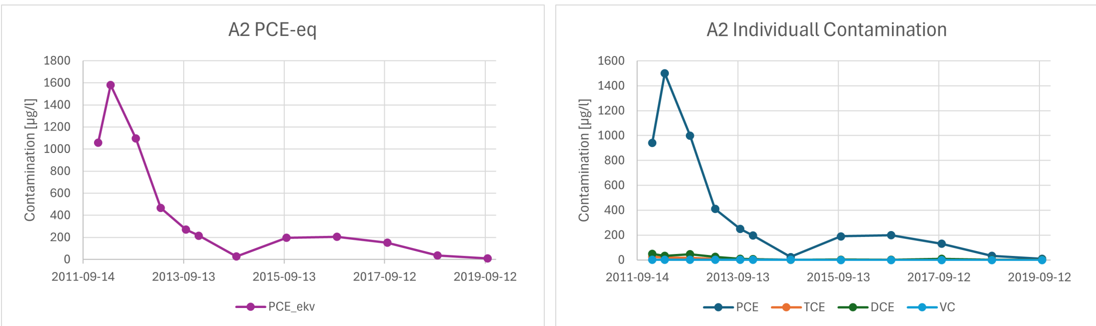


Figure 4.2: Comparison between PCE_{eq} concentrations (left) and individual chlorinated ethene concentrations (right) for well A2.

At Location B, wells B2 illustrate an important limitation of using only PCE_{eq} for trend analysis (Figure 4.3). In the final sampling event for both wells, PCE_{eq} decreases, whereas the individual constituent concentrations more clearly indicate that this decline is primarily caused by ND values. In fact, the PCE concentration in well B2 is increasing. Although all constituent concentrations in this case remain below guideline values, the example demonstrates a possible scenario in which trends in individual contaminants may be obscured when analysing only PCE_{eq}.

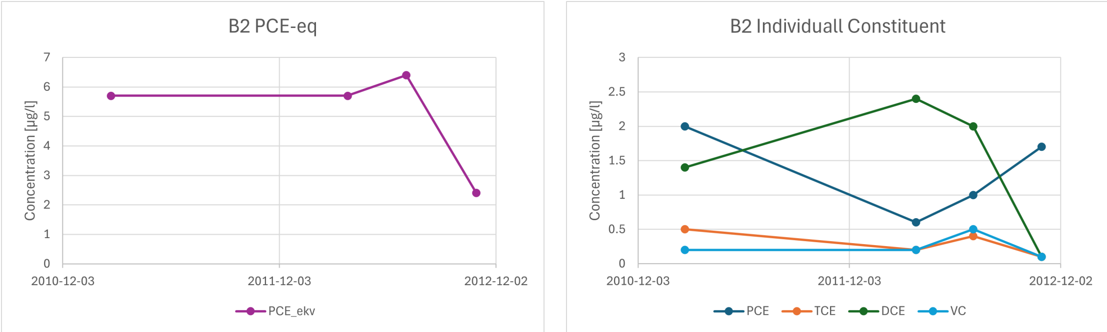


Figure 4.3: Comparison between PCE_{eq} concentrations (left) and individual chlorinated ethene concentrations (right) for well B2.

A similar pattern can be observed at well E6, where PCE_{eq} decreases as a result of declining PCE and TCE concentrations, while DCE and VC concentrations increase. Since the PCE_{eq} concentration remains above guideline values, it is unlikely that the well would be excluded from further monitoring. However, by analysing only PCE_{eq}, changes in the composition and behaviour of the contamination remain unidentified.

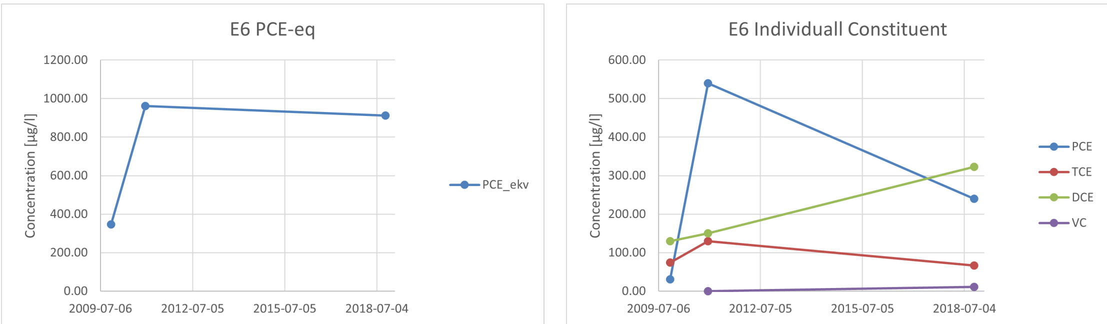


Figure 4.4: Comparison between PCE_{eq} concentrations (left) and individual chlorinated ethene concentrations (right) for well E6.

In Location D, at for example well D2, both the PCE_{eq} and the individual chlorinated ethenes show a clear decrease in measured concentrations over the past three years, indicating a downward development despite the absence of formal trend analysis. However, one important aspect is often overlooked: the ratio between the chlorinated ethenes. In D2, the proportion of DCE is increasing, meaning that although the summed concentration of chlorinated ethenes is decreasing, the composition of the plume is shifting toward daughter products with different regulatory thresholds. A comparable pattern is observed in wells F7 and E13 (Appendix B.3) where VC concentrations are increasing instead. Given that VC the most strict applicable guideline limit (Livsmedelsverket, 2022), even small increases may be of regulatory significance. These developments demonstrate that the overall toxicity profile of the plume may change and in some cases increase, even as total concentrations decline, underscoring the importance of monitoring constituent ratios, particularly in post-remediation settings where daughter-product formation is expected.

These examples illustrate that PCE_{eq} can mask changes in plume composition, which is central to evaluating whether the metric provides a reliable basis for monitoring de-

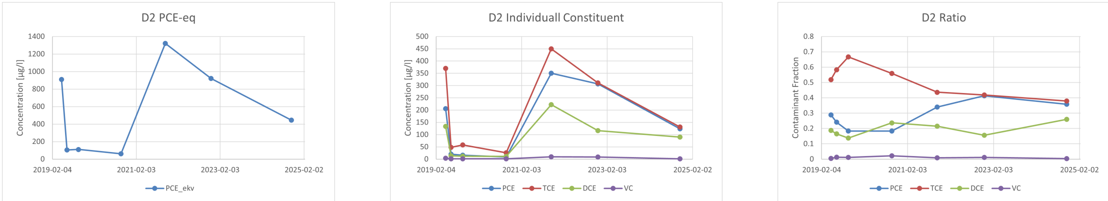


Figure 4.5: Comparison between PCE_{eq} concentrations (left), individual chlorinated ethene concentrations (middle), and constituent ratios (right) for well D2.

cisions. As a whole, these examples show that PCE_{eq} generally reflects overall concentration trends but may obscure important changes in plume composition, and shifts toward more toxic daughter products. Therefore, relying solely on PCE_{eq} risks misinterpreting plume behaviour, particularly in wells influenced by degradation processes or analytical censoring.

4.2.2 The Ricker Method

Figure 4.6 is an interpolation illustration acquired from the Surfer software, and is meant to give an idea of how these two dimensional plots can look. These are just examples of how contaminant plumes can vary for three points in time and how gridding is handled for data with varying well distribution. The optimal way to utilise this method would be to e.g. make a time lapse when evaluating how the plume behaves, suggestively with real-world maps as a backdrop. See Appendix C.1 for more interpolation maps.

The plume mass variations are plotted in Figure 4.7, which can be considered to provide an easier and more intuitive perception when compared to the multiple-well plots seen in Figure 3.3. Viewing plots for total plume mass also allow for incorporation of constituent plume masses, which can give insight of e.g. the total toxicity of the contamination. The VC plume can be seen to increase along with the total plume, which can be a sign of the functioning degradation of parent substances, but also an indication of a higher risk level since VC is considered more toxic than the other compounds. Further benefits of observing individual substance behaviours, which also apply for site-wide investigations, are discussed in Section 4.2.1.1

Also, viewing graphs for plume area, average plume concentration, and total plume mass together can contextualise the subsurface situation in new ways (Appendix C.2). For location C, all recent and consecutive sampling dates were plotted in Figure 4.8 as an example. Average concentration seems to have reached a stable trend at high values, but at the same time total plume mass indicates a decrease and plume area seems stable or even slightly decreasing. Although, the planar area is influenced by the number and placement of wells, and inadequate GMP layout can result in constrained results such as estimated planar area being equal to the total area of the grid (Appendix C.3). This effectively means that the plume could be larger in reality, and since both average concentration and plume mass calculations rely on the area, all results are compromised. The general method is therefore insightful, but the risk for inaccurate results are high for these limited datasets, especially if interpolation produces unstable grids.

Since kriging was recommended by a guide (Golden Software, 2023) that can reasonably be assumed to be created with larger datasets in mind, tests aimed at evaluating other interpolation options were carried out. The techniques tested were kriging, near-

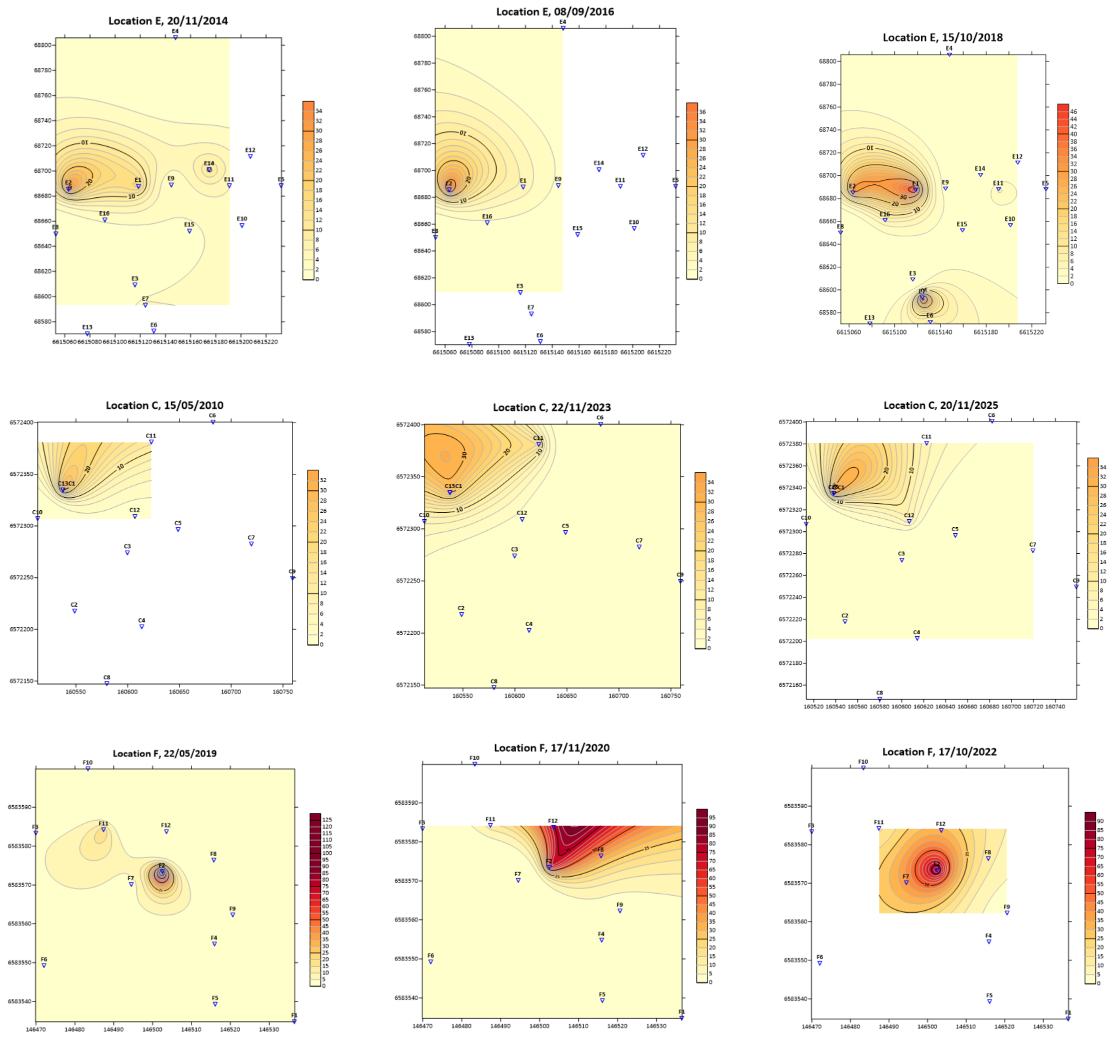


Figure 4.6: Interpolation illustration using kriging for Location E (first row), Location C (middle row), and Location F (last row) for three dates.

est neighbour, and natural neighbour, and as can be seen in Figure 4.9 they all yield very different results in total plume mass. Kriging and nearest neighbour behave in similar ways; both plume mass estimations increase and decrease for many of the same dates. Natural neighbour estimates quite small plume areas and volumes, resulting in unreasonably low total plume mass values, an indication of failures in gridding performance. For location F, the largest plume masses calculated (expressed in PCEeq weights) for kriging, natural neighbour, and nearest neighbour was 0.99kg, 0.23kg, and 2.84kg respectively, see Figure 4.9. It is apparent that the choice of interpolation method will have severe affect on the perception of plume size for monitoring programmes of this size.

The graphs in Figure 4.10 suggest that the nearest neighbour method tend to always yield the highest result out of the three investigated interpolation techniques, and natural neighbour the lowest. This is partly believed to be a consequence of the characteristics of the monitoring programmes, but also a direct correlation to the nature of the techniques. Nearest neighbour, for example, involves simple assumptions explained in Section 2.3.2, which often causes areas to receive high average concentration estimations. Similar distributions could therefore be found in monitoring programmes with a

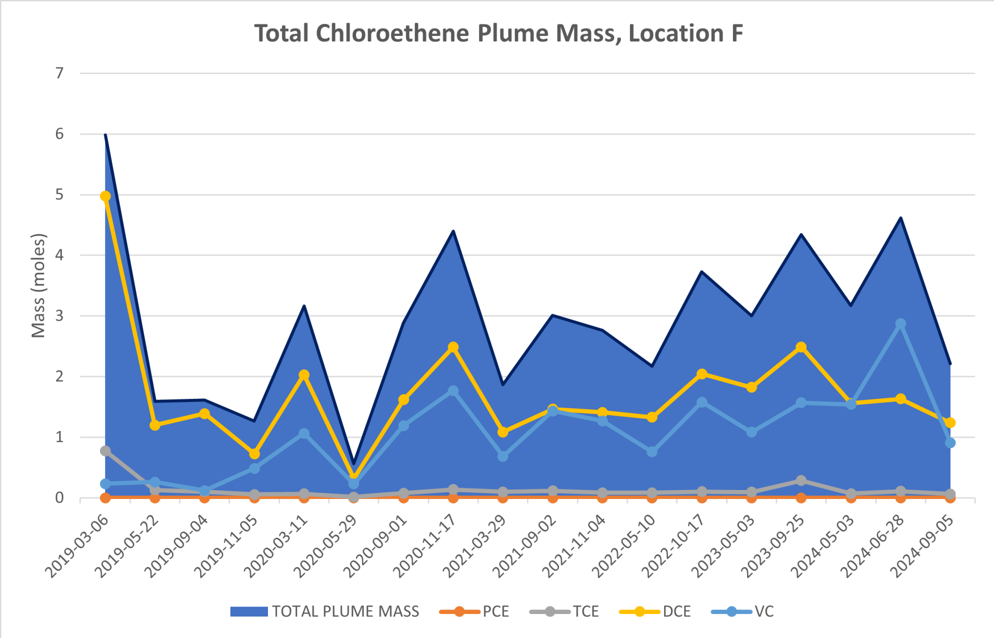


Figure 4.7: Total and individual chlorinated ethenes plume masses over time for Location F.

significantly larger number of wells, but most likely with the correlation that increasing the number of wells decreases plume mass differences between interpolation methods.

To try to determine which interpolation method most likely resembled real plume behaviour the closest, cross validations were examined for the log-transformed data. Because Surfer only performs cross validation on the raw data before applying its own transformations, the dataset was manually log10-transformed prior to upload. Natural neighbour could only produce estimations and errors for a few wells for each date, always for under half of the sampled wells and sometimes none at all. This is believed to be a consequence of the technique's subset assumptions, and is interpreted as further proof of the unsuitability for these types of datasets. It was therefore excluded for further cross validation observations.

The RMSE values for kriging and nearest neighbour indicate that kriging is the better plume predictor for almost all dates for Location F (Appendix C.4). Only three dates are showing more favorable RMSE values for nearest neighbour, although this is most likely influenced by the fact that only 3 sampling wells are present and have specific placements that are favorable for the high estimations that nearest neighbour produces. Neither technique is yielding RMSE values less than 20% of the range for any dates, but kriging is showing errors that more often fits into an acceptable range (20-40%) (Appendix C.4) Conclusively, the dataset is too limited for cross validation to confidently support either interpolation method, but kriging is more often better fitted.

As explained in the method (Section 3.4.2) the default linear variogram was chosen for all kriging interpolations for all datasets. The variograms show a very weak spatial structure with high variance and no clear spatial correlation range, indicating that the dataset is dominated by noise or outliers and is poorly suited for kriging (Appendix C.5). Since vast understanding of variograms is needed for proper modelling, as well as knowledge of various dataset needs, the default linear variogram model was deemed to

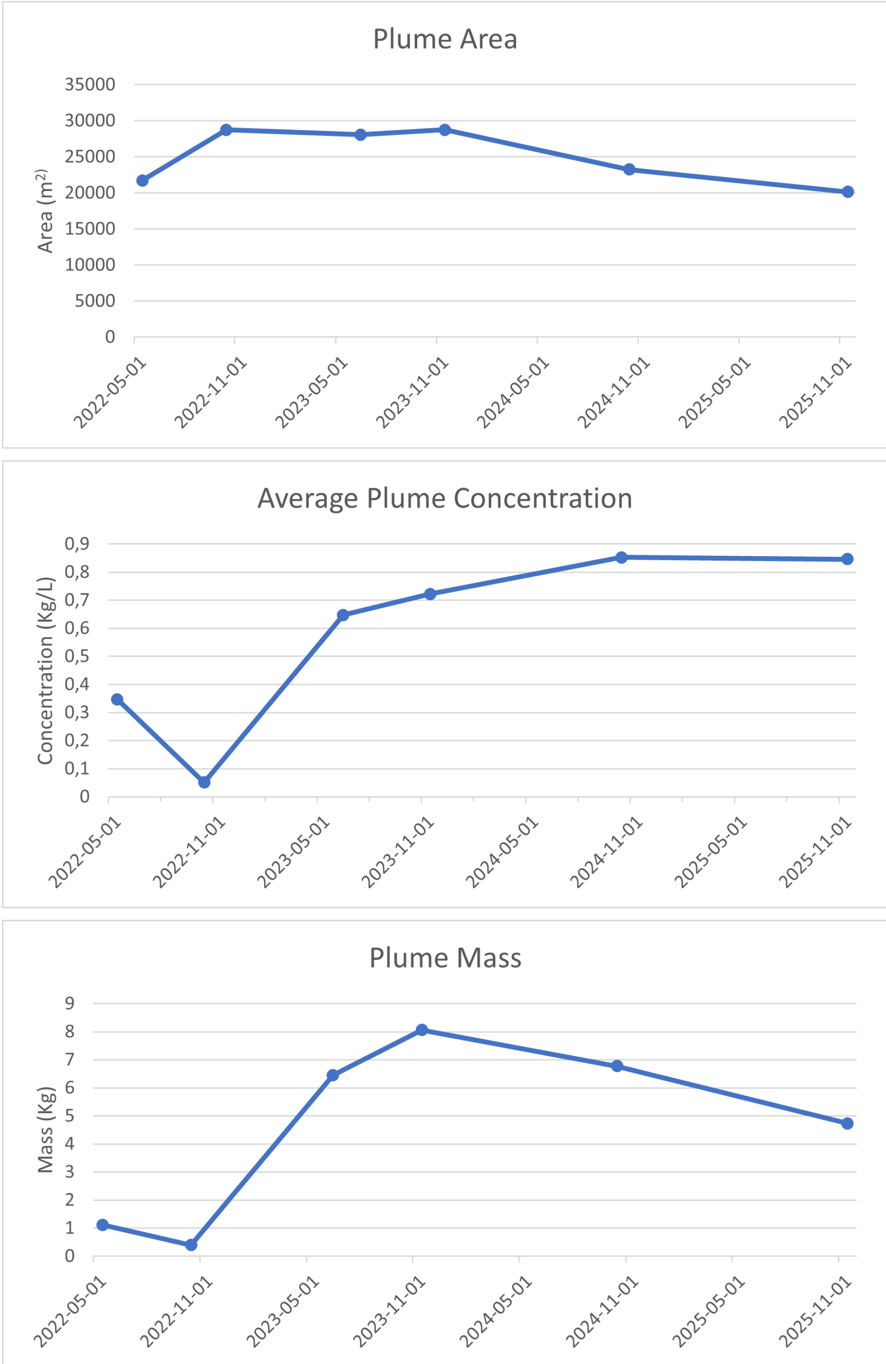


Figure 4.8: Graphs for plume area, average plume concentration, and plume mass for sampling dates from 2022 at Location C.

be enough for the surface level tests of this thesis. Also, the conclusion for the datasets in general is that the small number of measurements for each date impairs any ability of variogram modelling to provide significantly improved results.

4.2.2.1 Scenario Results

Tests were conducted to evaluate robustness of interpolation methods as well as importance of data-point density via theoretical scenarios explained in Section 3.4.2.1. Figure 4.11 indicates that removal of a well (F7) close to the contamination source (F2)

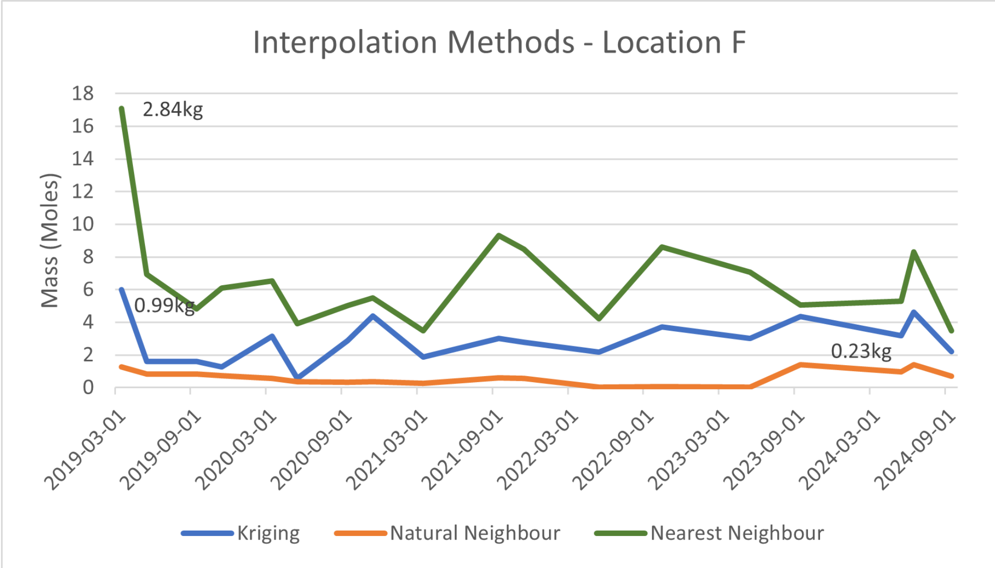


Figure 4.9: Plots of calculated plume masses using kriging, natural neighbour, and nearest neighbour for Location F. Each plots peak mass is included in the graph, expressed in kg of PCEeq

Figure 4.10: Calculated plume masses using kriging, natural neighbour, and nearest neighbour for Locations B, C, E, and F for three dates each.

severely affects any gridding methods ability to predict the spread. The 'F7 Removed' scenario reveals that density of wells surrounding the source can greatly improve plume mass estimations even if they exhibit low concentrations, and can arguably be considered more important because of it. From 2021-09-02, F7 was removed in the real monitoring programme, most likely due to effective remediation. This is common prac-

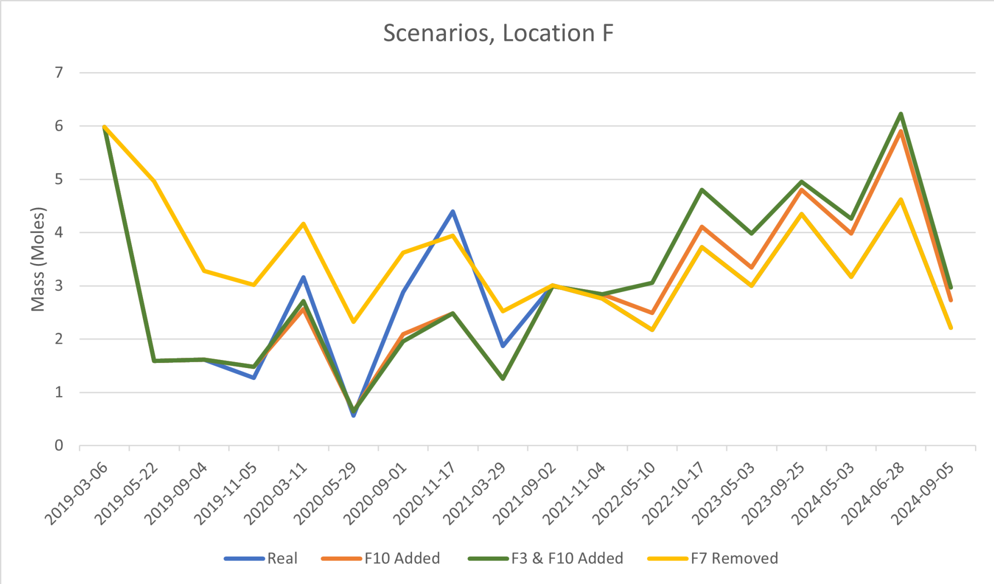


Figure 4.11: Plume mass calculations for different scenarios at Location F using the kriging interpolation method.

tice in order to save time and money, and although previous measurements can be used for e.g. an average value in interpolation, there is risk of wrongful plume variations being interpreted.

