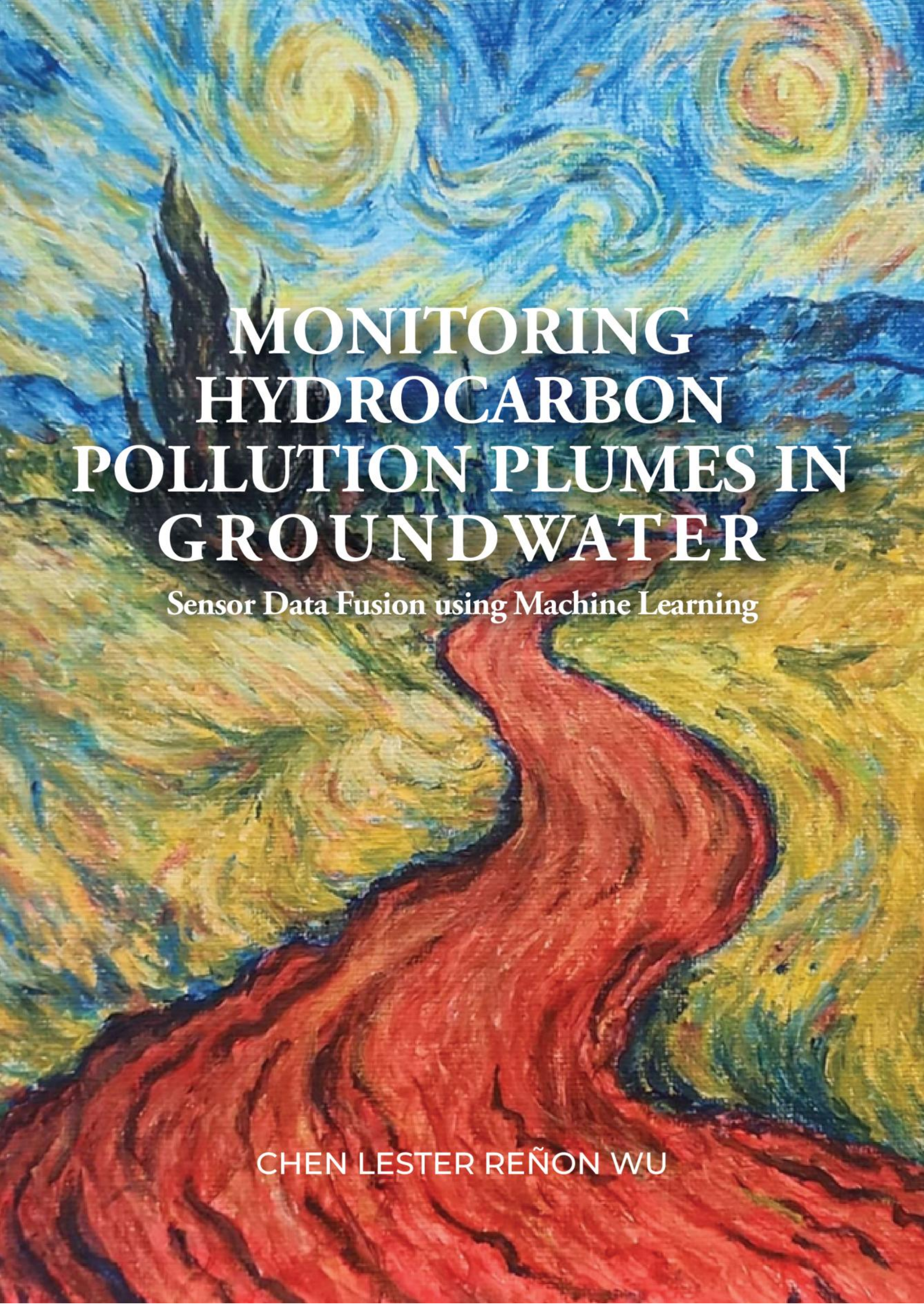
**Document Version**
Final published version

**Citation (APA)**
Wu, C. L. (2026). *Monitoring Hydrocarbon Pollution Plumes in Groundwater: Sensor Data Fusion using Machine Learning*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:87edfb29-91ff-4b45-bff4-3456189a7621



# MONITORING HYDROCARBON POLLUTION PLUMES IN GROUNDWATER

## Sensor Data Fusion using Machine Learning

CHEN LESTER REÑON WU

# Monitoring Hydrocarbon Pollution Plumes in Groundwater

## Sensor Data Fusion using Machine Learning

# Monitoring Hydrocarbon Pollution Plumes in Groundwater

## Sensor Data Fusion using Machine Learning

**Dissertation**

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus
Prof.dr.ir. H. Bijl,
Chair of the board for doctorates,
to be defended publicly on
Tuesday 12 May 2026 at 15:00

by

**Chen Lester Reñon WU**

**This dissertation has been approved by the promotors.**

**Composition of the doctoral committee:**

| | |
|---|---|
| Rector Magnificus, | chairperson |
| Prof.dr.ir. L.C. Rietveld | Delft University of Technology, promotor |
| Dr. B.M. van Breukelen | Delft University of Technology, promotor |
| Dr.ir. R.M. Wagterveld | Wetsus, external advisor |

**Independent members:**

| | |
|---|---|
| Prof.dr.ir. M. Bakker | Delft University of Technology |
| Prof.dr.ir. T.J. Heimovaara | Delft University of Technology |
| Prof.dr. J. Griffioen | Utrecht University |
| Dr. R. Taormina | Delft University of Technology |
| Prof.dr. H.M. Jonkers | Delft University of Technology, reserve member |

Keywords: reactive transport model, BTEX, virtual sensor, in-situ water quality, aquifer contamination

Printed by:
Cover by: Janine Lapiz

# Contents

# Summary

Oil remains the largest contributor to global energy consumption, powering vehicles, heating homes, and fueling the economy. This dependence on oil has led to widespread groundwater contamination, primarily through the release of light non-aqueous phase liquid (LNAPL). These contaminants typically enter the subsurface from leaking underground storage tanks, pipelines, and refineries. Once in the subsurface, LNAPLs migrate through soil and form an immiscible layer that floats on the water table. The formed layer acts as a persistent source of dissolved petroleum hydrocarbons (PHCs) and creates contaminant plumes that last for decades.

PHC compounds are highly toxic and carcinogenic, posing risks to human health and the environment upon exposure. Although groundwater serves as a major drinking water source, drinking water wells are monitored only quarterly or yearly for PHC contamination. Conventional monitoring relies on time-consuming manual sampling and costly laboratory analysis. Sensors exist for directly detecting and measuring PHC levels in the field, but most of these sensors remain experimental and/or expensive.

Microbial communities in aquifers can degrade and thereby attenuate PHCs, making monitored natural attenuation a preferred long-term management strategy for PHC contaminated sites. This is especially the case when active remediation (e.g., pump-and-treat or excavation) is technically challenging or costly. However, there is still a gap for cost-effective tools for real-time, continuous groundwater monitoring for PHC contamination.

PHC biodegradation has been reported