Similarly, testing the effect of re-adding representative measurements for wells that had been removed resulted in noteworthy plume mass differences. The wells investigated for this scenario were instead perimeter-wells, i.e. sampling points meant to verify plume cutoff. Like F7, F10 was most likely removed due to low concentrations, however, F3 exhibited concentrations far above recommended levels (2.1.2). Generally, the plumes for scenario 'F10 Added' and 'F3 & F10 Added' were adjacent in size. 2020-03-11 and 2020-09-01 are dates where neither F3 or F10 were sampled, and the scenarios indicate that there is not a significant difference in gridding results between only adding a perimeter-well with low concentrations or also adding one with higher measured values. These two sampling dates have 6 well measurements before the adding-scenarios, which is most likely the reason why the comparison differs for dates after 2021, since the real dataset only has 3 well measurements for most of those sampling times. This highlights the increased importance of keeping/adding wells the fewer a monitoring programme has. It should also be stated that F3 and F10 are located in the same general direction from the source well F2, and instead adding two wells in opposite directions would expectedly show greater differences between scenarios.

4.2.3 Overview of Approaches

The trend significance for total plume mass and singular well concentrations underscore the variability in certainty of the Ricker Method depending on location. Although the datasets are alike, some have proven to produce plume mass calculations that more reasonably align with its belonging well measurement trajectories over time. For example, 10 wells at Location C show non-significant trends, two wells (C12 and C13) display statistically supported increases, and one well (C5) a probable decrease (Table 4.4). This correlates with the associated total plume mass which is showing a significant trend at 90% confidence, but with an uncertain Sen Slope interval which fails to detect the trend direction. Meanwhile, Location F site-wide plume mass calculations suggest an increase in total plume, while the only wells with any statistically supported significance (F1, F2, F7, F9, F11) are showing decreasing trends. Even though the Mann-Kendall test shows above 90% confidence coefficient for the plume mass trend, this more likely is a result of poor gridding performance resulting in non-trustworthy estimated plume behaviour, especially during later dates when the number of wells are limited.

Inspecting the constituent sum PCE_{eq} can give a simple and clear first impression of a sampling wells trend, but should always be accompanied by further viewpoints. The most easily obtained while still providing significant insights being single constituent variations and ratio changes. This combination is useful especially for high concentration wells, e.g. near the contamination source, to understand if the spots posing the largest risk are still exhibiting an increase in threat levels. The clear downside to this well-by-well methodology is the difficulties in conveying behaviour of the entire site-wide plume, since it does not necessarily follow a clear trajectory that can be illustrated by plotting well measurements over time. This is especially emphasised for datasets containing irregular sampling frequency, since the perception becomes tied to scattered circumstances during various conditions.

Although the Ricker Method showed some clear issues when faced with sparse data, both temporally and spatially, the results that could be obtained from it still provided some insights where the ordinary well-by-well workflow could not. Following the minimal-input modelling used in gridding still allows for a simplistic procedure for career professionals, however, potentially time consuming. As discussed in section 4.2.2, viewing total plume mass along side individual substance plume mass can provide key insights into toxicity variations for entire sites, and observing plume area, average concentration, and plume mass variations gives a combined and intuitive understanding of plume behaviour.

If Ricker Method calculations are to be made for a monitoring programme with a constrained amount of wells, kriging will most likely produce the closest real-situation plume masses, however, with expected large uncertainty. Since nearest neighbour yields a form of worst-case-scenario perception, both methods could be used to gain an understanding of the variation range in plume mass for any given point in time. An option could also be to not place too much trust in the actual plume mass quantities but instead focus on behaviour over time. Viewing the increases and decreases of the site wide plume in addition to constituent behaviours of individual wells could provide greater contamination spread insights, and therefore be the basis for more well-founded decision making.

4.3 Software Findings

The most promising software were identified through key scientific publications on groundwater monitoring. One of the most notable findings from this stage of the project was that, contrary to our initial hypothesis, the number of dedicated time-series visualisation and temporal trend-detection tools for groundwater contamination was very limited. In practice, it appears far more common to rely on general-purpose statistical environments, particularly R and Python for trend analysis.

According to Holbert (n.d.), R has become the preferred computing environment for a substantial portion of the statistical community, and it's use as an industry standard in data analytics continues to expand. Moreover, R is not only widely adopted within statistics, it has also been embraced enthusiastically within the environmental sciences (Jones et al., 2014). The strength of R lies in its high degree of flexibility, enabling users to extend the language almost without limitation through packages that integrate seamlessly and function as if they were part of the native base system (Holbert, n.d.). Despite its widespread adoption and analytical strength, R was excluded from further evaluation because the thesis applies inclusion criteria that restrict the analysis to GUI-based tools suitable for users without experience in code-driven environments.

4.3.1 Selected Software

In early stages of this thesis work two potentially promising software were provided for us, GWSDAT and MAROS. GWSDAT met all essential requirements while MAROS failed at the technical requirements by not being actively maintained. Therefore, MAROS was excluded from further evaluation. In the end, the three software chosen were GWSDAT, ProUCL, and AquaChem.

Ground Water Spatiotemporal Data Analysis Tool (GWSDAT) is an open-source software developed through a collaboration between Shell and the University of Glasgow for analysing and visualising trends in groundwater quality monitoring data. It is available as an online tool and an Excel Add-in, but in this project only the Excel Add-in is evaluated. All data are processed locally, as stated in the document: "All data sets are retained and analysed locally with no information being sent over the internet" (Jones et al., 2026). The core feature of GWSDAT is its spatiotemporal model, which simultaneously estimates contaminant concentrations in space and time and provides suggestions for well-removal.

ProUCL is an open-source statistical software package developed by the U.S. Environmental Protection Agency (2022) for environmental data analysis, with a focus on datasets containing non-detects, outliers, and skewed distributions. It provides a wide range of parametric and non-parametric methods, including Upper Confidence Limit calculations, trend tests, and goodness-of-fit evaluations. ProUCL automatically identifies appropriate statistical models, supports robust handling of censored data, and generates summary tables and plots suitable for regulatory decision-making. It is widely used for groundwater monitoring, risk assessment, and compliance evaluations.

AquaChem 14.0 is a Windows-based geochemical commercial software developed by Waterloo Hydrogeologic (2026) for water quality and groundwater monitoring projects. The software supports environmental investigations by providing tools for groundwater contamination assessment, trend analysis, visualisation, and water chemistry modelling. One of AquaChem's main strengths is its focus on practical analytical tools specifically relevant to GMP.

4.3.2 Results of Functional Capabilities

First the functional capabilities were identified, the result are presented in Table 4.5. When comparing the functional capabilities of the selected software, it becomes clear that GWSDAT meets more of the desirable features identified by industry professionals. Both GWSDAT and AquaChem support the three highest-ranked functions from the survey (guideline highlight, handling NDs, GIS map import) (Appendix D.1), which strengthens their relevance for practical groundwater assessments. However, the tools differ in how they implement certain features. One example is the variation in ND methodologies. GWSDAT and AquaChem apply the commonly used substitution approach, whereas ProUCL offers more advanced statistical methods such as Kaplan–Meier and Regression on Order Statistics. These approaches can provide more robust estimates when the proportion of ND values is high.

Function	**GWSDAT**	**ProUCL**	**AquaChem**
Multiple trend tests	Yes	Yes	Yes
Indicates if trend is significant	Yes	Yes	Yes
Handling of ND's	Yes	Yes	Yes
Identification of outliers	No	Yes	Yes
GIS map import available	Yes	No	Yes
Spatial well removal support	Yes	No	No
Unit conversion	Yes	No	Partly
Interpolation	Yes	No	No
Guideline highlights	Yes	No	Yes
Mass-flux features available	No	No	No
Targeted at chlorinated solvents	Yes	No	Yes
Price	Free	Free	Cost

Table 4.5: Comparison of software functions between GWSDAT, ProUCL, and AquaChem with colour-coded feature availability.

Another distinction not captured in the table is the way each software displays guideline highlighting. GWSDAT is the only tool that presents all contamination guidelines within a single, consolidated table. Guideline highlighting was rated as the most important feature in the survey, as it enables users to quickly identify exceedances relative to regulatory thresholds. In addition, both GWSDAT and AquaChem support GIS map import. However, their purpose is very different since GIS maps in GWSDAT act as a back drop for interpolation, while AquaChem utilises them as illustrative tools for reports. Unit conversion was marked as "Partly" for AquaChem due to the feature did not function as intended.

None of the evaluated software packages performed mass-flux calculations. Although this was not identified as a priority by the industry professionals in the survey, mass-flux estimation is becoming increasingly common in contemporary research and site-management strategies, as noted by recent studies such as Bøllingtoft (2025). This suggests that mass-flux capability may become a more sought after feature in future software evaluations.

4.3.3 Results of Quantitative Outputs

The quantitative comparison in Table 4.6 highlights clear differences in the statistical depth offered by the three software tools. While all programs provide the basic Mann–Kendall S-value, the range of additional trend-related metrics varies substantially. GWSDAT delivers only the minimum required output, whereas ProUCL expands this with standardised value of S, approximate p-values, Sen's slope, and corresponding confidence intervals. AquaChem, however, provides the most extensive suite of quantitative indicators, including seasonal Mann-Kendall, Spearman correlation, and multiple Mann-Kendall based statistics. This broader output enables a more nuanced interpretation of trend behaviour.

Table 4.6: Quantitative output capabilities of GWSDAT, ProUCL, and AquaChem.

Function	**GWSDAT**	**ProUCL**	**AquaChem**
OLS	-	Yes	Yes
OLS p-value	-	Yes	Yes
R^2	-	Yes	Yes
Local linear regression	Yes	-	-
Mann-Kendall statistic	Yes	Yes	Yes
Standardised Mann-Kendall statistic	-	Yes	Yes
Mann-Kendall tau	-	-	Yes
Mann-Kendall Seasonal	-	-	Yes
p-value (tabulated)	-	Yes	-
p-value (approx.)	Yes	Yes	Yes
Sen's slope	-	Yes	Yes
Confidence interval for Sen's slope	-	Yes	Yes
Spearman	-	-	Yes

When evaluating the trend significance, the outputs were assessed using PCE_{eq}. Trend significance was evaluated using the Mann–Kendall P-value with a 95 % confidence level. For GWSDAT and AquaChem p-value (approx.) is used, whereas in ProUCL p-value depends on the number of sampling points. In Table 4.7, the statistical significance of each well was evaluated for all three programs.

Table 4.7: Summary of Mann–Kendall test results using GWSDAT, ProUCL, and AquaChem.

Software	Increasing	Decreasing	Not significant
GWSDAT	2	6	51
ProUCL	2	8	49
AquaChem	2	7	50

Table 4.7 demonstrates the differences in the software's significance identifications. For Location E, well E15 shifts from not statistically significant in GWSDAT to a significantly decreasing trend in both ProUCL and AquaChem. For Location F, well F1 is not statistically significant in GWSDAT or AquaChem but becomes significantly increasing in ProUCL. All remaining wells show consistent classifications across the programmes, indicating that differences in trend identification are limited relative to the total number of wells.

Furthermore, to compare the software in more detail, the Mann–Kendall S-values and p-values were evaluated (Appendix D.5). When considering the full dataset, one well, A1 at Location A, shows a difference in S-value across all programmes. However, for Location F, 6 of the wells showed different S-value for AquaChem, while GWSDAT and ProUCL calculated the same result. For the p-values, none of the wells produce identical values across all three software. However, for 7 wells the p-values are exactly the same in ProUCL and AquaChem. When the p-values are rounded to two decimals, an additional 10 wells show matching values. To illustrate the variation in p-values between the programmes, a box plot was created (Figure 4.12).

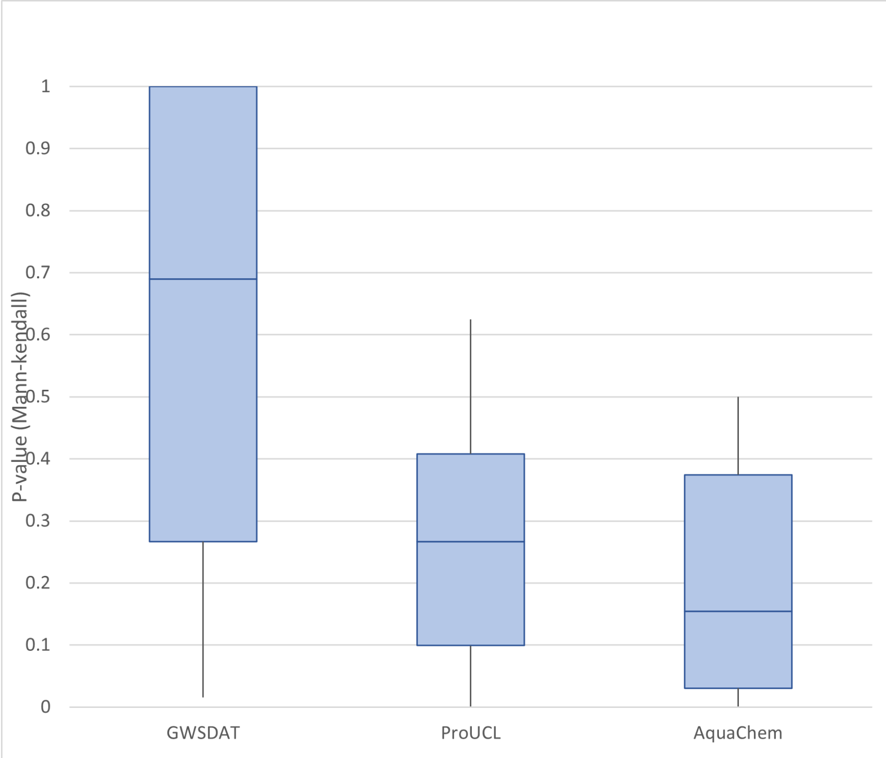


Figure 4.12: Variation in Mann-Kendall p-value for GWSDAT, ProUCL, and AquaChem for all wells in all locations.

As shown in Figure 4.12, GWSDAT tends to generate higher Mann–Kendall p-values. In fact, GWSDAT reports the highest P-value in 51 of the 59 trends. Higher p-values should logically result in fewer significant trends, and GWSDAT does produce the highest number of “not significant” wells. However, the difference in the number of significant trends between the programmes is surprisingly small. This raises the question of whether GWSDAT increases p-value across the entire range or mainly amplifies p-values that are already high.

To investigate this, the distribution in p-values was compared for all wells identified as significant in at least one of the programmes in Figure 4.13. ProUCL consistently has the lowest p-values across all summary statistics. AquaChem lies slightly higher. GWSDAT shows the highest overall distribution, with both a higher median and a wider spread. Its minimum is fixed at 0.01 because GWSDAT reports all p-values below this threshold as 0.01. In the boxplot, these censored values appear at the lower bound, which slightly compresses the lower tail of the GWSDAT distribution. The higher median and wider spread for GWSDAT support the conclusion that GWSDAT shifts p-values upward across the full range, rather than only amplifying already high values.

The results confirm the overall pattern. For all wells in which all three programs identify a significant trend, GWSDAT consistently reports the highest p-value. For borderline cases, the behaviour differs between wells. For E15, GWSDAT’s higher p-value shifts the classification from significant to not significant while both ProUCL and AquaChem still identify a decreasing trend. For F1, only ProUCL identifies a sig-

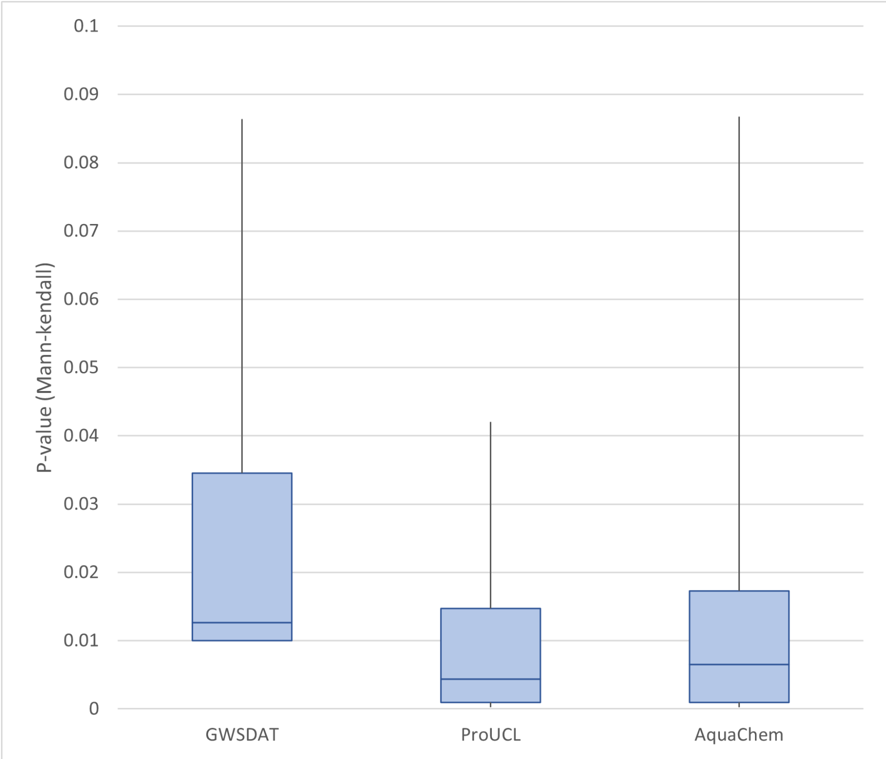


Figure 4.13: Variation in Mann-Kendall p-value for GWSDAT, ProUCL, and AquaChem for wells that were identified as having a significant trend by at least one of the software.

nificant trend, whereas both GWSDAT and AquaChem classify the well as not significant, with AquaChem reporting the highest P-value. The underlying reasons for these differences in S-values and p-values remain unclear.

4.3.4 Results of Subjective Criteria

In the MCA all subjective criteria were scored on a –5 to +5 scale, where 0 indicates neutral performance relative to each respondent's current workflow, +5 represents the best possible performance, and –5 the worst. For each criterion, the individual scores from all respondents were averaged and then multiplied by the corresponding criterion weight to obtain the weighted score. These weighted scores were summed to produce the total performance score for each software (see Table 4.8).

GWSDAT performs best in the final ranking. However, all total weighted scores are close to zero, indicating that none of the software solutions are assessed as particularly strong or weak overall. Instead, the ranking reflects relatively small differences in how each tool performs across the evaluated criteria.

To illustrate which criteria have the greatest influence on the performance ranking, the weighted scores for each software are shown in Figure 4.14. In the figure, each bar represents the weighted score for a criterion, positive bars indicate criteria where the software performs better than current practice, while negative bars indicate poorer performance.

The figure shows that GWSDAT's strongest contribution comes from guideline values. GWSDAT also performs consistently well across several plot-related criteria, including

Table 4.8: Final weighted score results from the MCA with assigned performance rankings.

Criteria	Weight	GWSDAT		ProUCL		AquaChem	
		Score	W. Score	Score	W. Score	Score	W. Score
Initial time investment	12%	-0.33	-0.04	0.67	0.08	-2.67	-0.31
Data modifiability	19%	0.00	0.00	-0.67	-0.13	-0.33	-0.06
Plots – Features	4%	1.67	0.07	-1.00	-0.04	0.83	0.03
Plots – Easy to produce	3%	1.17	0.03	1.50	0.04	0.50	0.01
Plots – Quality	4%	2.00	0.09	-1.83	-0.08	0.67	0.03
Plots – Add/remove series	5%	1.67	0.08	0.17	0.01	0.33	0.02
Plots – Multiple Plots	4%	0.00	0.00	-0.67	-0.02	1.17	0.04
Statistical measures	7%	-0.17	-0.01	2.83	0.21	1.83	0.14
Table	17%	-0.67	-0.12	-0.67	-0.12	-1.33	-0.23
Guideline values	11%	2.17	0.24	0.83	0.09	0.67	0.08
Intuitiveness	13%	1.33	0.18	0.83	0.11	-1.67	-0.22
Total	**100%**		**0.53**		**0.15**		**-0.48**
Performance Rank		1		2		3	

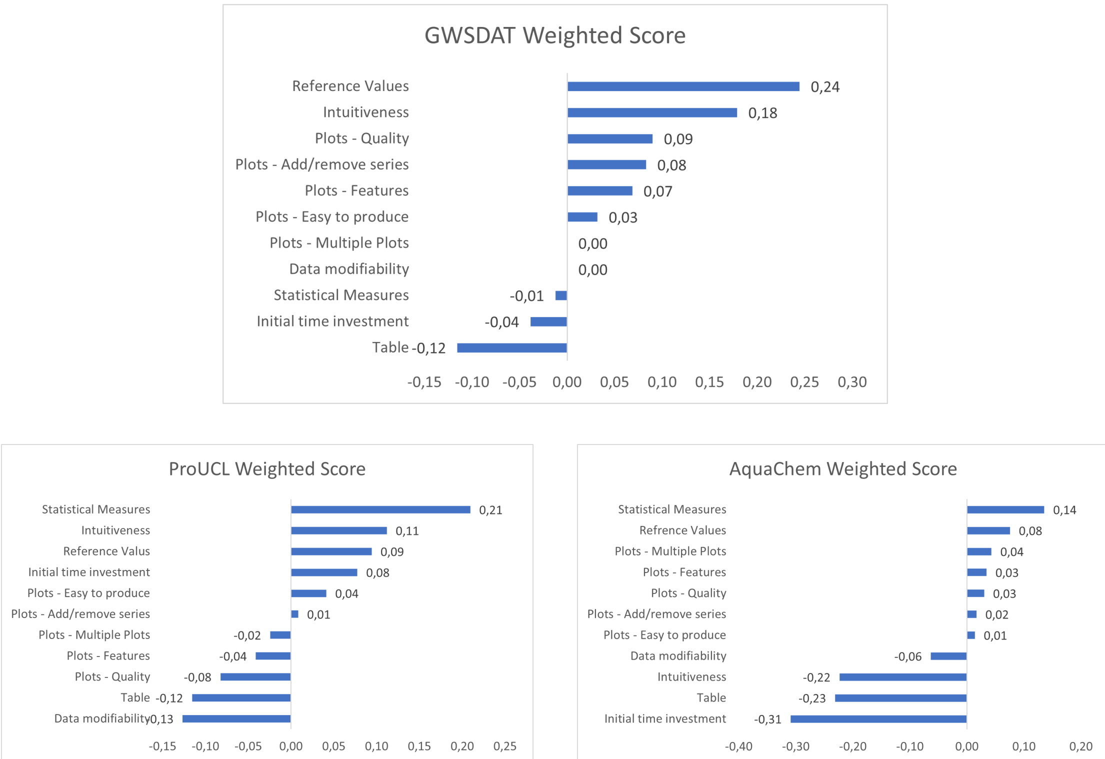


Figure 4.14: Weighted score for GWSDAT (top left), ProUCL (top right), and AquaChem (bottom).

plot quality, features, and the ability to add or remove data series. Two criteria, Plots Multiple Plots and Data Modifiability, are neutral, indicating that respondents saw no meaningful difference compared to their current workflow. The negative contributions are concentrated in three areas: table handling, initial time investment, and statistical measures. Among these, table handling has the largest negative impact, suggesting that users found GWSDAT less efficient or intuitive for tabular data review.

ProUCL ranks second in the MCA, with a total score that is positive but close to zero. This indicates that respondents perceive the software as offering some advantages over current practice, but these strengths are balanced by several notable weaknesses, which can clearly be seen in Figure 4.14. The results is an almost even distribution of positive and negative weighted scores. The statistical measures are the strongest positive contributor to ProUCL's total score. This aligns with the software's design focus on advanced statistical analysis.

The negative contributions are concentrated in table handling, data modifiability, and plot quality. Table handling and data modifiability are the two most influential negative criteria, indicating that respondents found ProUCL less flexible for editing or reviewing data compared to their current workflow. Plot quality also contributes negatively, the visual output is perceived as less effective than in other tools.

AquaChem ranks third in the MCA, with a total score that is negative, indicating that respondents generally perceive AquaChem as performing worse than current practice across several criteria. Figure 4.14 shows that initial time investment has the largest and most negative influence on the total score, meaning that respondents found AquaChem demanding to set up and use compared with their existing tools.

AquaChem performs moderately well on several plot-related criteria, including plot features, quality, and the ability to add or remove series, as well as extra statistical measure and guideline values. These positive contributions suggest that users appreciate the software's visualisation capabilities once data are properly configured. However, despite having the largest number of criteria contribute positively, the magnitude of the negative criteria is large enough to pull the total score below zero.

4.3.4.1 Sensitivity Analysis of the MCA

The first sensitivity test was targeted at the proportional weight variation. In Figure 4.15, the change in total software score when each individual criteria weight increase $+10\%$ are shown. As can be seen changes in the final score are minimal for all weight changes across each of the three software. Initial time investment has the largest overall affect, specifically on AquaChem, but even that is contained to a minor score variation of -0.14 compared to the score scale. When decreasing the individual criteria weight the graph is mirrored.

The performance rank when testing for individual scores are shown in Table 4.9. These results illustrate that the ranking of the three software tools varies depending on which respondent's scores are used.

Table 4.9: Performance rankings of GWSDAT, ProUCL and AquaChem when only accounting for one persons input at a time, for three out of five evaluators.

	Performance Rank		
	GWSDAT	ProUCL	AquaChem
Person A	1	3	2
Person B	1	2	3
Person C	3	2	1

This clear variation between respondents indicates that the scores are less robust than the criterion weights. When combined with the proportional weight-change sensitivity

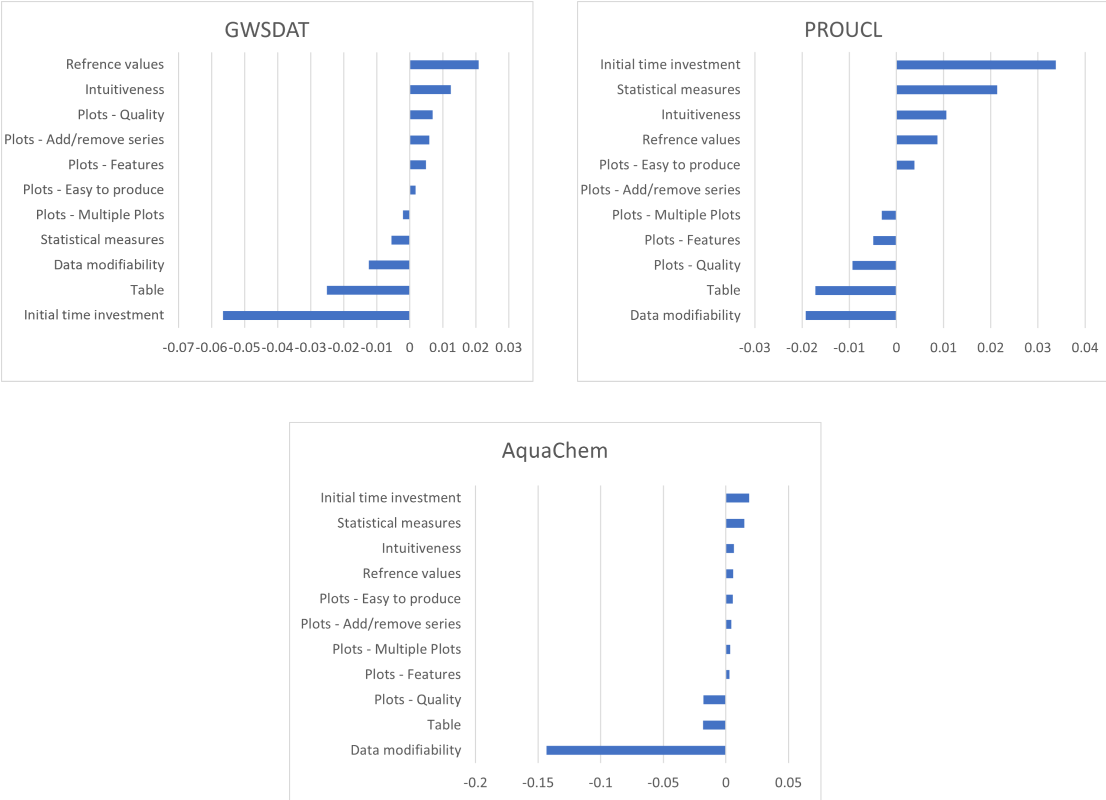


Figure 4.15: Effect on total software score from a 10% increase in the weighted score of each individual criterion, for GWSDAT (top left), ProUCL (top right), and AquaChem (bottom).

analysis, the findings suggest that the final MCA outcome is more sensitive to individual opinions than to moderate changes in weighting. In other words, the personal preferences and experiences of each respondent has a strong influence on the resulting software ranking. At the same time, the individual scoring results provide valuable insight into the differences in expectations and priorities among professionals working within the same field. Despite all working with GMP evaluation, the respondents evaluate the software tools differently, highlighting the diversity of user needs and preferences that an MCA must account for.

The results from the absolute weight shift test are presented in Table 4.10. AquaChem shows improved performance in all three scenarios, with the largest increase occurring when the weight of initial time investment is reduced. This is expected, as AquaChem requires more time to initiate a project compared with the other tools. ProUCL, in contrast, tends to decrease in total weighted score, reflecting its strong dependence on the criteria related to time investment and intuitiveness. ProUCL is therefore the most sensitive to absolute weight shifts, with changes ranging from –33% to +100%. GWSDAT remains comparatively stable, with changes between –11% and +25%.

Table 4.10: Results from the absolute weight shift test for GWSDAT, ProUCL, and AquaChem. Green cells indicate an increase from the original software score, and red a decrease.

	Final weighted score		
	GWSDAT	ProUCL	AquaChem
Orginal	0.53	0.15	-0.48
Decreasing time investment	0.66	0.10	-0.19
Decreasing intuitiveness	0.47	0.08	-0.30
Increasing easy to produce plot	0.57	0.30	-0.42

4.3.5 Final Software Recommendations

The overall conclusion from the software evaluation is that none of the assessed tools can be considered fully comprehensive for evaluating time-series. When examining the three complementary evaluation areas: (i) functional capabilities, (ii) quantitative outputs, and (iii) subjective scoring using a MCA, it becomes clear that GWSDAT performs strongest in functional capabilities and subjective scoring, whereas ProUCL and AquaChem provide more extensive quantitative statistical outputs.

GWSDAT meets more of the desirable features identified by industry professionals. Importantly, it is the only tool offering spatial interpretation of plume behaviour. A limitation of GWSDAT is the relatively shallow statistical output. Broader statistical reporting enables more nuanced interpretation of trend behaviour. GWSDAT also tends to generate higher Mann–Kendall p-values, which increases the likelihood of classifying trends as non-significant. GWSDAT ranks highest in the MCA final results, however, all total weighted scores lie close to zero, indicating that none of the evaluated software solutions stand out as particularly strong or weak overall.

The sensitivity analysis shows that the MCA outcome is strongly influenced by individual preferences. Despite working within similar roles, respondents evaluated the software differently, highlighting the diversity of user needs. This reinforces that no single tool will be optimal for all practitioners. Given the differing views among practitioners and the absence of a single tool that performs well across all required functions, our overall recommendation is (where possible) to develop a dedicated software solution. Since R is already widely used for statistical calculations, one practical approach is to combine scripted statistical routines with a more user-friendly platform such as GWSDAT for plotting and interpolation.

Based on both the evaluation and the survey, several functions emerge as essential for a new tool: guideline-based flagging, handling of non-detects, GIS map import, flexible plotting, data modifiability, and clear tabular outputs. Although respondents ranked statistical measures as less important, the literature consistently shows that statistical interpretation is central to understanding temporal trends. A new software tool should therefore include a robust statistical module capable of handling the characteristics of groundwater contaminant data. This also means integrating appropriate trend-detection methods, including non-parametric tests suited for skewed and irregular datasets, to ensure that temporal evaluations are both reliable and scientifically defensible.

Since the theoretical background has already shown that groundwater data often require non-parametric trend methods (U.S. Environmental Protection Agency, 2009), it is important to emphasise that a new GMP software tool must include appropriate trend-testing functionality. In addition to monotonic trend tests such as Mann–Kendall and Sen's slope, tools for identifying non-monotonic behaviour, as implemented in GWSDAT through local linear regression, are valuable for capturing variation in temporal patterns.

Several additional statistical components are also necessary. Measures such as mean, standard deviation and coefficient of variation support interpretation of variability and potential heteroskedasticity (U.S. Environmental Protection Agency, 2009), while statistical power is highlighted by Naturvårdsverket (2008) as important for evaluating the reliability of detected trends. Furthermore, systematic handling of non-detects and outliers must be integrated. Although this thesis applied simple substitution and did not perform outlier testing, both Naturvårdsverket (2008) and U.S. Environmental Protection Agency (2009) emphasise that these methods are essential when working with raw laboratory data and should therefore be included in future software development.

5 Conclusion

As a whole, there are several opportunities for alterations in current workflow that can provide a better understanding of contamination behaviour. As the results indicate, adapting to only one of the discussed methods or software will not allow for a sufficient overall picture. Instead, combining techniques and procedures that cover different aspects is the recommended approach. For example, combining the findings of the approaches and the software evaluation (Section 4.2 and 4.3). A software that covers all the most important features identified in Section 4.3.5 for standard individual-well time-series analysis while also having interpolation and Ricker Method calculations implemented would be ideal. This would not only assist GMP evaluators, but also streamline their ability to efficiently present their decisions to relevant stakeholders and authorities. GWSDAT already includes plume mass estimations, but was found to be too limited in e.g. choices surrounding the plume calculations and the amount of output parameters that could be used for reviewing the performance. There is potential there for more user-modifiability, however, this could ultimately go against the straightforward simplicity that GWSDAT seems to be going for.

For the Ricker Method, it can be concluded that the placement of wells is crucial for the outcome of the calculations. We identify that a sufficient, however, undetermined number of wells, are needed surrounding a source to adequately estimate the magnitude of plume concentration drop off in a close area around the source. Furthermore, including wells that are placed in the outer perimeter relative to the source allows for necessary estimations of plume cut-off. Any additional wells in expected groundwater flow directions will also yield increasing levels of certainty. If this can not be achieved due to e.g. site limitations or economic/time feasibility, sticking to well-by-well observations remains the best option in order to avoid producing misinterpreting interpolation results. However, as mentioned, involving Mann-Kendall significance tests and other statistical measures, as well as making sure to review individual constituent behaviours is essential to meet the foundational requirements of a GMP.

The sampling frequency of a GMP has a direct influence on the ability to detect trends, since short or irregular datasets tend to produce non-significant outcomes even when real changes are present. Several studies emphasise this limitation, with one investigation finding that more than seven years of quarterly monitoring data was required to identify a statistically significant concentration trend in most wells, highlighting how sparse sampling can obscure underlying behaviour. A second challenge is the reliance on monotonic trend assumptions, which rarely reflect the behaviour of environmental datasets that fluctuate seasonally or respond to operational changes. One way to avoid this is by testing intervals of data, e.g. only including measurements after remediation. Another way is to use non-monotonic trend test, however, these are more uncommon in practice.

Together, the limitations of the Ricker Method and the Mann-Kendall test shows that the outcome of a trend evaluation depends not only on the method but also on the design of the GMP, including both well layout and sampling frequency, which are crucial for producing a substantiated interpretation. Therefore, for monitoring programmes in Sweden to be able to rely more on quantitative, statistical methods, each programme needs to be designed with a well network and/or sampling frequency that can support the chosen method. When performing site-specific evaluations and providing decision support, additional contextual information is required, such as the timing of remediation activities, hydrogeological changes, operational history, or shifts in analytical methods. Without this context, it becomes difficult to distinguish true environmental trends from external disturbances.

5.1 Extensions of This Work

There are multiple ways to utilise the Ricker Method further than what is explored in this thesis work. For example movement of the center of a plume mass over time or the spacial change indicator meant to clearly illustrate which parts of a plume that have increased or decreased since a set date. These are both very informational tools which can highlight contamination movements and remediation performance, but ultimately fell outside of feasibility given the time schedule of this work.

The interpolations for the Ricker Method was only carried out on the sum of constituents in order to save time. A potential solution to the issue of planar areas being the maximum size of the grid area could be to carry out the plume mass interpolations for each substance individually, as some Ricker literature suggests. This could hypothetically work better for sampling dates with few wells that do not represent a wide area of the site, since each constituent plume might indicate a smaller contamination spread. Also, although the Ricker Method scenario tests gave meaningful context to the importance of a GMP well-layout, they were too few to properly draw more specific conclusions. More tests of the same manner would be helpful in determining the minimum amount of wells to recommend for both the 'close to source' set of wells and the 'outer perimeter' set of wells. Further tests on this could produce a framework for optimal Ricker Method workflow specifically for small datasets, but could also end in recommendations that would be too time consuming for practitioners.

Another extension of this work could be a comparison between the Ricker plume-mass method, implemented through kriging, and the "Plume Diagnostics" in GWSDAT. GWSDAT computes the plume metrics of mass, area and average concentration from its spatiotemporal P-spline smoother, which generates a smooth estimate of concentration in space and time and then derives plume area and mass from the resulting surface (Jones et al., 2026). Such a comparison would have allowed evaluation of how a model-based smoother differs from a geometry-based interpolation workflow. However, this analysis was not feasible within the project time frame because the P-spline modelling approach used by GWSDAT was only identified during the software evaluation phase, leaving insufficient time to implement a full cross-method comparison.

6 References

Aziz, J. J., Ling, M., Rifai, H. S., Newell, C. J., & Gonzales, J. R. (2003). Maros: A decision support system for optimizing monitoring plans. *Groundwater*, *41*(3), 355–367.

Benavides Höglund, N., Sparrenbom, C., & Hugman, R. (2023). A probabilistic assessment of surface water-groundwater exchange flux at a pce contaminated site using groundwater modelling. *Frontiers in Earth Science*, *11*, 1168609.

Bøllingtoft, A. (2025). Geostatistical approaches to contaminant mass discharge quantification and uncertainty estimation in groundwater.

Bøllingtoft, A., Troldborg, M., Riis, C., Tuxen, N., & Bjerg, P. L. (2026). Temporal evolution of contaminant mass discharge: Effect of source remediation at contaminated sites. *Groundwater Monitoring & Remediation*.

Byl, T. D., & Williams, S. D. (2000). *Biodegradation of chlorinated ethenes at a karst site in middle tennessee* (Vol. 99). US Department of the Interior, US Geological Survey.

Cremeans, M., Devlin, J., McKnight, U. S., & Bjerg, P. L. (2018). Application of new point measurement device to quantify groundwater-surface water interactions. *Journal of Contaminant Hydrology*, *211*, 85–93.

Cwiertny, D. M., & Scherer, M. M. (2010). Chlorinated solvent chemistry: Structures, nomenclature and properties. In *In situ remediation of chlorinated solvent plumes* (pp. 29–37). Springer.

Dean, M. (2022). A practical guide to multi-criteria analysis. *UCL: London, UK*, *33*, 142.

Denham, M. E., Amidon, M. B., Wainwright, H. M., Dafflon, B., Ajo-Franklin, J., & Eddy-Dilek, C. A. (2020). Improving long-term monitoring of contaminated groundwater at sites where attenuation-based remedies are deployed. *Environmental Management*, *66*(6), 1142–1161.

Ding, X.-H., Feng, S.-J., Zhao, L.-H., Zhang, S., Chen, H.-X., Sun, X.-L., & Zhou, H.-T. (2025). Back diffusion of chlorinated solvents from unsaturated low permeability zones. *Journal of Hydrology*, *656*, 133007.

Directive (EU) 2020/2184. (2020). Directive (eu) 2020/2184 of the european parliament and of the council of 16 december 2020 on the quality of water intended for human consumption [Accessed: 2026-05-16].

European Chemicals Agency. (n.d.-a). Tetrachloroethylene: Identity information [EC number 204-825-9, CAS number 127-18-4. Accessed: 2026-02-17].

European Chemicals Agency. (n.d.-b). Trichloroethylene: Substance overview [EC number 201-167-4, CAS number 79-01-6. Accessed: 2026-02-17].

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, *21*(1), 196–216. Retrieved March 19, 2026, from http://www.jstor.org/stable/3035587

Föreningen, S. G. (2011). Klorerade lösningsmedel i mark och grundvatten. *SGF Rapport*, *2*.

Golden Software. (2023). Calculate contamination plume stability statistics in surfer [Accessed: 2026-05-18]. https://support.goldensoftware.com/hc/en-us/articles/236054748-Calculate-contamination-plume-stability-statistics-in-Surfer

Golden Software. (2024a). Cross validation [Accessed: 2026-05-18].

Golden Software. (2024b). Default linear variogram [Accessed: 2026-06-05].

Golden Software. (2024c). Variogram overview [Accessed: 2026-06-05].

Golden Software. (2025). *Understanding surfer gridding methods*. Golden Software Support. https://support.goldensoftware.com/hc/en-us/articles/231348728-Understanding-Surfer-gridding-methods

Golden Software. (n.d). *The ricker method for plume stability analysis* [Accessed: 2026-03-24]. https://www.goldensoftware.com/ricker-method-for-plume-stability-analysis/

Grath, J., Scheidleder, A., Uhlig, S., Weber, K., Kralik, M., Keimel, T., & Gruber, D. (2001). The eu water framework directive: Statistical aspects of the identification of groundwater pollution trends, and aggregation of monitoring results. *Final report. Austrian Federal Ministry of Agriculture and Forestry, Environment and Water Management*.

Holbert, C. (n.d.). Groundwater statistics using trendmk a brief introduction to r.

Jones, W. R., Rock, L., Miller, C., Low, M., Alexander, C., Bowman, A., & Radvanyi, P. (2026, January 30). *Gwsdat version 3.3 user manual* [Version 3.3]. Shell Research Limited and The University of Glasgow. https://gwsdat.net

Jones, W. R., Spence, M. J., Bowman, A. W., Evers, L., & Molinari, D. A. (2014). A software tool for the spatiotemporal analysis and reporting of groundwater monitoring data. *Environmental Modelling Software*, *55*, 242–249. https://doi.org/https://doi.org/10.1016/j.envsoft.2014.01.020

Kendall, M. G., & Gibbons, J. D. (1962). Rank correlation methods.

Lin, N., Yun, X., Zhang, Y., Wu, Y., Du, W., & Batterman, S. (2025). State of the science on 1, 4-dioxane: Emission sources, global contamination, and regulatory disparities. *Environmental Science & Technology*, *59*(38), 20125–20142.

Ling, M., Rifai, H. S., & Newell, C. J. (2005). Optimizing groundwater long-term monitoring networks using delaunay triangulation spatial analysis techniques. *Environmetrics: The official journal of the International Environmetrics Society*, *16*(6), 635–657.

Livsmedelsverket. (2022). *Livsmedelsverkets föreskrifter om dricksvatten* (Föreskrifter No. LIVSFS 2022:12). Livsmedelsverket. https://www.livsmedelsverket.se/4af3b1/globalassets/om-oss/lagstiftning/dricksvatten---naturl-mineralv---kallv/livsfs-2022-12_web_t.pdf

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, 245–259.

McCarty, P. L. (2010). Groundwater contamination by chlorinated solvents: History, remediation technologies and strategies. In *In situ remediation of chlorinated solvent plumes* (pp. 1–28). Springer.

McGuire, J. T., Smith, E. W., Long, D. T., Hyndman, D. W., Haack, S. K., Klug, M. J., & Velbel, M. A. (2000). Temporal variations in parameters reflecting terminal-electron-accepting processes in an aquifer contaminated with waste fuel and chlorinated solvents. *Chemical Geology*, *169*(3-4), 471–485.

McGuire, T. M., Newell, C. J., Looney, B. B., Vangelas, K. M., & Sink, C. H. (2004). Historical analysis of monitored natural attenuation: A survey of 191 chlorinated solvent sites and 45 solvent plumes. *Remediation Journal: The Journal of Environmental Cleanup Costs, Technologies & Techniques*, *15*(1), 99–112.

McHugh, T. E., Adamson, D. T., Actkinson, B. W., & Newell, C. J. (2025). Determining pfaa plume stability condition quickly and efficiently. *Groundwater Monitoring & Remediation*, *45*(1), 68–79.

Mchugh, T. E., Kulkarni, P. R., Newell, C. J., & Britt, S. L. (2015). *Methods for minimization and management of variability in long term groundwater monitoring results* (tech. rep.).

McLean, M., Evers, L., Bowman, A., Bonte, M., & Jones, W. (2019). Statistical modelling of groundwater contamination monitoring data: A comparison of spatial and spatiotemporal methods. *Science of The Total Environment*, *652*, 1339–1346. https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.10.231

Naturvårdsverket. (2008, March). *Miljökontroll av omgivningspåverkan vid efterbehandling* (Report No. 5803). Naturvårdsverket. Stockholm. https://www.naturvardsverket.se/4a436f/globalassets/media/publikationer-pdf/ovriga-pub/hallbar-sanering/undersokningsmetoder/620-5803-6.pdf

Naturvårdsverket. (2016). *Nationell plan för fördelning av statliga bidrag för efterbehandling av förorenade områden* (tech. rep. No. Rapport 6720) (Reviderad version augusti 2016). Naturvårdsverket. https://www.diva-portal.org/smash/get/diva2:956367/FULLTEXT02.pdf

Naturvårdsverket. (2025). Förorenade områden – ett problem för miljö och hälsa [Granskad: 8 december 2025]. Retrieved January 27, 2026, from https://www.naturvardsverket.se/amnesomraden/fororenade-omraden/om-fororenade-omraden/

Newell, C. J., Cowie, I., McGuire, T. M., & McNab Jr, W. W. (2006). Multiyear temporal changes in chlorinated solvent concentrations at 23 monitored natural attenuation sites. *Journal of Environmental Engineering*, *132*(6), 653–663.

Pino, D. S., Bertolo, R., Fregona, L. G. G., & Pak, T. (2026). Overview of groundwater management at sites contaminated by chlorinated solvents in brazil. *Environmental Earth Sciences*, *85*(1), 21.

Ricker, J. A. (2008). A practical method to evaluate ground water contaminant plume stability. *Groundwater Monitoring & Remediation*, *28*(4), 85–94.

Ricker, J. A., Goist, T. O., & Winchell, D. C. (2020). Molarity provides better clarity–using molar-based data when evaluating chlorinated volatile organic compound impacted sites. *Environmental Forensics*, *21*(1), 71–78.

Sen, P. K. (1968). Estimates of the regression coefficient based on kendall's tau. *Journal of the American statistical association*, *63*(324), 1379–1389.

Sharma, D. K., Chatterjee, M., Kaur, G., & Vavilala, S. (2022). 3 - deep learning applications for disease diagnosis. In D. Gupta, U. Kose, A. Khanna, & V. E. Balas (Eds.), *Deep learning for medical applications with unique data* (pp. 31–51). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-824145-5.00005-8

Sveriges geologiska undersökning. (2024). *Halogenerade lösningsmedel – gränsvärden och tillståndsklasser* [Senast ändrad 2024-12-19]. Retrieved February 25, 2026, from https://www.sgu.se/anvandarstod-for-geologiska-fragor/bedomningsgrunder-for-grundvatten/grundvattnets-kvalitet--organiska-amnesgrupper/halogenerade/halogenerade-losningsmedel--gransvarden-och-tillstandsklasser/

Sveriges geologiska undersökning (SGU). (2024, December). Sgu-fs 2024:1 — föreskrifter om ändring i SGU-FS 2023:1 om kartläggning, riskbedömning och klassificering av status för grundvatten [Utkom från trycket den 12 december 2024]. https://www.sgu.se

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes mathematicae*, *12*(85), 173.
Tobiszewski, M., & Namieśnik, J. (2012). Abiotic degradation of chlorinated ethanes and ethenes in water. *Environmental Science and Pollution Research*, *19*(6), 1994–2006.
U.S. Environmental Protection Agency. (2005). *Roadmap to long-term monitoring optimization* (tech. rep. No. EPA 542-R-05-003). U.S. Environmental Protection Agency. https://www.epa.gov/sites/default/files/2015-08/documents/roadmap_ltmo_542-r-05-003.pdf
U.S. Environmental Protection Agency. (2009). *Statistical analysis of groundwater monitoring data at rcra facilities: Unified guidance* (tech. rep. No. EPA 530-R-09-007). U.S. Environmental Protection Agency. https://archive.epa.gov/epawaste/hazard/web/pdf/unified-guid-toc.pdf
U.S. Environmental Protection Agency. (2022, April). *Proucl version 5.2 technical guide* (tech. rep. No. EPA/600/R-07/041) (Updated April 2022). Office of Research and Development. Washington, DC. https://www.epa.gov/land-research/proucl-software
Waterloo Hydrogeologic. (2026). Aquachem: Water quality data analysis and reporting software [Accessed: 2026-04-26].
Wikipedia contributors. (2024). *Ordinary least squares* [Accessed: 2026-05-18]. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/Ordinary_least_squares
World Health Organization. (2022). *Guidelines for drinking-water quality: Fourth edition incorporating the first and second addenda* [CC BY-NC-SA 3.0 IGO]. https://iris.who.int/handle/10665/352532
Zdaniuk, B. (2024). Ordinary least-squares (ols) model. In *Encyclopedia of quality of life and well-being research* (pp. 4867–4869). Springer.

Appendix
A Data

This appendix presents the groundwater monitoring data used throughout the study.

A.1 Input data

This appendix presents the full dataset used in the evaluations of the methods for all locations. The tables include measured concentrations of PCE, TCE, DCE, and VC, as well as calculated PCE-equivalent values for each sampling event. All concentration values presented in this appendix are reported in micrograms per liter (μg/L),

Non-detect values reported as ND<x in the original laboratory results are included in the tables exactly as reported. In the main analysis, ND values were handled according to the procedure described in the Method section.

A.1.1 Location A

Data measurements and calculated PCE-equivalent values for Location A.

Sample Well	Sample Date	PCE	TCE	DCE	VC	PCE_eq
A3	2011-01-01	439	3,3	3	ND<1	448
A3	2012-01-01	560	8	7	ND<1	582
A3	2012-04-01	1500	17	10	ND<1	1538
A3	2012-10-01	2100	12	8	ND<1	2129
A3	2013-04-01	1960	10,9	7,11	ND<1	1986
A3	2013-10-01	1340	11,6	11,9	ND<10	1375
A3	2014-01-01	828	8,03	4,23	ND<1	845
A3	2014-10-01	1140	18,1	15,5	ND<10	1189
A3	2015-10-01	1700	12	5,6	ND<0,02	1725
A3	2016-10-01	2900	29	10	ND<0,02	2954
A3	2017-10-01	4000	17	7,6	ND<0,02	4034
A3	2018-10-01	2000	7,5	3,4	ND<0,02	2015
A3	2019-10-01	1800	5,5	2,6	ND<0,02	1811
A3	2020-10-01	600	3,1	1,5	ND<0,02	607
A3	2021-10-01	1400	2,9	1,5	ND<0,02	1406
A3	2022-10-01	962	4	1,99	ND<0,1	971
A3	2023-10-01	1020	2,69	0,603	ND<0,1	1024
A4	2011-01-01	0,2	ND<0,1	ND<0,1	ND<1	0,2
A4	2012-01-01	7,9	0,17	ND<1	ND<1	8
A4	2012-04-01	850	0,8	21	ND<1	887
A4	2012-10-01	470	6,9	27	ND<1	525
A4	2013-04-01	430	3,3	11,8	ND<1	454
A4	2013-10-01	97,5	0,7	1,3	ND<1	101
A4	2014-01-01	122	0,7	0,52	ND<1	124
A4	2014-10-01	155	0,93	0,32	ND<1	157
A4	2015-10-01	110	0,69	0,23	ND<0,02	111
A4	2016-10-01	180	1,7	0,17	ND<0,02	182
A4	2017-10-01	ND<0,02	ND<0,02	ND<0,02	ND<0,02	ND<0,02
A4	2018-10-01	170	1,2	0,5	ND<0,02	172
A4	2019-10-01	120	0,43	0,54	ND<0,02	121
A4	2020-10-01	200	2,4	4,1	ND<0,02	210
A4	2021-10-01	250	0,58	0,83	ND<0,02	252
A4	2022-10-01	153	1,85	1,18	ND<0,1	157
A4	2023-10-01	88,1	0,474	0,194	ND<0,1	89
A1	2011-01-01	ND<0,2	ND<0,1	ND<0,1	ND<1	0,2
A1	2012-01-01	42	0,35	ND<0,1	ND<1	42
A1	2012-04-01	6	0,2	ND<1	ND<1	6,3
A1	2012-10-01	0,4	ND<0,1	ND<1	ND<1	0,4
A1	2013-04-01	ND<0,2	ND<0,1	ND<0,1	ND<1	0,2
A1	2013-10-01	ND<0,2	ND<0,1	ND<0,1	ND<1	0,2
A1	2014-01-01	ND<0,2	ND<0,1	ND<0,1	ND<1	0,2
A1	2014-10-01	ND<0,2	ND<0,1	ND<0,1	ND<1	0,2
A1	2015-10-01	0,03	ND<0,02	ND<0,02	ND<0,02	0,03
A1	2016-10-01	ND<0,02	ND<0,02	ND<0,02	ND<0,02	ND<0,02
A1	2017-10-01	ND<0,02	ND<0,02	ND<0,02	ND<0,02	ND<0,02
A2	2012-01-01	940	25	50	ND<1	1057
A2	2012-04-01	1500	19	32	ND<1	1579
A2	2012-10-01	1000	16	45	ND<1	1097
A2	2013-04-01	411	9,9	25,1	ND<1	467
A2	2013-10-01	250	4,64	9,3	ND<1	272

A2	2014-01-01	197	5,47	6,97	ND<1	216
A2	2014-10-01	23,7	0,25	2,19	ND<1	28
A2	2015-10-01	190	1,8	2,5	ND<0,02	197
A2	2016-10-01	200	1,4	1,8	ND<0,02	205
A2	2017-10-01	130	5,9	8,2	ND<0,02	151
A2	2018-10-01	33	0,96	1,6	ND<0,02	37
A2	2019-10-01	9,3	0,6	0,82	ND<0,02	11

A.1.2 Location B

Data measurements and calculated PCE-equivalent values for Location B.

Sample Wells	Sample Date	PCE	TCE	DCE	VC	PCE_eq
B1	2011-02-22	ND<0,1	0,2	ND<0,1	ND<0,1	1
B1	2012-03-27	ND<0,1	0,2	1,8	ND<0,1	3,9
B1	2012-07-04	0,8	0,4	2,2	0,3	6
B1	2012-10-30	ND<0,1	ND<0,1	ND<0,1	ND<0,1	0,8
B2	2011-02-22	2	0,5	1,4	0,2	5,7
B2	2012-03-27	0,6	0,2	2,4	0,2	5,7
B2	2012-07-04	1	0,4	2	0,5	6,4
B2	2012-10-30	1,7	ND<0,1	ND<0,1	ND<0,1	2,4
B3	2011-02-22	15	9,3	16	0,2	50,9
B3	2012-03-27	95	11	10	ND<0,1	126,5
B3	2012-07-04	20	5,7	8,5	ND<0,1	42,2
B3	2012-10-30	80	23	25	0,4	154,1
B3	2017-04-26	6300	350	460	29	7615,1
B3	2020-03-26	2400	380	1400	22	5346
B3	2024-04-09	57,4	18,9	63,9	0,844	193,81
B4	2011-02-22	90	29	120	3,2	342,5
B4	2012-03-27	13	6,3	140	0,8	265,9
B4	2012-07-04	75	25	180	2,8	425,2
B4	2012-10-30	390	37	180	4,6	763,3
B4	2013-04-11	160	21	48	8,2	292,5
B4	2013-10-08	67	12	20	0,77	118,9
B4	2014-04-23	230	38	630	1,2	1374,3
B4	2014-09-30	1200	69	72	4,7	1425
B4	2015-03-30	290	24	130	ND<0,5	547,2
B4	2016-05-03	ND<0,02	120	180	21	527,5
B4	2017-04-26	240	33	74	1,1	415,8
B4	2020-03-26	340	49	190	11	764
B4	2024-04-09	130	21,7	40,4	0,292	228,44
B5	2011-02-22	910	270	750	120	2861,9
B5	2012-03-27	330	140	2700	170	5611,7
B5	2012-07-04	1500	660	2000	220	6470,9
B5	2012-10-30	620	130	320	8,9	1361,7
B5	2013-04-11	600	310	480	12	1849,6
B5	2013-10-08	64	59	100	11	345,3
B5	2014-04-23	390	220	740	320	2793,3
B5	2014-09-30	500	280	1600	360	4560,8
B5	2015-03-30	1600	240	980	280	4333,1
B5	2016-05-03	ND<0,02	350	990	150	2543,6
B5	2017-04-26	1100	300	850	42	3053,3
B5	2020-03-26	410	110	330	66	1291,9
B5	2024-04-09	1730	345	685	118	3730,18
B6	2011-02-22	180	120	730	110	1881,9
B6	2012-03-27	84	160	460	13	1111,4
B6	2012-07-04	100	120	1900	530	4931,8
B6	2012-10-30	10	5,7	360	49	766,14
B6	2013-04-11	13	4,3	150	46	399,3
B6	2013-10-08	3,5	2	16	10	60,18

B6	2014-04-23	15	24	240	240	1095,7
B6	2014-09-30	72	150	570	160	1666,4
B6	2015-03-30	33	11	120	94	502,6
B6	2016-05-03	40	220	590	180	1811,4
B6	2017-04-26	24	94	430	44	999,2
B6	2020-03-26	120	69	435	130	1293,12
B6	2024-04-09	140	61,2	330	77,3	1041,868511
B7	2011-02-22	55	10	44	ND<0,1	143,66
B7	2012-03-27	70	18	69	0,9	214,26
B7	2012-07-04	88	26	120	5,5	342,84
B7	2012-10-30	170	22	58	ND<0,1	298,02
B7	2017-04-26	310	75	120	0,83	614,3914611
B7	2020-03-26	120	9	16	0,031	159,12
B7	2024-04-09	79,5	7,26	6,43	ND<0,1	99,99722545
B8	2011-02-22	350	6,4	4,9	ND<1	370,9
B8	2012-03-27	4,2	1,8	13	0,4	30
B8	2012-07-04	5,1	2,3	21	0,6	45,8
B8	2012-10-30	18	3,8	8	0,2	37,2
B8	2013-04-11	23	6,9	12	0,3	53,3
B8	2013-10-08	40	6,5	10	0,26	66,2
B8	2014-04-23	48	6,9	9,9	ND<0,1	74,1
B8	2015-03-30	55	6,7	11	ND<0,5	83,2
B8	2016-05-03	41	3,6	5,3	ND<0,02	54,7
B8	2017-04-26	36	9,5	17	0,051	77,5
B8	2020-03-26	29	3	7,2	0,2	45,7
B8	2024-04-09	5,63	0,89	1,89	ND<0,1	10,3

A.1.3 Location C

Data measurements and calculated PCE-equivalent values for Location C.

Sample Wells	Sample Date	PCE	TCE	DCE	VC	PCE_ekv
C1	2022-05-12	1600	860	624	5,7	3767,597992
C1	2022-10-27	ND<0,1	ND<0,1	23	0,41	40,53847729
C1	2023-06-16	1400	890	1709,9	11	5476,768772
C1	2023-11-22	1000	690	1509	17	4496,700595
C1	2024-10-24	1300	840	2009,5	24	5860,57732
C1	2025-11-20	410	320	2507,8	55	5248,890814
C10	1997-05-18	51	28	81	0	224,8702177
C10	2010-05-15	3,8	0,9	0,5	ND<0,1	5,923511015
C10	2011-11-15	1,2	0,5	1,6	ND<0,1	4,700124103
C10	2012-02-15	ND<0,1	0,4	22	0,7	40,03911522
C10	2022-05-10	ND<0,1	ND<0,1	0,68	ND<0,1	1,408763281
C10	2022-10-27	ND<0,1	ND<0,1	0,32	ND<0,1	0,793042217
C10	2023-06-17	6,3	6,5	71,23	0,95	138,8497125
C10	2023-11-22	82	140	452,9	13	1067,762597
C10	2024-10-25	190	160	281,4	16	915,6361645
C10	2025-11-20	ND<0,1	ND<0,1	0,37	ND<0,1	0,878559031
C11	2010-05-15	1100	270	361,9	3,1	2067,905244
C11	2011-11-15	42	13	451,8	0,5	832,460933
C11	2012-02-15	1700	580	322	4,5	2994,56329
C11	2022-05-10	ND<0,1	0,13	17	0,53	30,74574691
C11	2022-10-27	ND<0,1	0,14	69,24	1,1	121,6184296
C11	2023-06-17	330	200	642,5	10	1707,797482
C11	2023-11-22	770	580	643,8	11	2632,192708
C11	2024-10-25	27	20	210,9	14	450,0869652
C11	2025-11-20	6	110	532,9	22	1114,607938
C12	1997-05-18	10	2	2	0	15,94445674
C12	2010-05-15	7,4	4,8	67,3	0,6	130,1543943
C12	2011-11-15	0,8	0,6	5	ND<0,1	10,2414567
C12	2012-02-15	39	25	141	1,4	315,4186389
C12	2022-05-10	0,73	0,92	23,11	2,3	47,51825238
C12	2022-10-27	1,2	1,9	30,17	1,4	58,91236083
C12	2023-06-17	24	21	130,54	1,3	277,215673
C12	2023-11-22	9,3	11	93,69	14	220,5614199
C12	2024-10-25	320	240	331,7	18	1237,923046
C12	2025-11-20	120	170	472,4	25	1208,804517
C11	1997-05-12	420	77	25		559,9240972
C11	1997-10-27	370	78	24		509,475653
C11	1997-12-15	750	120	53		992,0748726
C11	1998-02-23	840	150	55		1123,352308
C11	1998-04-20	1000	180	73		1351,995123
C11	1998-06-15	940	170	64		1263,983176
C11	1998-09-07	1200	200	81		1590,915655
C11	1999-04-19	1600	230	83		2032,19309
C11	1999-10-11	1500	190	130		1962,103212
C11	2000-05-02	2400	290	110		2954,085694
C11	2000-10-16	1300	240	110		1790,99109
C11	2001-05-07	1600	250	100		2086,506648
C11	2001-10-15	1600	270	120		2145,951216
C11	2002-04-29	1200	260	170		1818,849109
C11	2002-10-14	1700	300	150		2335,118067
C11	2003-05-12	1600	320	140		2243,252545
C11	2004-05-17	1600	320	160		2277,459271
C11	2005-10-17	2000	350	110		2629,799219
C11	2008-09-08	1700	380	160		2453,172796
C11	2008-10-14	1900	410	190		2742,339647

C13	1996-09-18	3800	240	31		4155,874523
C13	1997-02-12	5500	390	64		6101,599433
C13	1997-03-17	2800	200	35		3112,240186
C13	1997-05-26	2300	260	33		2684,533038
C13	1997-08-12	3300	250	28		3663,362435
C13	1997-10-27	4000	220	29		4327,216009
C13	1997-12-15	4100	260	43		4501,636401
C13	1998-02-23	3100	230	35		3450,096948
C13	1998-04-20	3400	250	27		3761,652099
C13	1998-06-15	4100	280	28		4501,219198
C13	1998-09-07	3800	410	590		5326,474163
C13	1999-04-19	3400	300	77		3910,263517
C13	1999-10-11	2100	1000	1000		5072,228368
C13	2000-05-02	4600	590	160		5618,170132
C13	2000-10-16	2000	1200	940		5121,986605
C13	2001-05-07	3200	680	100		4229,120241
C13	2001-10-15	4000	610	150		5026,304611
C13	2002-04-29	3800	790	160		5070,548547
C13	2002-10-14	3200	760	250		4586,622051
C13	2003-05-12	4100	930	120		5478,799986
C13	2004-05-17	3700	1000	160		5235,545884
C13	2005-10-17	4600	1200	190		6439,234388
C13	2008-09-08	4300	910	190		5773,285685
C13	2008-10-14	4800	1200	210		6673,441113
C13	2010-05-15	4300	660			5132,848771
C9	2011-11-15	ND<0,1	0,2	0,3	ND<0,1	1
C9	2012-02-15	ND<0,1	ND<0,1	ND<0,1	ND<0,1	0,3
C9	2022-05-10	ND<0,1	0,24	ND<0,1	ND<0,1	0,7
C9	2022-10-27	ND<0,1	0,66	0,11	ND<0,1	1,4
C9	2023-06-16	ND<0,1	0,66	ND<0,1	ND<0,1	1,27
C9	2023-11-20	ND<0,1	0,28	ND<0,1	ND<0,1	0,79
C2	2022-05-10	ND<0,1	0,41	6,6	0,39	13,1
C2	2022-10-27	23	27	111	0,39	248,6
C2	2023-06-16	ND<0,1	0,89	9,4	ND<0,1	17,6
C2	2023-11-20	ND<0,1	0,78	8,7	0,15	16,5
C2	2024-10-24	ND<0,1	0,62	8,4	0,15	15,8
C2	2025-11-20	ND<0,1	0,61	10	0,15	18,5
C3	2022-05-10	ND<0,1	ND<0,1	1,5	0,24	3,5
C3	2022-10-28	ND<0,1	ND<0,1	23	1,2	42,8
C3	2023-06-17	ND<0,1	ND<0,1	3,2	3	13,7
C3	2023-11-20	ND<0,1	ND<0,1	2	12	35,5
C3	2024-10-24	ND<0,1	ND<0,1	3,6	13	40,9
C3	2025-11-20	ND<0,1	ND<0,1	0,44	1,3	4,5
C4	2022-05-10	ND<0,1	ND<0,1	0,3	ND<0,1	0,9
C4	2022-10-27	ND<0,1	ND<0,1	ND<0,1	ND<0,1	0,5
C4	2023-06-16	0,1	ND<0,1	ND<0,1	ND<0,1	0,6
C4	2023-11-20	ND<0,1	ND<0,1	ND<0,1	ND<0,1	0,5
C4	2024-10-24	ND<0,1	ND<0,1	ND<0,1	ND<0,1	0,5
C4	2025-11-20	ND<0,1	ND<0,1	ND<0,1	ND<0,1	0,5
C5	2022-05-10	ND<0,1	ND<0,1	5,8	0,67	12
C5	2022-10-28	ND<0,1	ND<0,1	7,3	0,56	14,3
C5	2023-06-17	ND<0,1	ND<0,1	8	0,52	15,3
C5	2023-11-22	ND<0,1	ND<0,1	3,5	0,28	7
C5	2024-10-24	ND<0,1	ND<0,1	2,2	0,52	5,4
C5	2025-11-20	ND<0,1	ND<0,1	1,7	0,62	4,8
C6	2022-05-12	ND<0,1	ND<0,1	ND<0,1	ND<0,1	0,5
C6	2022-10-27	0,14	0,12	0,82	ND<0,1	2
C6	2023-06-17	ND<0,1	ND<0,1	ND<0,1	ND<0,1	0,5
C6	2023-11-22	ND<0,1	ND<0,1	0,46	0,11	1,4

C6	2024-10-24	ND<0,1	ND<0,1	ND<0,1	ND<0,1	0,5
C7	2022-05-10	ND<0,1	ND<0,1	0,54	ND<0,1	1,3
C7	2022-10-27	ND<0,1	ND<0,1	0,55	ND<0,1	1,5
C7	2023-06-16	ND<0,1	ND<0,1	0,4	ND<0,1	1
C7	2023-11-20	ND<0,1	ND<0,1	0,99	0,14	2
C7	2024-10-24	ND<0,1	ND<0,1	0,62	ND<0,1	1
C7	2025-11-20	0,46	0,85	6,7	0,14	14
C8	2011-11-15	2,4	1,5	164	2,5	291
C8	2012-02-15	ND<0,1	0,9	258	7,8	463
C8	2022-05-10	ND<0,1	ND<0,1	28,7	89	285,4188869
C8	2022-10-27	ND<0,1	ND<0,1	12,6	21	77,46699095
C8	2023-06-17	ND<0,1	ND<0,1	25,3	26	112,456482
C8	2023-11-21	ND<0,1	ND<0,1	42,5	42	184,5618961

A.1.4 Location D

Data measurements and calculated PCE-equivalent values for Location D.

Sample location	Sample date	PCE	TCE	DCE	VC	PCE_eq
D1	2019-04-25	2580	671	188.88	ND<10	3763.32
D1	2019-06-11	8400	3650	539	ND<100	14061.79
D1	2019-09-19	7920	2560	481.39	ND<10	11988.03
D1	2020-09-23	4700	1500	1511.5	25	9245.30
D1	2021-10-13	1300	410	213.45	1.4	2186.36
D1	2022-11-16	593	179	17.85	ND<1	850.80
D1	2024-10-16	499	180	49.957	1.86	816.59
D2	2019-04-25	206	370	133.14	3.6	910.33
D2	2019-06-11	19.9	48.2	13.6	ND<1	105.33
D2	2019-09-19	15.9	57.8	11.9	ND<1	110.54
D2	2020-09-23	8.5	26	11	ND<1	61.46
D2	2021-10-13	350	450	222.12	9.2	1322.38
D2	2022-11-16	307	311	115.906	8.39	920.08
D2	2024-10-16	124	131	89.805	1.18	446.11
D3	2019-04-25	1620	467	329.84	12.7	2807.39
D3	2019-06-11	3760	1010	251.74	13.1	5500.23
D3	2019-09-19	1610	519	614.52	12.1	3348.42
D3	2020-09-23	2100	580	734.9	41	4198.02
D3	2021-10-13	2100	670	495.7	19	3844.06
D3	2022-11-16	828	280	107	3.8	1374.54
D3	2024-10-16	2110	830	536.38	22.9	4135.96
D4	2019-04-25	0.21	6.9	120.77	16	257.97
D4	2019-06-11	ND<0.02	ND<0.02	0.29	ND<0.02	0.55
D4	2019-09-19	ND<0.02	3.2	98.42	8.3	194.43
D4	2020-09-23	0	5.1	46.66	2.9	93.95
D4	2021-10-13	0.54	20	200.81	5.8	384.69
D4	2022-11-16	0.622	42.7	142.821	5.59	313.67
D4	2024-10-16	0.13	139	583.059	17.2	1218.62
D5	2019-04-25	0.089	0.089	0.03	ND<0.02	0.28
D5	2019-06-11	ND<0.02	ND<0.02	0.03	ND<0.02	0.10
D5	2019-09-19	ND<0.02	ND<0.02	0.07	ND<0.02	0.17
D5	2020-09-23	0.024	0.021	0.5	ND<1	2.23
D5	2021-10-13	0.75	0.56	1.92	0.046	4.86
D5	2022-11-16	ND<0.1	ND<0.1	0.15	ND<0.1	0.50
D5	2024-10-16	ND<0.1	ND<0.1	0.15	ND<0.1	0.50
D6	2019-04-25	ND<0.02	ND<0.02	0.03	ND<0.02	0.10
D6	2019-06-11	ND<0.02	ND<0.02	0.03	ND<0.02	0.10
D6	2019-09-19	ND<0.02	ND<0.02	0.03	ND<0.02	0.10

A.1.5 Location E

Data measurements and calculated PCE-equivalent values for Location E.

Sample Well	Sample Date	PCE	TCE	DCE	VC	PCE_eq
E1	2000-04-14	391	128	93	ND<1	714.67
E1	2000-08-28	1687	354	344	ND<1	2725.26
E1	2000-10-19	1380	406	217	ND<1	2266.63
E1	2000-12-19	1418	297	108	ND<1	1980.61
E1	2001-11-14	3354	324	141	ND<1	4007.13
E1	2002-12-18	730	140	42	ND<10	1008.90
E1	2003-11-28	1100	240	260		1847.65
E1	2004-03-30	475	58	30		599.52
E1	2004-12-15	350	77	180		755.09
E1	2005-11-24	3500	470	440		4845.84
E1	2007-01-17	420	98	180		851.59
E1	2009-11-09	1300	200	460		2344.78
E1	2014-11-20	1610	285	463	ND<10	2779.37
E1	2016-09-08	419	75,4	215	ND<10	898.27
E1	2018-10-15	6640	413	328	ND<10	7744.32
E1	2021-12-09	1700	160	310	1	2441.38
E1	2022-11-02	1090	183	318	ND<1	1687.15
E2	2000-04-14	999	204	250	ND<1	1687.15
E2	2000-08-28	2923	347	1082	ND<1	5214.88
E2	2000-10-19	5379	872	800	ND<1	7851.04
E2	2000-12-19	234	48	70	ND<1	417.36
E2	2001-11-14	894	218	360	ND<1	1787.99
E2	2002-12-18	880	290	140	ND<10	1515.85
E2	2003-11-28	9300	1200	650		11926.35
E2	2004-03-30	73	21	41		169.64
E2	2004-12-15	270	110	460		1195.72
E2	2005-11-24	3500	470	930		5684.05
E2	2007-01-17	940	300	940		2926.61
E2	2009-11-09	3900	1000	1200		7238.75
E2	2014-11-20	1420	749	1740	ND<10	5374.21
E2	2016-09-08	3060	668	1230	100	6156.90
E2	2018-10-15	2680	842	854	14.2	5264.30
E2	2021-12-09	0.23	ND<0.1	7.7	ND<1	15.73
E2	2022-11-02	369	322	1560	143	3842.42
E3	2009-11-09	0.3	0.3	2.4		5.30
E3	2014-11-20	5.3	2.3	9.3	ND<0.5	25.90
E3	2016-09-08	0.76	0.88	4.34	ND<1	11.60
E3	2018-10-15	ND<0.2	4.38	14.9	ND<1	37.13
E3	2021-12-09	330	97	500	ND<1	1317.11
E3	2022-11-02	ND<0.2	0.467	7.16	ND<1	15.21
E4	2009-11-09	ND<0.2	0.4	3.3		6.52
E4	2014-11-20	0.84	0.12	2	ND<0.5	5.25
E4	2016-09-08	2.74	1.51	5.66	ND<1	15.83
E4	2018-10-15	1.87	0.16	0.82	ND<1	4.97
E5	2009-11-09	15	0.5	2.5		19.99
E5	2010-12-20	0.3	0.5	26	0.4	46.55
E5	2021-12-09	1800	170	310	1.1	2554.78
E5	2022-11-02	2.7	0.707	1.7	ND<1	8.77
E6	2009-11-09	E1	74	130		347.12
E6	2010-12-20	540	130	150	ND<0.2	962.22
E6	2018-10-15	240	66.7	323	10.8	911.96
E7	2009-11-09	7900	360	190		8687.90
E7	2010-12-20	10000	660	320	0.6	11387.41

E7	2014-11-20	184	166	374	1.8	1043.22
E7	2018-10-15	7930	226	85.9	ND<10	8378.18
E8	2009-11-09	0.5	0.3	0.4		1.65
E8	2014-11-20	10	0.3	4.1	ND<0.5	18.23
E8	2016-09-08	5.84	2.08	7.16	ND<1	22.21
E8	2018-10-15	6.36	0.32	0.93	ND<1	9.85
E8	2022-11-02	134	56	284	ND<1	695.08
E9	2009-11-09	0.3	ND<0.2	71		122.86
E9	2010-12-20	ND<0.1	0.6	93	0.9	162.80
E9	2014-11-20	11	3.1	190	11	371.52
E9	2018-10-15	38.3	2.08	89.1	6.9	214.41
E9	2021-12-09	0.35	0.8	11	ND<1	23.47
E9	2022-11-02	629	136	328	ND<1	1367.64
E10	2010-12-20	61	42	53	0.6	206.95
E10	2018-10-15	61.3	23	15.4	ND<1	118.56
E10	2021-12-09	2100	200	350	1.1	2961.06
E10	2022-11-02	76.4	28.8	32.2	ND<1	170.36
E11	2010-12-20	770	370	830	0.9	2665.67
E11	2014-11-20	140	76	250	1.9	671.01
E11	2018-10-15	75.4	3.17	429	2	829.98
E11	2022-11-02	2.47	4.23	234	ND<1	412.24
E12	2010-12-20	ND<0.2	0.5	22	9.3	63.23
E12	2018-10-15	4.54	0.72	41	10.3	103.94
E12	2021-12-09	2200	200	350	ND<1	3060.11
E12	2022-11-02	4.14	1.79	80.8	17.8	192.97
E13	2010-12-20	0.6	1.7	8.8	0.2	18.42
E13	2018-10-15	61.9	4.26	86	2.4	223.57
E13	2022-11-02	1.62	0.763	3.31	2.16	14.92
E14	2010-12-20	290	330	800	3.8	2092.77
E14	2014-11-20	120	190	810	37	1850.93
E14	2018-10-15	3.5	1.48	8.59	ND<1	21.56
E14	2021-12-09	800	240	1200	ND<1	3165.03
E14	2022-11-02	20.2	4.26	275	34.8	591.56
E15	2010-12-20	2.6	1.7	390	ND<0.1	673.40
E15	2014-11-20	8.9	2.8	94	35	267.18
E15	2018-10-15	50	3.39	5.79	23	125.50
E15	2021-12-09	0.73	0.92	12	ND<1	25.88
E15	2022-11-02	ND<0.2	0.215	10.4	39.1	122.85
E16	2018-10-15	439	148	250	ND<1	1061.29
E16	2021-12-09	390	110	540	ND<1	1462.62
E16	2022-11-02	422	204	470	ND<1	1491.92

A.1.6 Location F

Data measurements and calculated PCE-equivalent values for Location F.

Sample Well	Sample date	PCE	TCE	DCE	VC	PCE_eq
F8	2019-03-06	0.27	0.67	3.7	1.2	10.63
F8	2019-05-22	0.32	0.9	4.3	1.6	13.35
F8	2019-09-04	0.43	1.1	4.8	1.4	14.17
F8	2019-11-05	0.27	0.82	2.7	0.68	8.03
F8	2020-05-29	0.26	0.72	1.8	0.41	5.54
F9	2019-03-06	ND<0.1	0.11	140	11	270.34
F9	2019-05-22	ND<0.1	ND<0.1	160	24	339.58
F9	2019-09-04	ND<0.1	0.15	220	32	464.93
F9	2019-11-05	ND<0.1	0.11	130	26	293.59
F9	2020-03-11	ND<0.1	0.14	120	21	263.13
F9	2020-05-29	ND<0.1	0.12	130	15	264.18
F9	2020-09-01	ND<0.1	0.11	140	12	273.03
F9	2020-11-17	ND<0.1	ND<0.1	130	19	274.45
F9	2021-03-29	ND<0.1	0.11	110	11	218.82
F9	2021-09-02	ND<0.1	0.14	150	15	298.59
F9	2021-11-04	ND<0.1	0.15	140	23	302.72
F9	2022-05-10	ND<0.1	ND<0.1	87	11	179.02
F9	2022-10-17	ND<0.1	0.15	140	20	295.05
F9	2023-05-03	ND<0.1	ND<0.1	90	13	189.57
F9	2023-09-25	ND<0.1	0.12	100	20	225.85
F10	2019-03-06	ND<0.1	ND<0.1	ND<0.1	ND<0.1	0.25
F10	2019-05-22	ND<0.1	0.18	0.65	0.16	1.81
F10	2019-09-04	ND<0.1	ND<0.1	0.15	ND<0.1	0.50
F11	2019-03-06	0.58	1600	1400	5.5	4470.40
F11	2019-05-22	0.25	730	1300	74	3396.42
F11	2019-09-04	0.19	700	1400	57	3479.65
F11	2019-11-05	0.16	830	1600	43	3936.11
F11	2020-03-11	0.16	240	840	25	1839.95
F11	2020-05-29	0.11	320	980	30	2186.52
F11	2020-09-01	ND<0.1	260	760	26	1716.62
F11	2020-11-17	0.12	340	1000	37	2262.66
F11	2021-03-29	0.12	390	860	38	2087.08
F11	2021-09-02	0.17	370	1000	57	2359.21
F11	2021-11-04	0.11	260	820	61	1915.95
F11	2022-05-10	0.12	280	910	100	2196.89
F11	2022-10-17	0.15	370	920	99	2325.42
F11	2023-05-03	0.2	350	790	92	2060.86
F11	2023-09-25	0.13	250	590	64	485.38
F11	2024-05-03	ND<0.1	71	350	56	840.89
F11	2024-06-28	ND<0.1	160	280	110	980.92
F11	2024-09-05	ND<0.1	67	310	47	746.80
F12	2019-03-06	ND<0.1	0.17	ND<0.1	ND<0.1	0.40
F12	2019-05-22	ND<0.1	0.17	0.4	ND<0.1	1.08
F12	2019-09-04	ND<0.1	0.49	20	1.5	39.25
F12	2019-11-05	ND<0.1	0.11	1.8	0.75	5.65
F12	2020-05-29	ND<0.1	ND<0.1	0.32	ND<0.1	1.00
F12	2023-09-25	ND<0.1	0.16	ND<0.1	ND<0.1	0.56
F7	2019-05-22	ND<0.1	31	170	53	476.07
F7	2019-09-04	ND<0.1	6.3	230	220	991.83
F7	2019-11-05	ND<0.1	0.82	170	130	641.91
F7	2020-03-11	ND<0.1	5.4	93	93	419.19
F7	2020-05-29	ND<0.1	2.1	12	13	60.10
F7	2020-09-01	ND<0.1	1.6	3.5	2.9	17.24
F7	2020-11-17	ND<0.1	0.98	2.4	2.7	13.17
F7	2021-03-29	ND<0.1	0.65	1.4	2.9	11.20
F1	2019-03-06	ND<0.1	0.16	51	2	93.08

F1	2019-05-22	ND<0.1	0.24	76	5.8	146.39
F1	2019-09-04	ND<0.1	0.2	50	16	128.85
F1	2019-11-05	ND<0.1	0.18	33	14	94.30
F1	2020-03-11	ND<0.1	0.25	40	17	114.61
F1	2020-05-29	ND<0.1	0.16	18	5.3	45.49
F1	2020-09-01	ND<0.1	0.3	21	6.6	54.32
F1	2020-11-17	ND<0.1	0.23	16	5.4	42.36
F2	2019-03-06	0.78	1100	11000	330	21194.29
F2	2019-05-22	0.44	1100	11000	1600	24552.74
F2	2019-09-04	ND<0.1	120	6300	110	11255.69
F2	2019-11-05	ND<0.1	120	7300	3800	22750.24
F2	2020-03-11	ND<0.1	42	5200	2000	14291.56
F2	2020-05-29	ND<0.1	48	5100	2900	16503.31
F2	2020-09-01	ND<0.1	13	3300	2000	10987.03
F2	2020-11-17	ND<0.1	12	3500	2100	11587.13
F2	2021-03-29	ND<0.1	14	2300	1300	7413.30
F2	2021-09-02	ND<0.1	4.1	2200	2100	9354.58
F2	2021-11-04	ND<0.1	6.1	2200	1800	8560.75
F2	2022-05-10	ND<0.1	3.6	2300	1100	6868.93
F2	2022-10-17	ND<0.1	3.6	4400	2600	14447.19
F2	2023-05-03	ND<0.1	2.9	4100	1800	11813.19
F2	2023-09-25	ND<0.1	1.6	1700	1300	6370.94
F2	2024-05-03	ND<0.1	0.61	2700	2200	10460.87
F2	2024-06-28	ND<0.1	1.2	3000	4400	16818.40
F2	2024-09-05	ND<0.1	2.3	2000	1200	6621.86
F3	2019-03-06	ND<0.1	8.6	170	33	391.99
F3	2019-05-22	ND<0.1	24	180	69	525.86
F3	2019-09-04	ND<0.1	13	71	21	195.79
F3	2019-11-05	ND<0.1	7.2	42	14	119.37
F3	2020-05-29	ND<0.1	8.8	39	4.9	91.97
F3	2020-11-17	ND<0.1	5.9	42	22	138.89
F3	2021-03-29	ND<0.1	6.2	54	23	162.44
F3	2021-09-02	ND<0.1	5.4	68	23	186.29
F3	2021-11-04	ND<0.1	5.8	70	32	213.64
F4	2019-03-06	ND<0.1	ND<0.1	59	3	109.32
F4	2019-05-22	ND<0.1	ND<0.1	70	2.1	125.73
F4	2019-09-04	ND<0.1	ND<0.1	82	2.4	147.33
F4	2019-11-05	ND<0.1	ND<0.1	54	2.9	100.59
F4	2020-03-11	ND<0.1	0.16	87	3.7	159.62
F4	2020-05-29	ND<0.1	ND<0.1	68	2.5	123.52
F4	2020-09-01	ND<0.1	0.17	110	8.8	212.94
F4	2020-11-17	ND<0.1	0.14	100	13	207.01
F4	2021-03-29	ND<0.1	ND<0.1	57	0.96	100.50
F4	2021-09-02	ND<0.1	0.16	69	3.7	128.59
F4	2021-11-04	ND<0.1	0.17	74	5.7	142.70
F5	2019-03-06	ND<0.1	ND<0.1	0.15	ND<0.1	0.50
F5	2019-05-22	ND<0.1	ND<0.1	ND<0.1	ND<0.1	0.25
F5	2019-09-04	ND<0.1	ND<0.1	0.32	ND<0.1	0.79
F5	2019-11-05	ND<0.1	ND<0.1	0.25	ND<0.1	0.67
F5	2020-05-29	ND<0.1	ND<0.1	0.11	ND<0.1	0.43
F5	2020-11-17	ND<0.1	ND<0.1	ND<0.1	ND<0.1	0.25
F5	2021-09-02	ND<0.1	ND<0.1	ND<0.1	ND<0.1	0.25
F5	2021-11-04	ND<0.1	ND<0.1	0.15	ND<0.1	0.50
F6	2023-09-25	0.5	290	1100	33	2361.02
F6	2024-05-03	0.12	140	500	4.9	1051.23
F6	2024-06-28	0.25	200	700	9.3	1484.80
F6	2024-09-05	0.33	110	330	7.7	732.19

A.2 Data characteristics

This appendix presents the statistical characteristics of the groundwater monitoring datasets used in the study. The analysis include descriptive statistics and trend-related metrics calculated from the PCE-equivalent concentrations for each monitoring well.

Well ID	# of sampling dates	CV	SD	Mean	Skewness
A1	11	2.78	12.57	4.53	3.20
A2	12	1.16	513.7	443.30	1.36
A3	17	0.58	912.00	1567.00	1.31
A4	17	1.08	224.6	208.80	2.04
B1	4	0.85	2.492	2.925	0.572
B2	4	0.36	1.797	5.05	-1.79
B3	7	1.64	3176.00	1933.00	1.41
B4	13	0.72	412.4	576.20	1.30
B5	13	0.56	1770.00	3139.00	0.37
B6	13	0.89	1204.00	1351.00	2.39
B7	7	0.66	175.6	267.50	1.48
B8	12	1.19	94.26	79.08	3.16
C1	6	0.518	2147	4149	-1.838
C2	6	1.724	94.85	55.02	2.45
C3	6	0.78	18.31	23.48	-0.07
C4	6	0.275	0.16	0.58	2.15
C5	6	0.473	4.638	9.8	0.108
C6	5	0.705	0.691	0.98	1.016
C7	6	1.492	5.174	3.467	2.421
C8	6	0.6	141.4	235.7	0.645
C9	6	0.442	0.402	0.91	-0.337
C10	10	1.685	404.6	240.1	1.676
C11	29	0.815	1082	1328	0.325
C12	10	1.338	471.5	1238	1.561
C13	25	0.303	1705	3485	0.0697
D1	7	0.90	5541.00	6130.00	0.49
D2	7	0.90	500.6	553.70	0.47
D3	7	0.36	1288.00	3601.00	-0.47
D4	7	1.15	403.6	352.00	2.08
D5	7	1.42	1.759	1.24	1.90
D6	3	N/A	0.00	0.10	2.45
E1	17	0.79	0.73	2323.00	0.19
E2	17	0.81	1.674	4016.00	-1.83
E3	6	2.25	1.934	235.40	1.79
E4	4	0.64	0.538	8.14	1.73
E5	4	1.92	2.516	657.50	1.57
E6	3	0.46	0.574	740.40	-1.72
E7	4	0.60	1.107	7374.00	-1.91
E8	5	2.04	2.197	149.40	0.97
E9	6	1.32	1.338	377.10	-0.20
E10	4	1.62	1.474	864.20	1.85
E11	4	0.90	0.792	1145.00	1.18
E12	4	1.72	1.732	855.10	1.60
E13	3	1.40	1.506	85.63	1.69
E14	5	0.81	2.035	1544.00	-1.77
E15	5	1.05	1.199	243.00	-0.40
E16	3	0.18	0.191	1339	-1.711
F1	8	0.44	0.484	89.92	-0.36
F2	18	0.43	0.422	12881.00	0.19

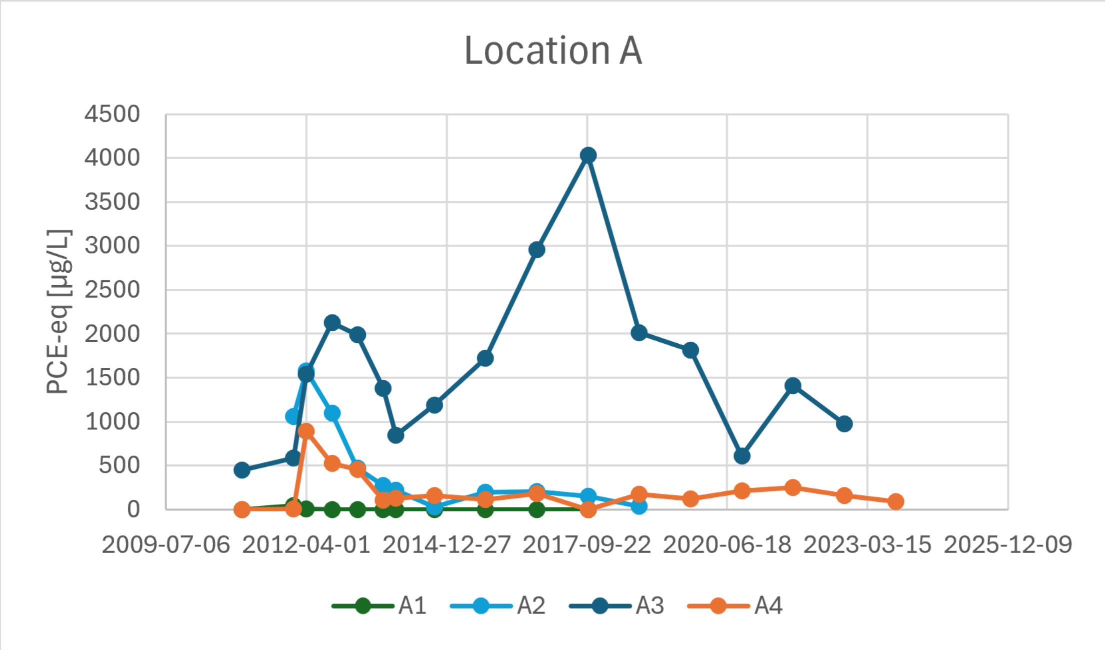
F3	9	0.63	0.552	225.10	0.72
F4	11	0.27	0.257	141.60	0.60
F5	8	0.45	0.471	0.46	-0.04
F6	4	0.50	0.499	1406.00	0.26
F7	8	1.11	1.905	328.80	-0.15
F8	5	0.35	0.388	10.34	-0.78
F9	15	0.25	0.233	276.90	0.32
F10	3	0.99	1.013	0.85	0.81
F11	18	0.50	0.593	2183.00	-0.84
F12	6	1.93	1.735	7.99	1.27

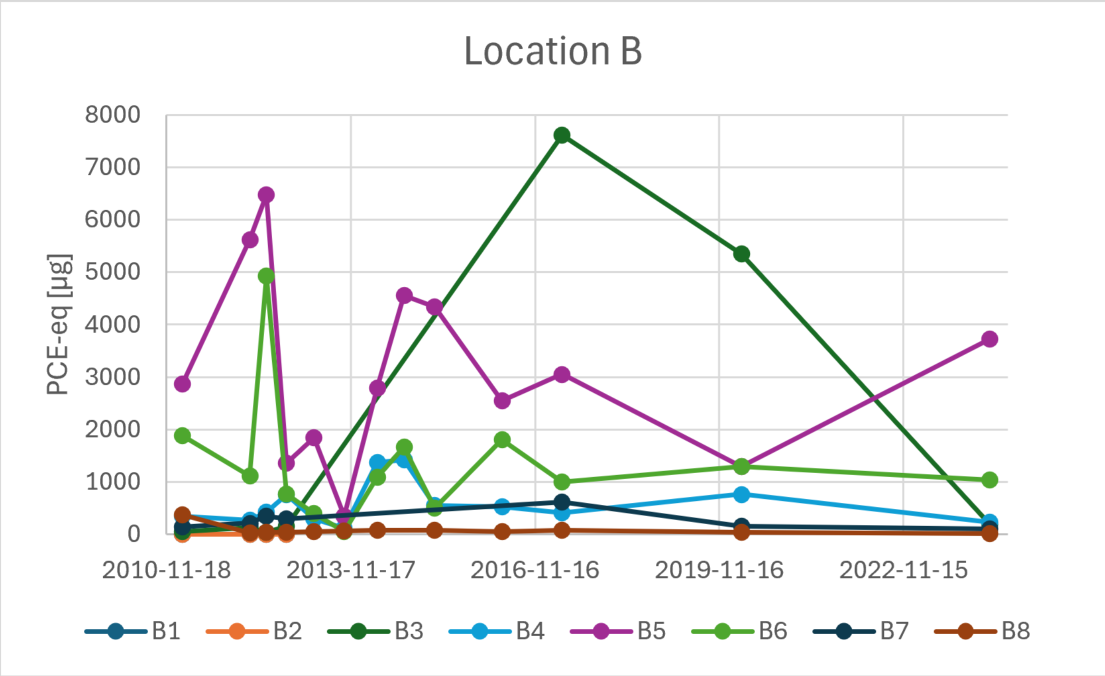
B Well plots & visual assessment

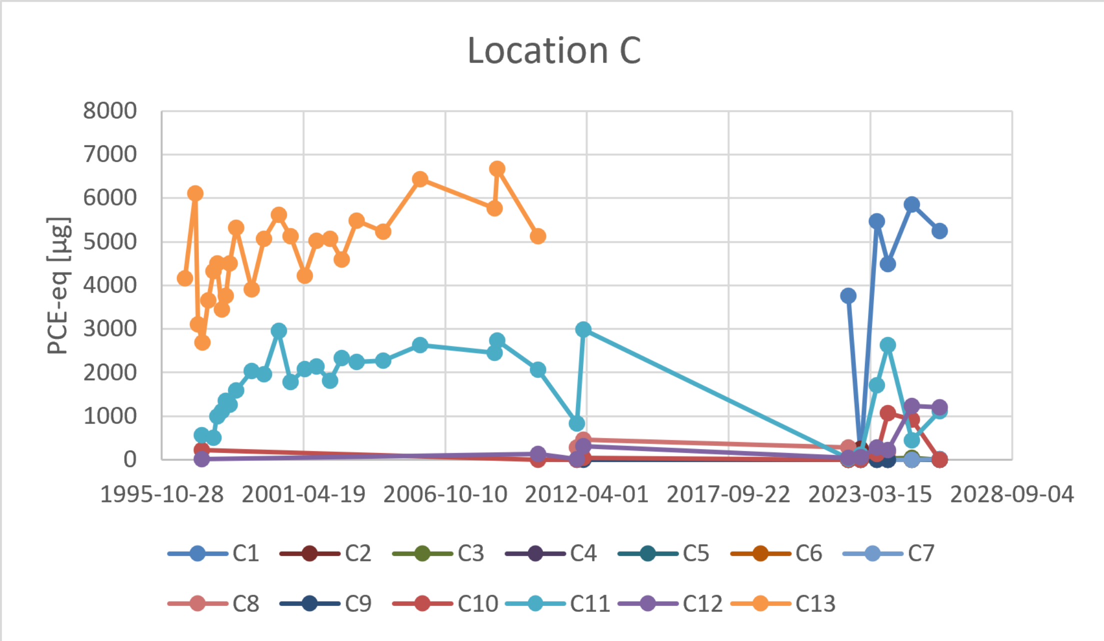
This appendix presents the graphical analyses used to support the assessment of groundwater contamination trends at the investigated locations.

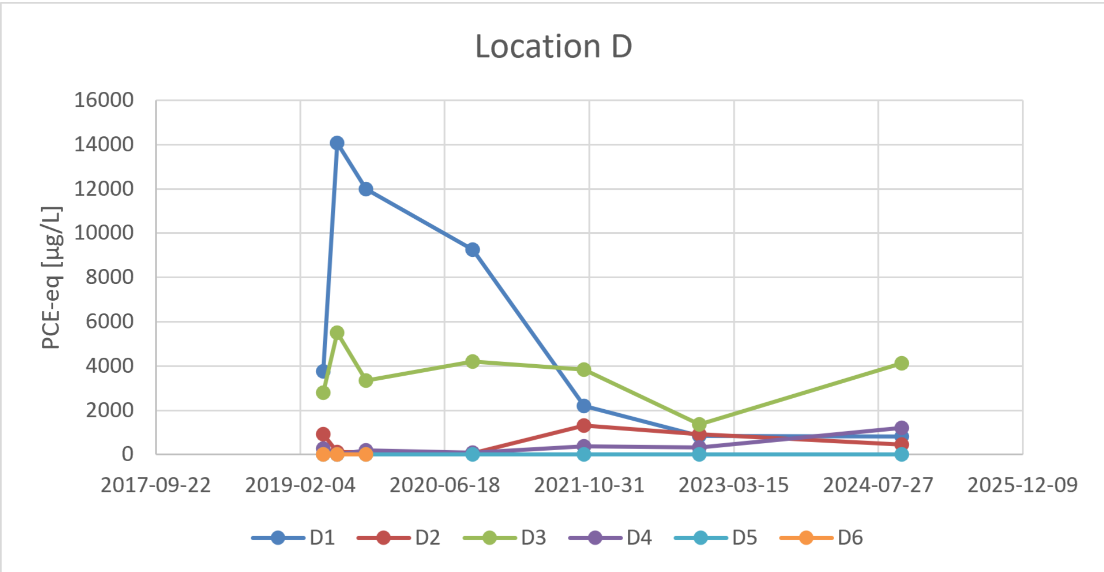
B.1 PCE-eq plots for all locations

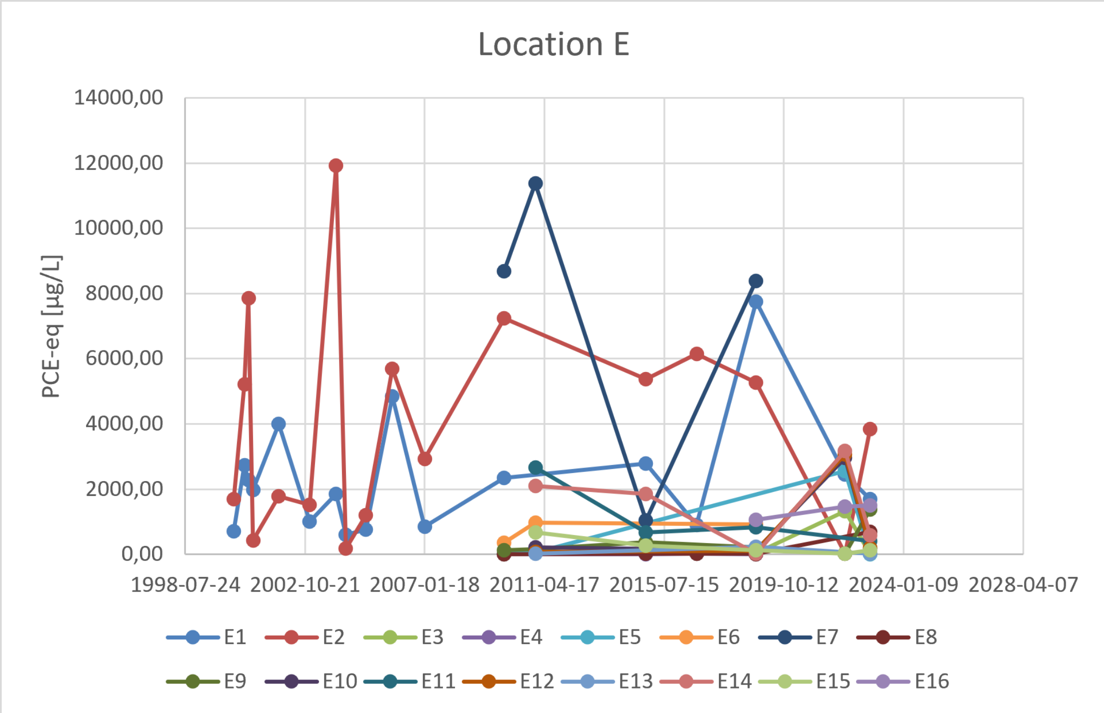
This section presents time-series plots of PCE-equivalent concentrations for all monitoring wells included in the study.

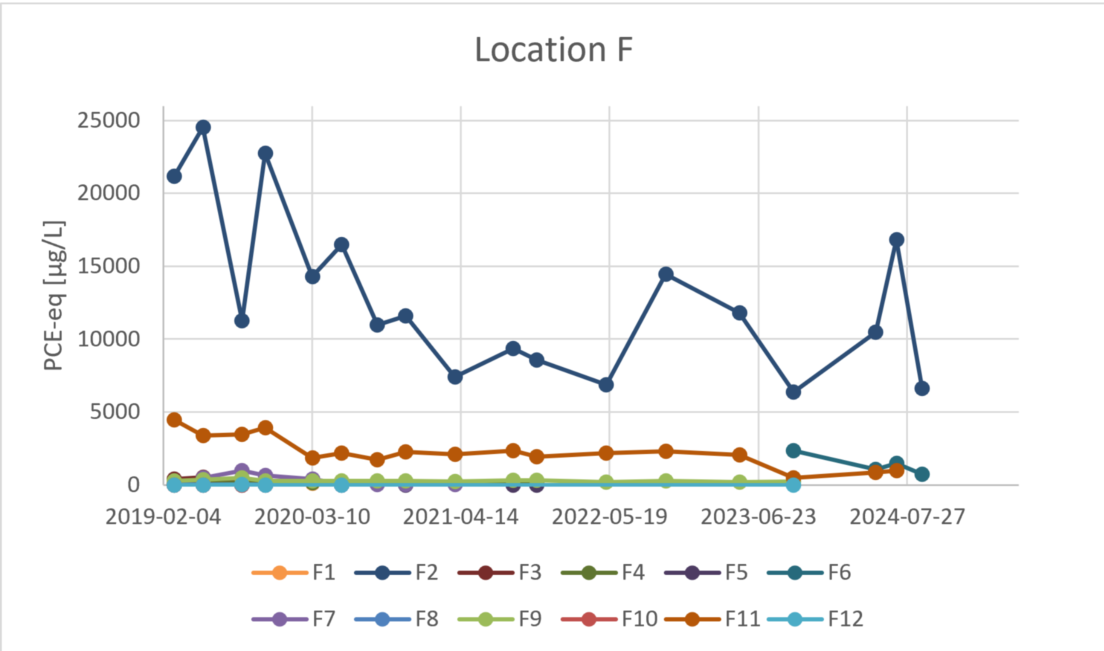





Location C
8000
7000
6000
5000
4000
3000
2000
1000
0
PCE-eq [µg]
1995-10-28
2001-04-19
2006-10-10
2012-04-01
2017-09-22
2023-03-15
2028-09-04
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13


Location D
16000
14000
12000
10000
8000
6000
4000
2000
0
PCE-eq [µg/L]
2017-09-22
2019-02-04
2020-06-18
2021-10-31
2023-03-15
2024-07-27
2025-12-09
D1
D2
D3
D4
D5
D6


Location E
14000,00
12000,00
10000,00
8000,00
6000,00
4000,00
2000,00
0,00
PCE-eq [µg/L]
1998-07-24 2002-10-21 2007-01-18 2011-04-17 2015-07-15 2019-10-12 2024-01-09 2028-04-07
E1 E2 E3 E4 E5 E6 E7 E8
E9 E10 E11 E12 E13 E14 E15 E16


Location F
25000
20000
15000
10000
5000
0
PCE-eq [µg/L]
2019-02-04 2020-03-10 2021-04-14 2022-05-19 2023-06-23 2024-07-27
F1 F2 F3 F4 F5 F6
F7 F8 F9 F10 F11 F12

B.2 Visual assessment results

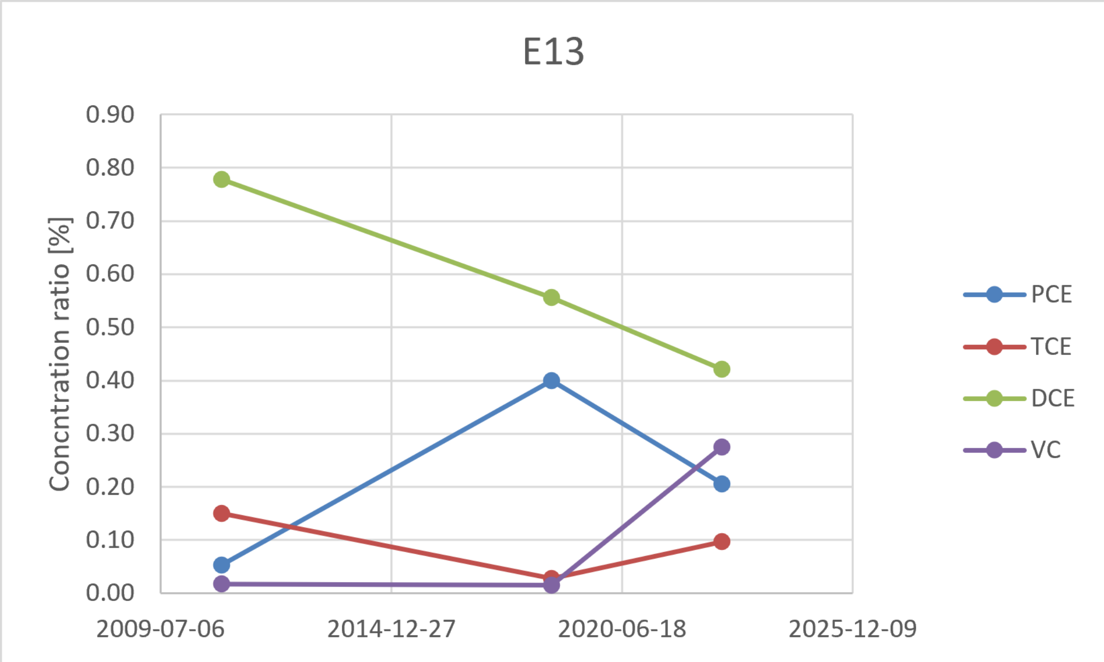
This section presents the results of the visual assessment of the PCE-equivalent concentration plots.

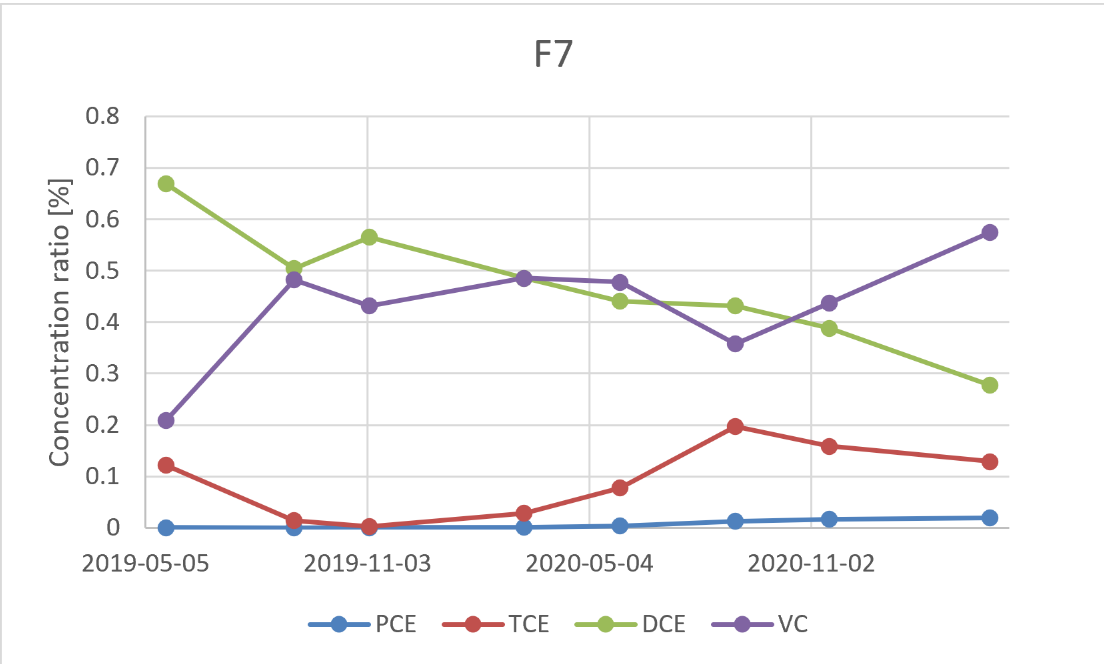
Well ID	Trend direction		Level of certainty	
	Person B	Person A	Person B	Person A
A1	Decreasing	Decreasing	Very sure	Very sure
A2	Decreasing	Decreasing	Very sure	Very sure
A3	Decreasing	Stable	Somewhat sure	Very unsure
A4	Stable	Stable	Somewhat unsure	Somewhat unsure
B1	Stable	Stable	Very unsure	Very unsure
B2	Stable	Stable	Very unsure	Very unsure
B3	Decreasing	Decreasing	Somewhat unsure	Very unsure
B4	Decreasing	Stable	Very unsure	Somewhat unsure
B5	Stable	Stable	Somewhat unsure	Somewhat sure
B6	Stable	Stable	Very sure	Somewhat sure
B7	Decreasing	Decreasing	Very unsure	Very unsure
B8	Decreasing	Decreasing	Very sure	Somewhat unsure
C1	Increasing	Stable	Somewhat sure	Somewhat unsure
C2	Stable	Stable	Somewhat sure	Somewhat sure
C3	Stable	Stable	Very unsure	Very unsure
C4	Stable	Stable	Very sure	Very sure
C5	Decreasing	Decreasing	Somewhat sure	Somewhat sure
C6	Decreasing	Stable	Somewhat unsure	Very sure
C7	Increasing	Stable	Very unsure	Somewhat unsure
C8	Decreasing	Decreasing	Very unsure	Somewhat unsure
C9	Increasing	Increasing	Very unsure	Somewhat unsure
C10	Stable	Increasing	Very unsure	Very unsure
C11	Stable	Stable	Very unsure	Very unsure
C12	Increasing	Increasing	Somewhat sure	Very unsure
C13	Increasing	Increasing	Very sure	Somewhat sure
D1	Decreasing	Decreasing	Very sure	Somewhat sure
D2	Decreasing	Decreasing	Very unsure	Somewhat unsure
D3	Stable	Stable	Very unsure	Somewhat sure
D4	Increasing	Increasing	Very sure	Very sure
D5	Stable	Stable	Very unsure	Very sure
D6	Stable	Stable	Somewhat sure	Very sure
E1	Increasing	Stable	Somewhat unsure	Somewhat unsure
E2	Stable	Decreasing	Somewhat unsure	Somewhat sure
E3	Stable	Stable	Somewhat sure	Somewhat unsure
E4	Stable	Stable	Very unsure	Somewhat sure
E5	Stable	Stable	Very unsure	Very unsure
E6	Decreasing	Stable	Very unsure	Very unsure
E7	Decreasing	Stable	Somewhat unsure	Very unsure
E8	Increasing	Increasing	Very unsure	Very unsure
E9	Increasing	Increasing	Very unsure	Very unsure
E10	Stable	Stable	Very unsure	Somewhat unsure
E11	Decreasing	Decreasing	Very sure	Somewhat unsure
E12	Stable	Stable	Very unsure	Somewhat unsure
E13	Stable	Stable	Very unsure	Very unsure
E14	Stable	Decreasing	Very unsure	Very unsure
E15	Decreasing	Decreasing	Somewhat sure	Somewhat sure
E16	Increasing	Increasing	Very sure	Somewhat unsure
F1	Decreasing	Decreasing	Somewhat sure	Somewhat sure

F2	Decreasing	Decreasing	Somewhat unsure	Somewhat sure
F3	Increasing	Increasing	Somewhat sure	Somewhat unsure
F4	Increasing	Stable	Somewhat unsure	Somewhat sure
F5	Decreasing	Stable	Very unsure	Very sure
F6	Decreasing	Decreasing	Somewhat sure	Somewhat sure
F7	Decreasing	Decreasing	Very sure	Somewhat sure
F8	Decreasing	Stable	Somewhat unsure	Somewhat sure
F9	Decreasing	Stable	Somewhat sure	Somewhat unsure
F10	Stable	Stable	Very unsure	Somewhat unsure
F11	Decreasing	Decreasing	Very sure	Very sure
F12	Stable	Stable	Somewhat unsure	Somewhat sure

B.3 Ratio of concentrations

Concentration ratio for wells E13 and F7. Example of wells where the ratio of VC increases.





C The Ricker Method

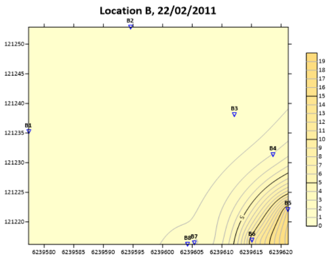
This appendix presents the results obtained using the Ricker method for the evaluation of groundwater contamination trends.

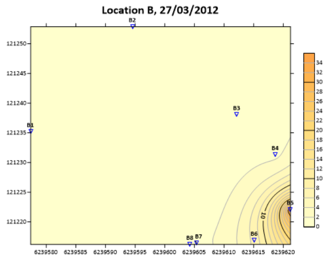
C.1 Kriging Interpolation

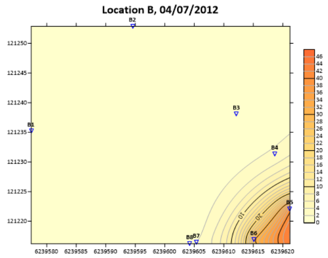
This section presents the kriging interpolation results generated for each investigated location.

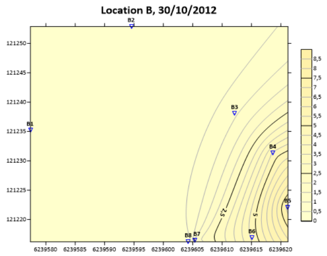
C.1.1 Location B

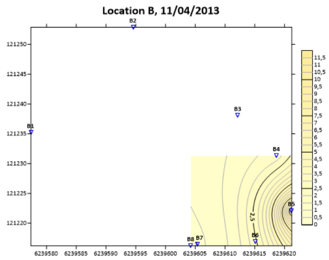
Illustrations of interpolation results using kriging for Location B.

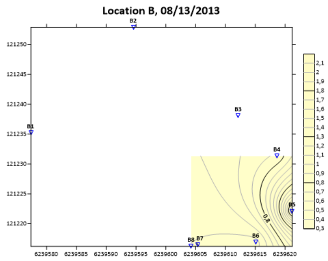


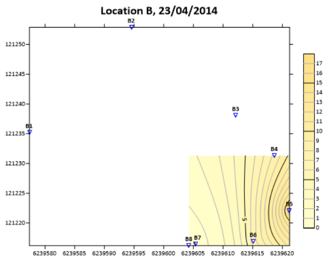


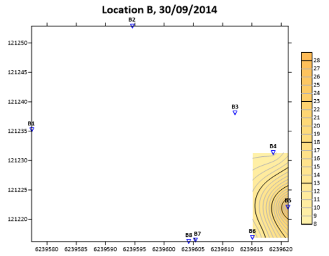


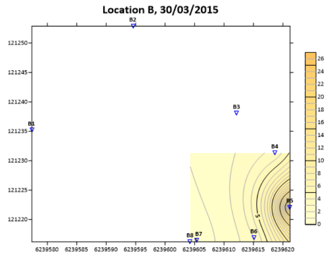


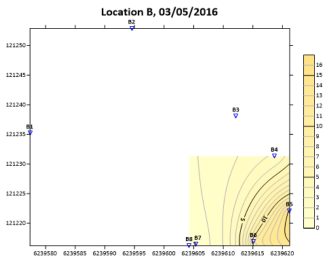


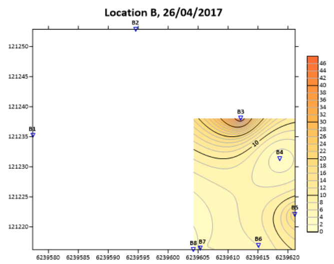


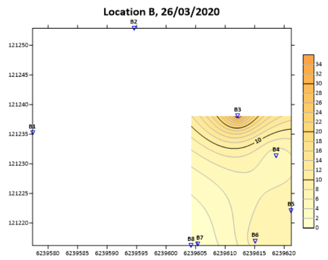


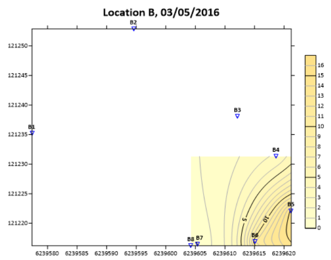






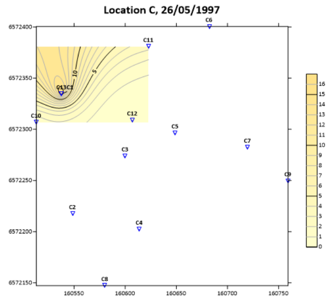


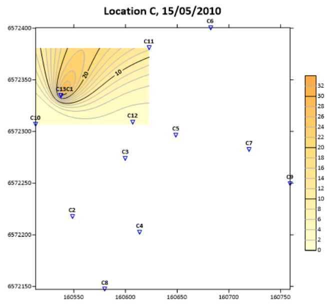


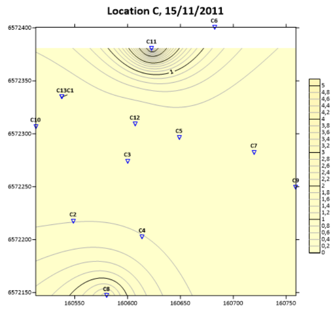


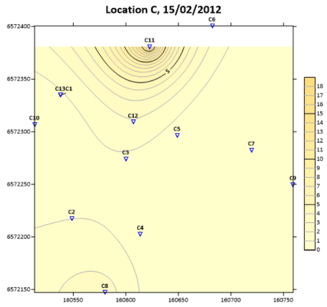
C.1.2 Location C

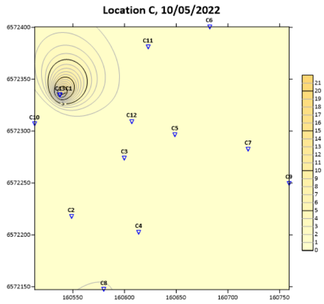
Illustrations of interpolation results using kriging for Location C.

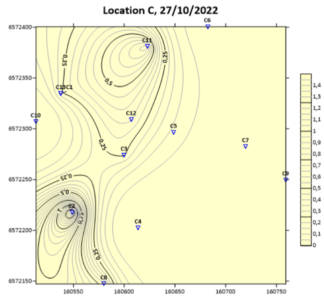


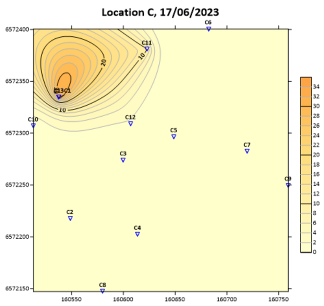


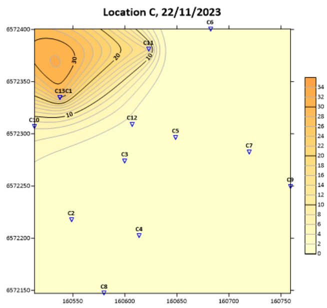


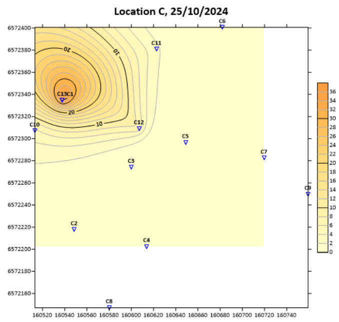


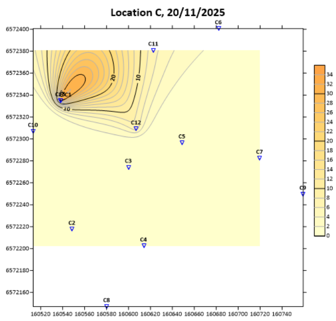






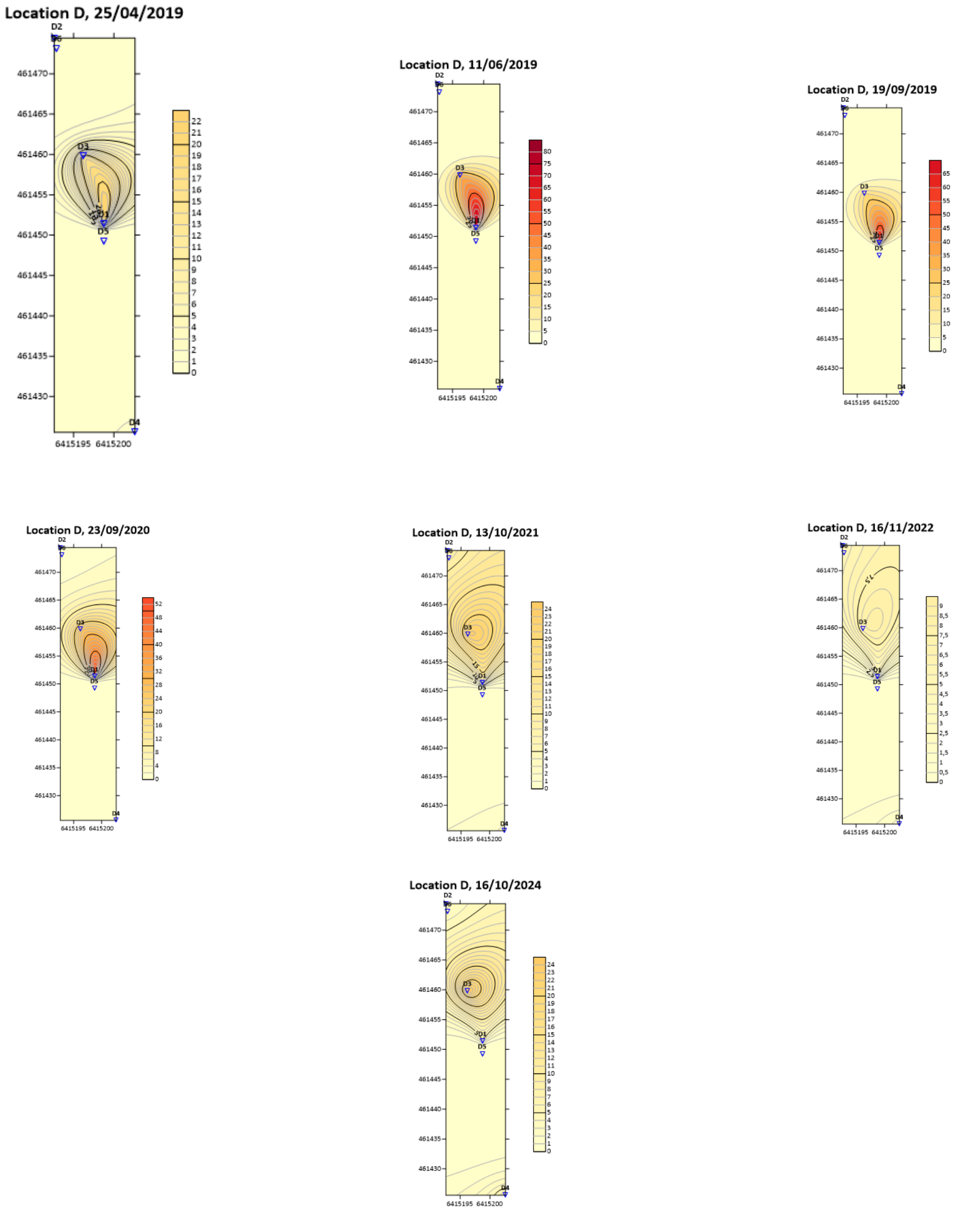






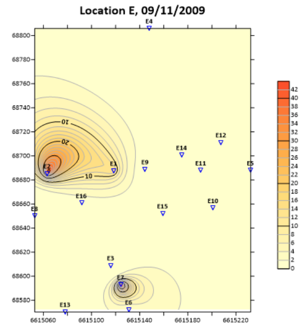
C.1.3 Location D

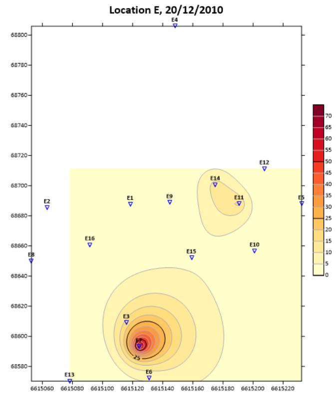
Illustrations of interpolation results using kriging for Location D.

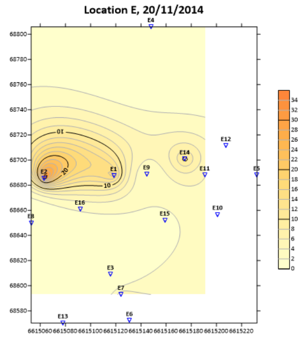


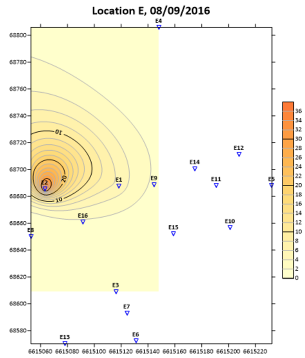
C.1.4 Location E

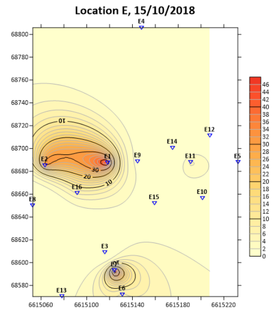
Illustrations of interpolation results using kriging for Location E.

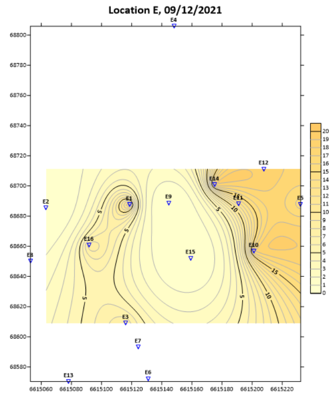


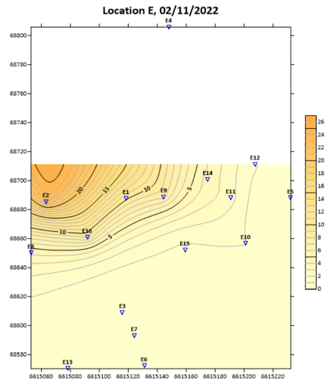






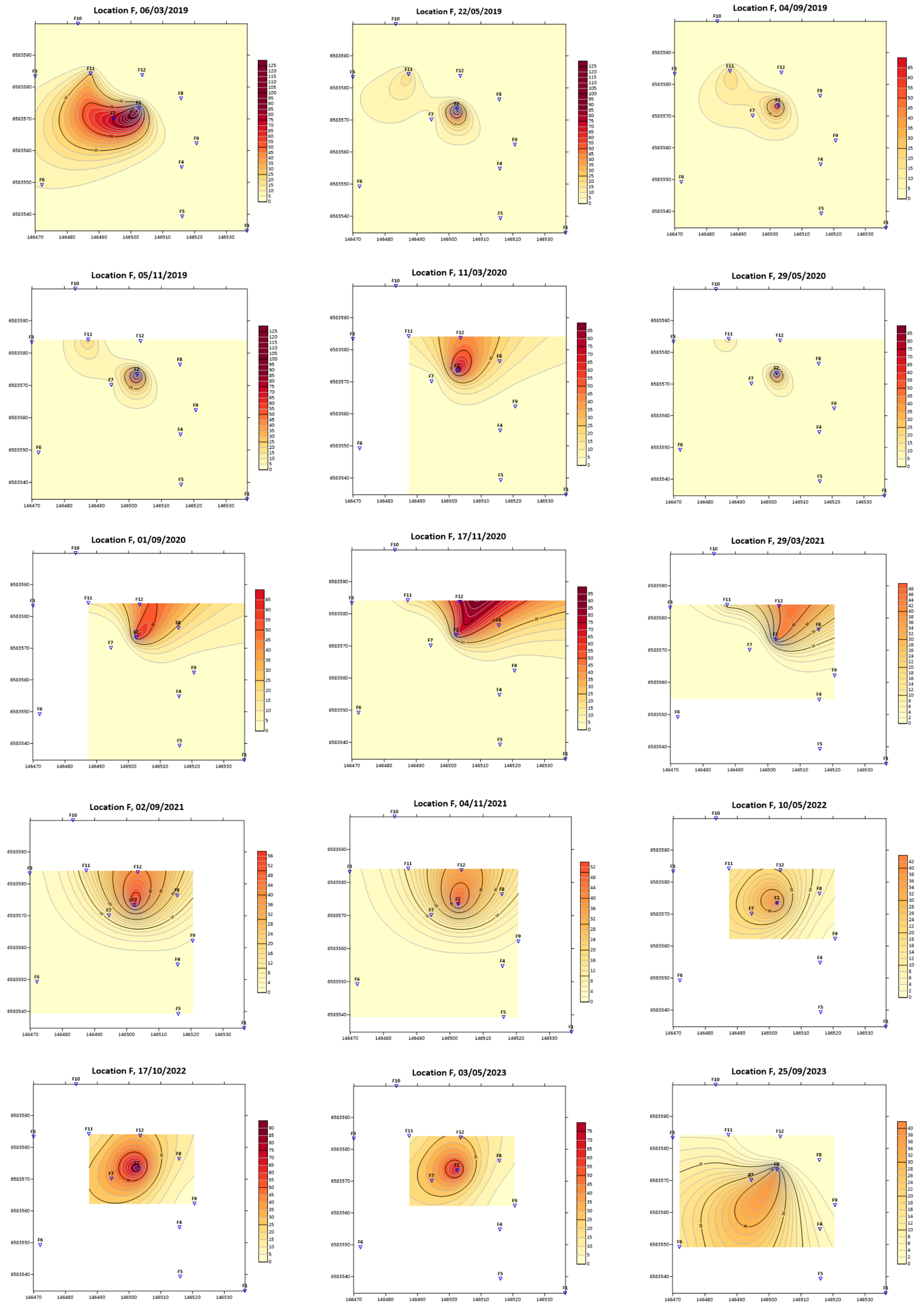






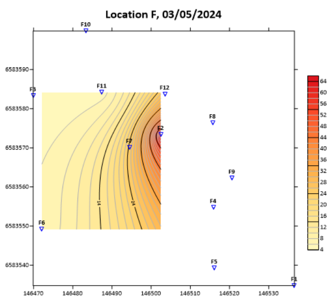
C.1.5 Location F

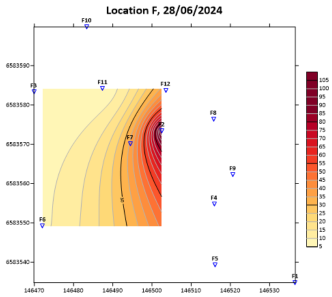
Illustrations of interpolation results using kriging for Location F.

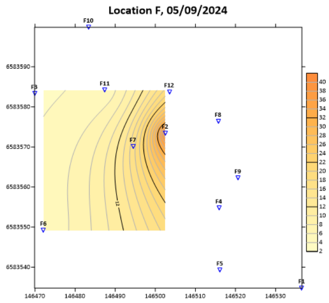


C.2 Plume area, average plume cocentration, and plume mass

This section presents the calculated plume characteristics derived from the interpolated concentration surfaces. For each location, plume area, average plume concentration, and estimated plume mass were determined.

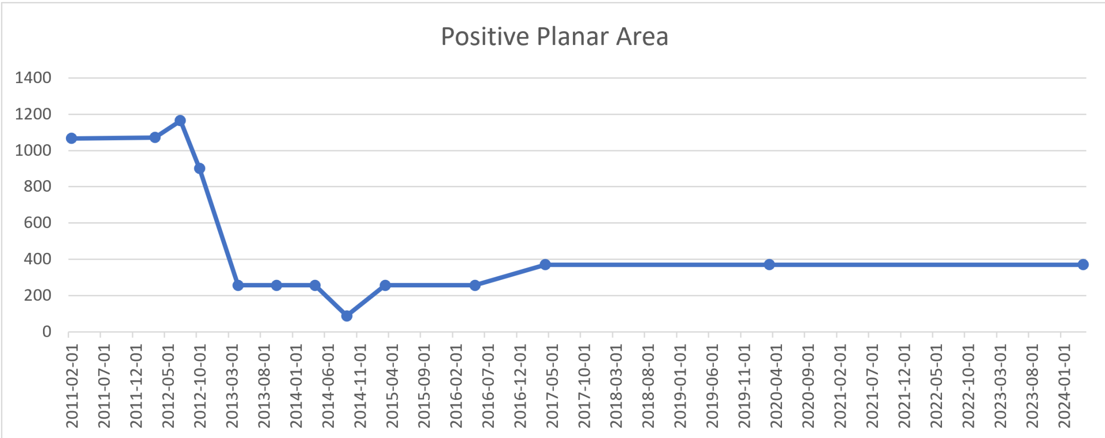


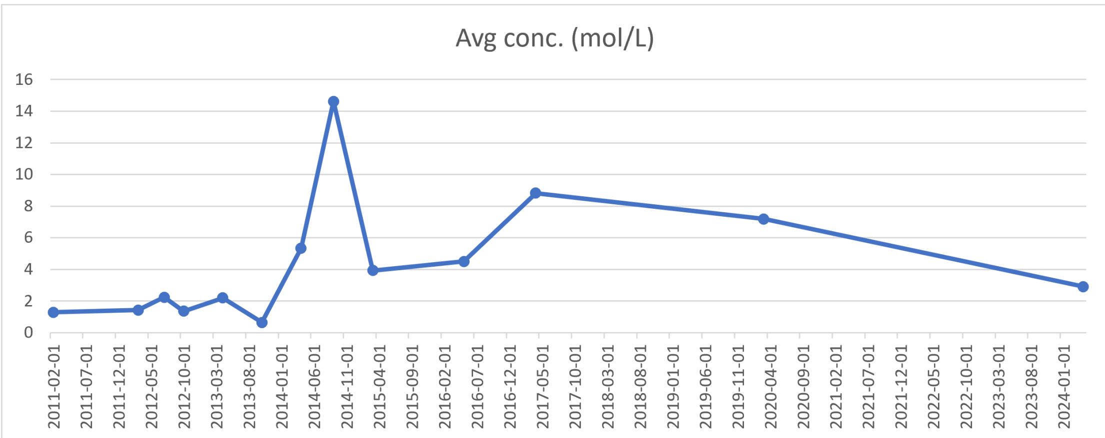


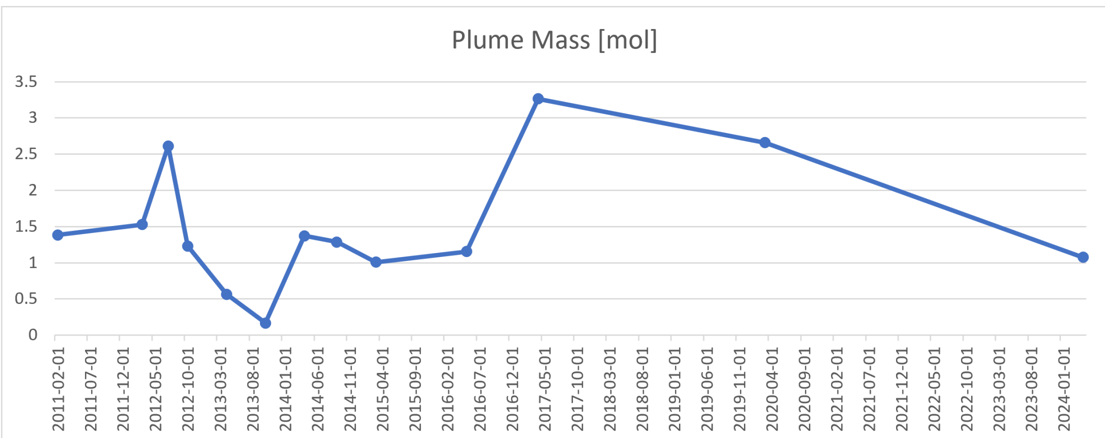


C.2.1 Location B

Plots for plume area, average plume concentration, and plume mass for Location B.

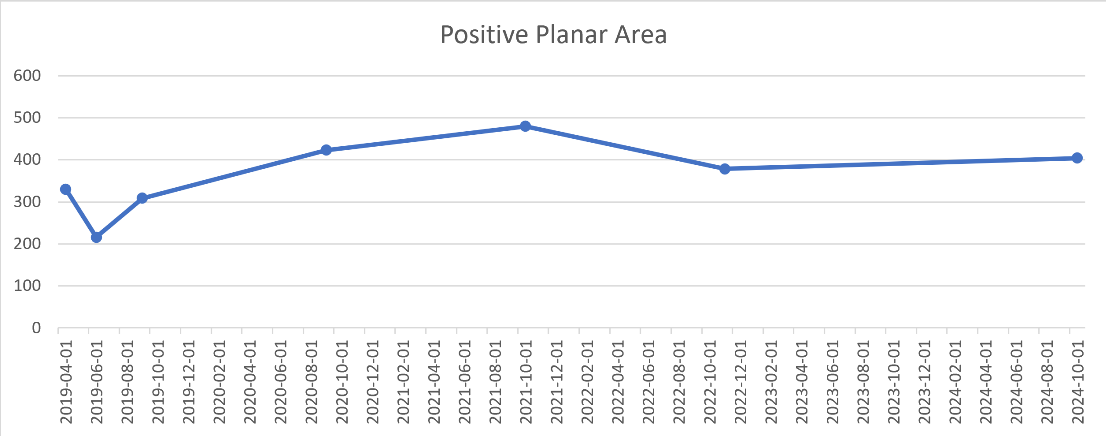


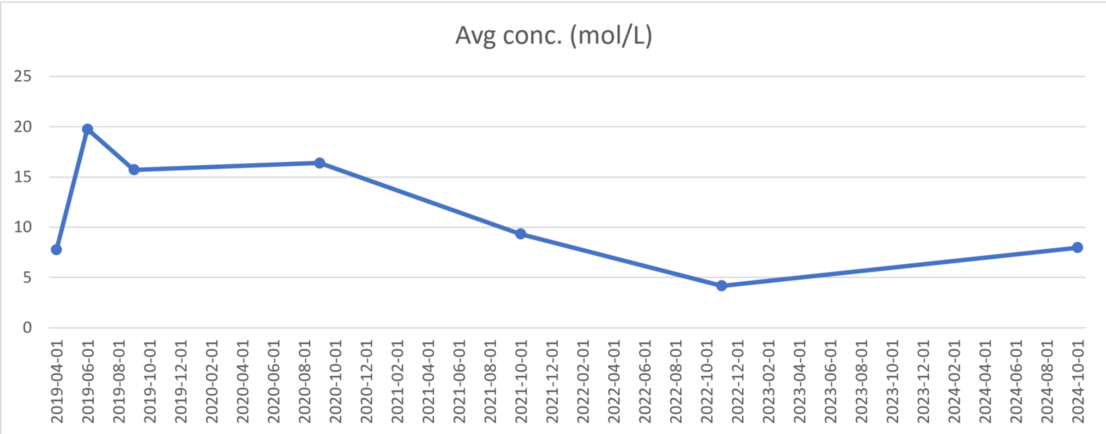


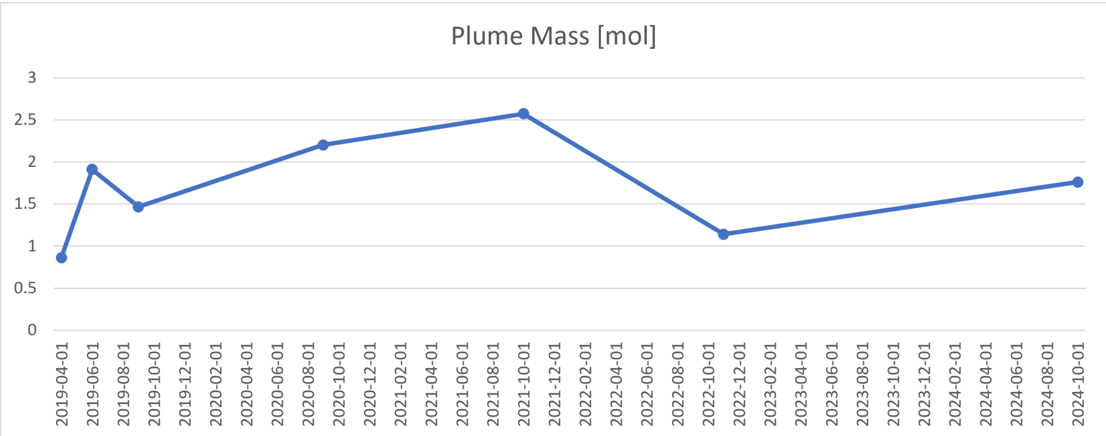


C.2.2 Location D

Plots for plume area, average plume concentration, and plume mass for Location D.

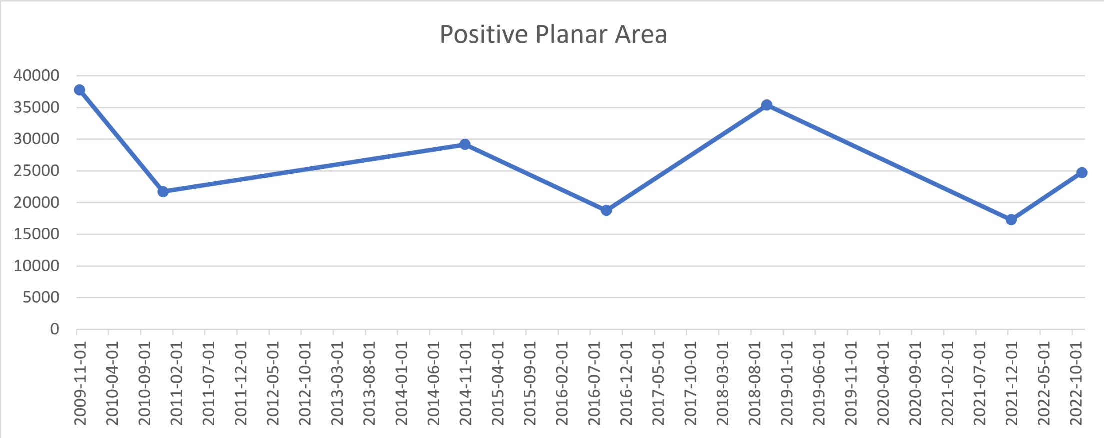


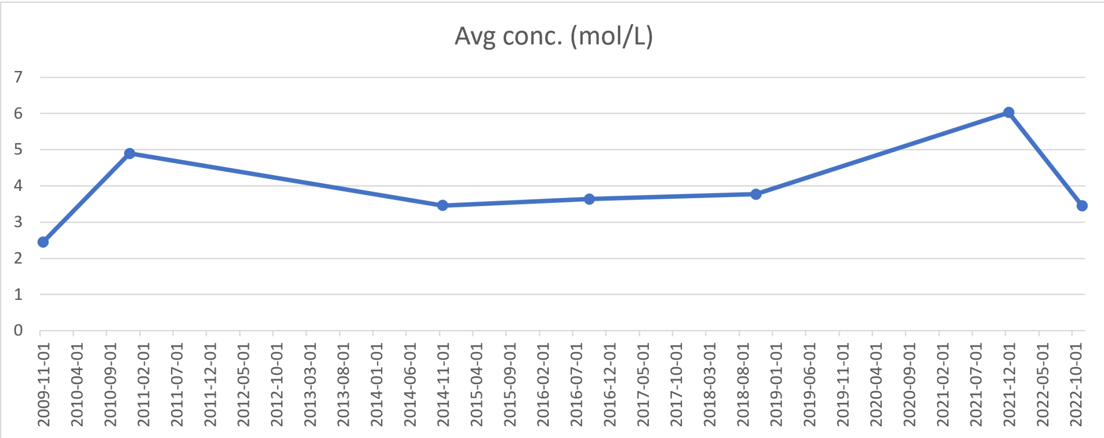


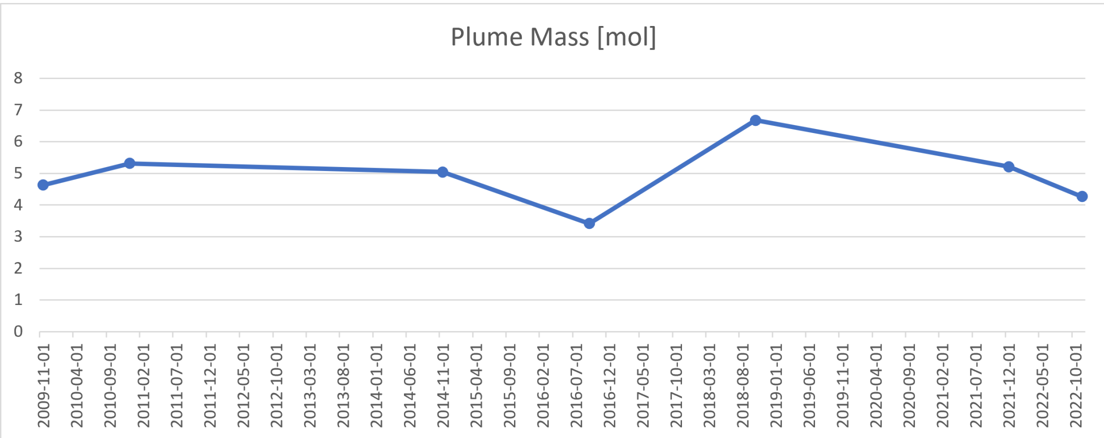


C.2.3 Location E

Plots for plume area, average plume concentration, and plume mass for Location E.

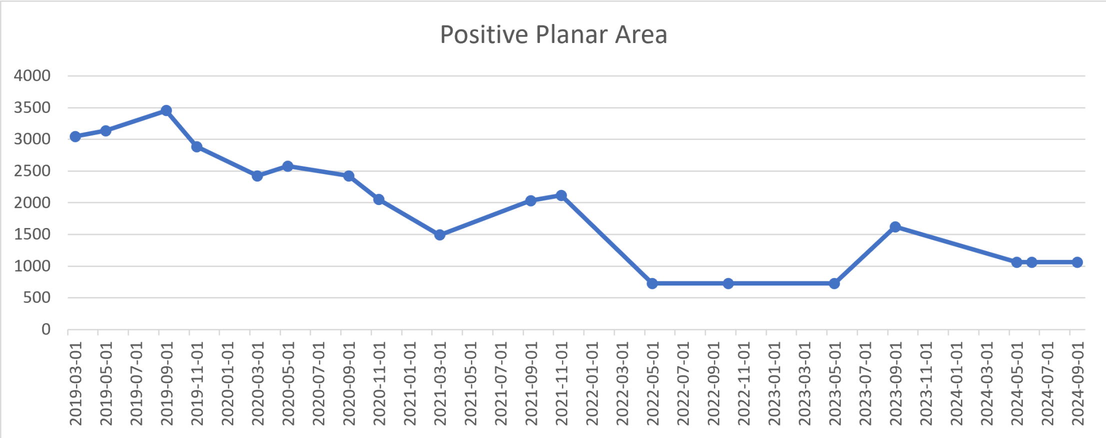


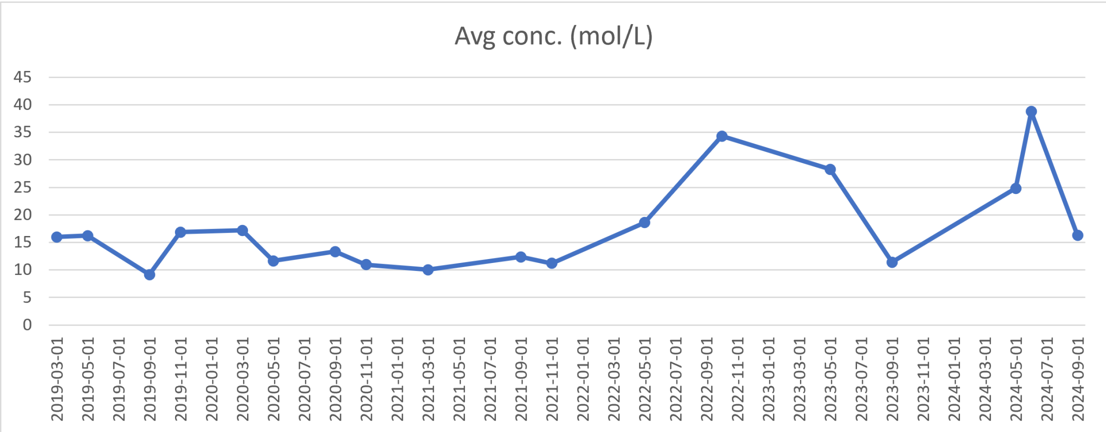


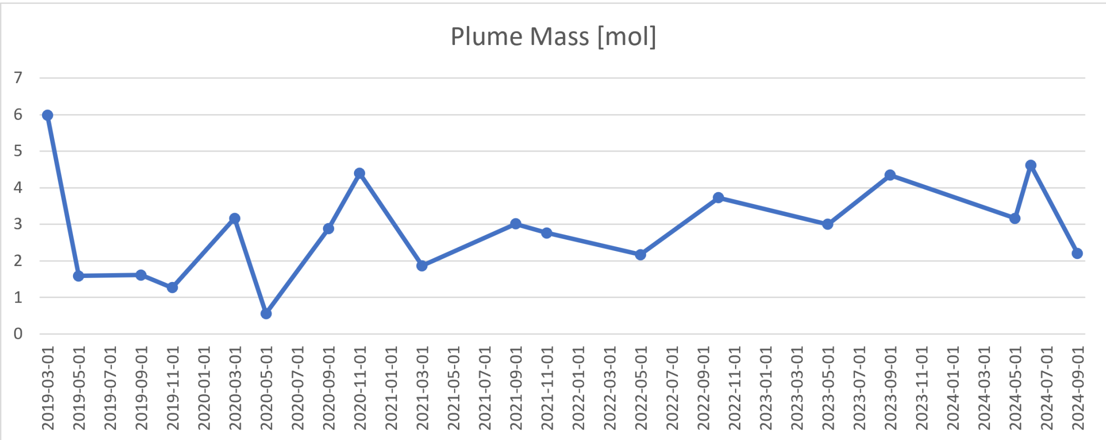


C.2.4 Location F

Plots for plume area, average plume concentration, and plume mass for Location F.







C.3 Planar areas and volumes

Planar areas and volumes extracted from Surfer using kriging as the interpolation method for Locations B, C, D, E, and F.

Location B		
Date	Positive Volume	Positive Planar Area
2011-02-22	1320,07	1066,42
2012-03-27	1462,71	1071,99
2012-07-04	2542,41	1164,93
2012-10-30	1177,88	900,59
2013-04-11	547,88	256,34
2013-10-08	153,11	256,34
2014-04-23	1355,46	256,34
2014-09-30	1279,46	87,87
2015-03-30	993,70	256,34
2016-05-03	1138,18	256,34
2017-04-26	3241,77	370,10
2020-03-26	2636,71	370,10
2024-04-09	1053,83	370,10

Location C		
Date	Positive Volume	Positive Planar Area
1997-05-26	41385,67	8104,78
2010-05-15	79881,05	8097,37
2011-11-15	9994,32	41688,17
2012-02-15	57098,78	46464,78
2022-05-10	25642,10	21724,26
2022-10-27	7843,31	28737,80
2023-06-17	153856,75	28049,58
2023-11-22	192815,24	28746,96
2024-10-25	162092,38	23238,23
2025-11-20	112863,15	20155,38

Location D		
Date	Positive Volume	Positive Planar Area
2019-04-25	1129,33	330,21
2019-06-11	2532,39	215,89
2019-09-19	1937,06	308,64
2020-09-23	2907,62	423,35
2021-10-13	3399,26	480,00
2022-11-16	1498,74	378,56
2024-10-16	2319,63	404,15

Location E		
Date	Positive Volume	Positive Planar Area
2009-11-09	90388,30	37764,14
2010-12-20	105058,87	21713,15
2014-11-20	99146,98	29171,27
2016-09-08	67179,27	18769,82
2018-10-15	131420,19	35384,99
2021-12-09	103121,61	17287,47
2022-11-02	83877,65	24736,82

Location F		
Date	Positive Volume	Positive Planar Area
2019-03-06	31732,48	3046,11
2019-05-22	8284,09	3136,94
2019-09-04	8391,14	3454,20
2019-11-05	6584,97	2886,52
2020-03-11	16715,50	2424,56
2020-05-29	2835,67	2578,35
2020-09-01	15224,95	2424,56
2020-11-17	23314,93	2051,99
2021-03-29	9873,51	1493,20
2021-09-02	15939,43	2032,32
2021-11-04	14619,97	2116,64
2022-05-10	11530,39	727,75
2022-10-17	19826,69	727,75
2023-05-03	15957,74	727,75
2023-09-25	23072,53	1621,31
2024-05-03	16828,90	1060,28
2024-06-28	24549,82	1060,28
2024-09-05	11718,60	1060,28

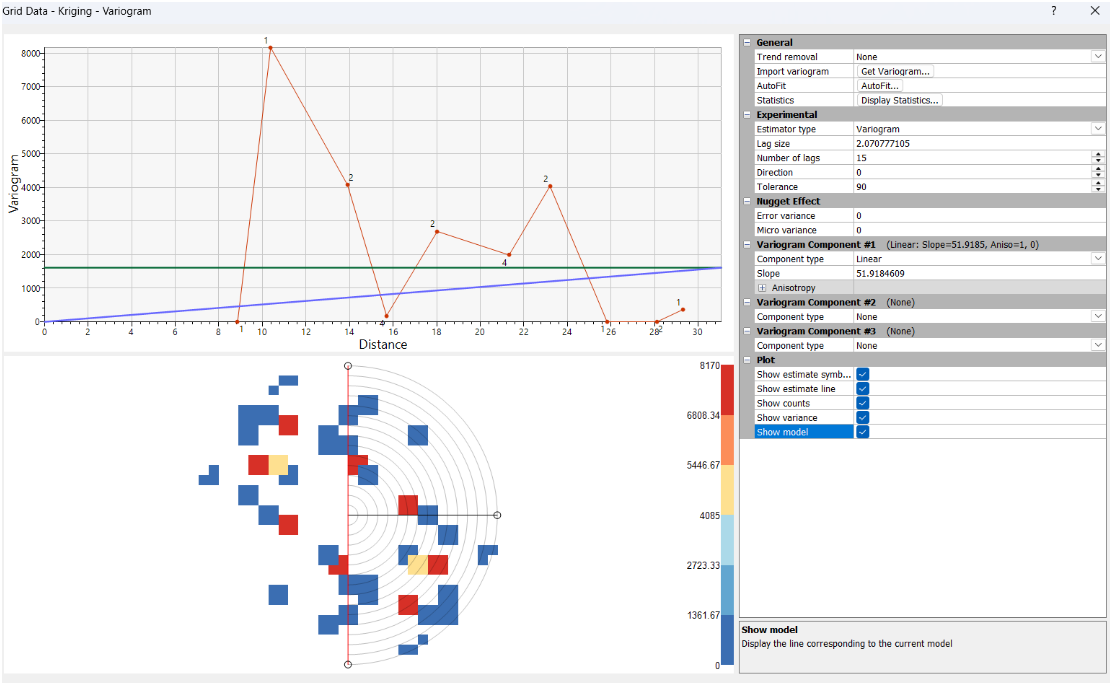
C.4 RMSE values

RMSE values from cross validation of Kriging and Nearest Neighbour for Location F, as well as the RMSE values divided by the belonging concentration range for each date. The lowest RMSE value per date is highlighted in green and the highest in red.

Date	**RMSE**		**RMSE Range**	
	Kriging	Nearest Neighbour	Kriging	Nearest Neighbour
2019-03-06	2.0865	3.1965	42%	65%
2019-05-22	1.8487	2.4501	37%	49%
2019-09-04	1.6542	1.9437	38%	45%
2019-11-05	1.5931	2.1083	35%	47%
2020-03-11	0.7613	0.9333	36%	45%
2020-05-29	2.2724	2.3751	50%	52%
2020-09-01	1.4011	1.8302	50%	65%
2020-11-17	1.4834	2.1620	32%	46%
2021-03-29	1.3537	1.9377	48%	69%
2021-09-02	1.2444	1.3200	27%	29%
2021-11-04	1.1338	1.1882	27%	28%
2022-05-10	1.1185	0.9999	71%	63%
2022-10-17	1.2838	1.1711	76%	69%
2023-05-03	1.3327	1.2070	74%	67%
2023-09-25	2.0672	2.9727	51%	73%
2024-05-03	0.8280	0.8957	76%	82%
2024-06-28	0.9208	1.0130	75%	82%
2024-09-05	0.7363	0.7739	77%	81%

C.5 Variogram

Default linear variogram for Location F, 2019-03-06.

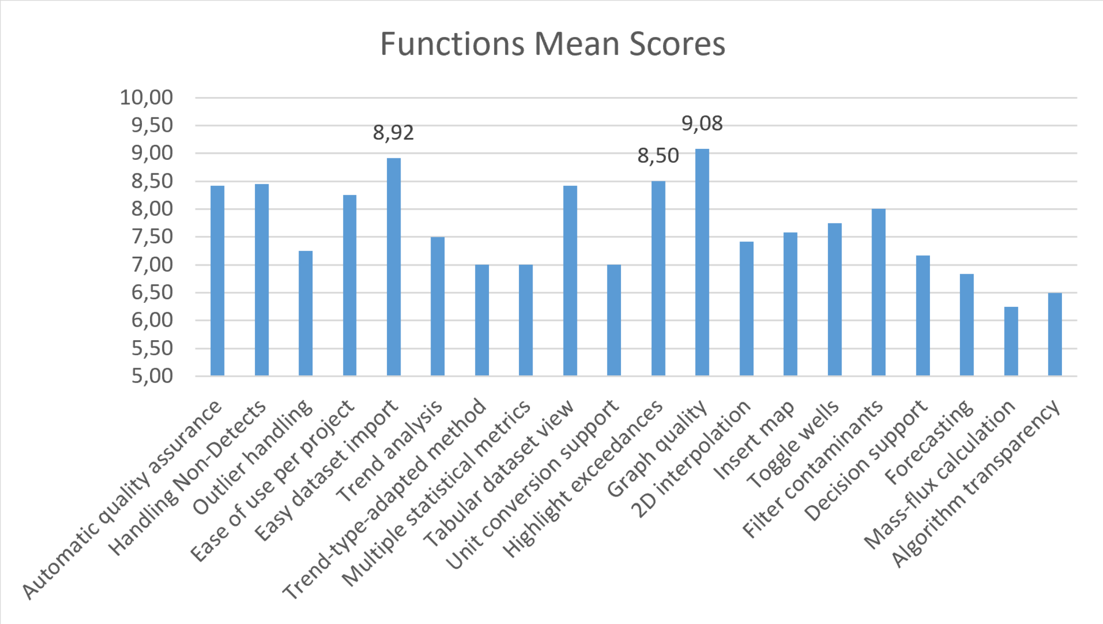


D Software

This appendix presents the supplementary material related to the software evaluation conducted as part of this study.

D.1 Result from survey

Mean score of the relevant functions or qualities received from the survey.



D.2 All identified software

Representation of all software identified during the software search. Green cells are fully evaluated software, yellow cells are software evaluated more than surface level, red cells are software not evaluated or only evaluated at surface level.

GWSDAT	Geostatistical Temporal/Spatial (GTS) Optimization Algorithm	Flowpy	CADstat	Qfield
GSI Maros	3- Tiered Monitoring Optimization (3TMO) tool	SPARROW Modeling Program	Tableau Public	JASP
PowerBI	HydroGeoSphere	Matrix Diffusion Toolkit	GenStat	AnAqSim
ModelMuse	TRENDS by T. Gerrodette	PHAST	SADA	EnviroSuite
FEFLOW	MONITOR by James P. Gibbs	GFLOW	Dynatrace	Grapher/Surfer
Visual MODFLOW Flex	EGRET	Enablon	FREEWAT	XLSTAT
ProUCL	EarthSoft EQuIS	MonitorPro (MP5)	Datadog	GSI Mann-Kendall Toolkit
ArcGIS Software	HydroGeo Analyst	Sphera	Mann-Kendall Automated (GitHub)	AquaChem
MIKE SHE	RBCA Tool Kit	Intelex	GMS (Groundwater Modeling System)	
MODFLOW 6: USGS Modular Hydrologic Model	Aermod View	Jamovi	Fracman	

D.3 Weights assigned by the respondents

Weights of software criteria from each individual input participant.

Weighing of Criteria:

Main criteria	Person 1	Person 2	Person 3	Person 4	Person 5	Weights (tot)	Weights (fin.)
Initial time investment	0,15	0,09	0,09	0,10	0,16	0,58	0,12
Data modifiability	0,20	0,18	0,17	0,24	0,16	0,95	0,19
Plots/Graphs	0,25	0,18	0,17	0,24	0,16	1,00	0,20
Additional statistical measures	0,05	0,09	0,09	0,05	0,09	0,37	0,07
Table view	0,15	0,23	0,17	0,19	0,13	0,87	0,17
Comparison values / Reference values	0,05	0,14	0,17	0,05	0,16	0,56	0,11
Intuitiveness	0,15	0,09	0,13	0,14	0,16	0,67	0,13

Subcriteria	Person 1	Person 2	Person 3	Person 4	Person 5	Weights (sub)	Weights (fin.)
Plots - Visual parameters	0,07	0,29	0,18	0,28	0,21	1,03	0,21
Plots - Simplicity of inclusion	0,07	0,12	0,18	0,11	0,21	0,68	0,14
Plots - Quality	0,29	0,12	0,24	0,28	0,21	1,12	0,22
Plots - Adding/removing data series	0,29	0,29	0,24	0,22	0,21	1,25	0,25
Plots - Multiple plots	0,29	0,18	0,18	0,11	0,17	0,92	0,18

D.4 Evaluation scores assigned by the respondents

Table D.1: Evaluation of GWSDAT based on five respondents

Criteria	P1	P2	P3	P4	P5
Initial time investment	-2	4	-1	-1	-2
Data modifiability	-1	2	0	0	-2
Plots – Visual parameters	2	3	0	1	1
Plots – Simplicity of inclusion	2	2	0	1	1
Plots – Quality	2	3	1	3	1
Plots – Add/remove data series	1	4	0	1	1
Plots – Multiple plots	2	4	2	2	1
Additional statistical measures	1	-3	1	-1	1
Table view	-4	0	-1	0	1
Reference values	2	2	1	2	1
Intuitiveness	1	5	1	-2	1

Table D.2: Evaluation of ProUCL based on five respondents

Criteria	P1	P2	P3	P4	P5
Initial time investment	-3	3	-1	0	2
Data modifiability	-1	-2	-2	-1	2
Plots – Visual parameters	-5	-2	0	0	1
Plots – Simplicity of inclusion	0	3	0	1	1
Plots – Quality	-4	-3	-1	-2	1
Plots – Add/remove data series	0	1	-1	-2	1
Plots – Multiple plots	-5	-3	1	-1	3
Additional statistical measures	4	4	2	3	4
Table view	-4	-2	0	2	0
Reference values	2	0	1	1	0
Intuitiveness	-1	-2	1	3	2

Table D.3: Evaluation of AquaChem based on five respondents

Criteria	P1	P2	P3	P4	P5
Initial time investment	-4	-3	-3	-3	1
Data modifiability	-2	1	-1	1	1
Plots – Visual parameters	2	2	0	1	1
Plots – Simplicity of inclusion	3	2	0	-1	1
Plots – Quality	3	-3	0	3	2
Plots – Add/remove data series	0	1	0	1	1
Plots – Multiple plots	4	2	-1	2	2
Additional statistical measures	4	3	1	3	4
Table view	-5	-2	-1	1	1
Reference values	0	2	1	1	1
Intuitiveness	-2	-2	-3	-2	3

D.5 S- and P-values

Mann-Kendall S-values and p-values produced by GWSDAT, ProUCL, and AquaChem for all wells.

	S				P-value			
Well ID	GWSDAT	ProUCL	AquaChem	Equal	GWSDAT	ProUCL	AquaChem	Equal
A1	-42	-38	-32	FALSE	<0.01	0,0012	0,0053	FALSE
A2	-52	-52	-52	TRUE	<0.01	0,0002	0,0002	FALSE
A3	10	10	10	TRUE	0,7110	0,3550	0,3554	FALSE
A4	1	1	1	TRUE	1,0000	0,5000	0,5000	FALSE
B1	0	0	0	TRUE	1,0000	0,6250	0,0000	FALSE
B2	-1	-1	-1	TRUE	1,0000	0,6250	0,5000	FALSE
B3	11	11	11	TRUE	0,1330	0,0680	0,0666	FALSE
B4	8	8	8	TRUE	0,6690	0,3350	0,3347	FALSE
B5	-8	-8	-8	TRUE	0,6690	0,3350	0,3347	FALSE
B6	-8	-8	-8	TRUE	0,6690	0,3350	0,3347	FALSE
B7	-1	-1	-1	TRUE	1,0000	0,5000	0,5000	FALSE
B8	0	0	0	TRUE	1,0000	-	0,0000	FALSE
C1	7	7	7	TRUE	0,2600	0,1360	0,1298	FALSE
C2	1	1	1	TRUE	1,0000	0,5000	0,5000	FALSE
C3	1	1	1	TRUE	1,0000	0,5000	0,5000	FALSE
C4	-7	-7	-7	TRUE	0,1760	0,1360	0,0880	FALSE
C5	-9	-9	-9	TRUE	0.133	0,0680	0,0664	FALSE
C6	-1	-1	-1	TRUE	1,0000	0,5920	0,5000	FALSE
C7	4	4	4	TRUE	0.566	0,2350	0,2830	FALSE
C8	-7	-7	-7	TRUE	0.26	0,1360	0,1298	FALSE
C9	3	3	3	TRUE	0.707	0,3600	0,3536	FALSE
C10	-1	-1	-1	TRUE	1,0000	0,5000	0,5000	FALSE
C11	86	86	86	TRUE	0,1110	0,3600	0,0554	FALSE
C12	25	25	25	TRUE	0,0318	0,0140	0,0159	FALSE
C13	152	152	152	TRUE	<0.01	0,0002	0,0002	FALSE
D1	-15	-15	-15	TRUE	0,0355	0,0150	0,0178	FALSE
D2	3	3	3	TRUE	0,7640	0,3860	0,3819	FALSE
D3	-1	-1	-1	TRUE	1,0000	0,5000	0,5000	FALSE
D4	11	11	11	TRUE	0,1330	0,0680	0,0666	FALSE
D5	8	8	8	TRUE	0,2880	0,1190	0,1438	FALSE
D6	0	0	0	TRUE	1,0000	N/A	0,0000	FALSE
E1	16	16	16	TRUE	0,5370	0,2680	0,2683	FALSE
E2	4	4	4	TRUE	0,9020	0,4510	0,4508	FALSE
E3	7	7	7	TRUE	0,2600	0,1360	0,1298	FALSE
E4	-2	-2	-2	TRUE	0,7340	0,3750	0,3670	FALSE
E5	0	0	0	TRUE	1,0000	0,6250	0,0000	FALSE
E6	1	1	1	TRUE	1,0000	N/A	0,5000	FALSE
E7	-2	-2	-2	TRUE	0,7340	0,3750	0,3670	FALSE
E8	6	6	6	TRUE	0,2210	0,1170	0,1103	FALSE
E9	5	5	5	TRUE	0,4520	0,2350	0,2262	FALSE
E10	0	0	0	TRUE	1,0000	0,6250	0,0000	FALSE
E11	-4	-4	-4	TRUE	0,3080	0,1670	0,1551	FALSE
E12	4	4	4	TRUE	0,3080	0,1670	0,1541	FALSE
E13	-1	-1	-1	TRUE	1,0000	N/A	0,5000	FALSE
E14	-2	-2	-2	TRUE	0,8060	0,4080	0,4032	FALSE
E15	-8	-8	-8	TRUE	0,0864	0,0420	0,0432	FALSE
E16	3	3	3	TRUE	1,0000	N/A	0,1481	FALSE
F1	-16	-16	-12	FALSE	0,0635	0,0310	0,0868	FALSE

F2	-65	-65	-65	TRUE	0,0153	0,0077	0,0077	FALSE
F3	-4	-4	0	FALSE	0,7540	0,3810	0,0000	FALSE
F4	9	9	13	FALSE	0,5330	0,2670	0,1751	FALSE
F5	-6	-6	-6	TRUE	0,5210	0,2740	0,2605	FALSE
F6	-4	-4	-4	TRUE	0,3080	0,1670	0,1541	FALSE
F7	-24	-24	-24	TRUE	<0.01	0,0010	0,0022	FALSE
F8	-4	-4	0	FALSE	0,4620	0,2420	0,0000	FALSE
F9	-27	-27	-23	FALSE	0,1980	0,0991	0,1381	FALSE
F10	1	1	1	TRUE	1,0000	N/A	0,5000	FALSE
F11	-83	-83	-87	FALSE	<0.01	0,0009	0,0006	FALSE
F12	-1	-1	-1	TRUE	1,0000	0,5000	0,5000	FALSE

DEPARTMENT OF ARCHITECTURE AND
CIVIL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY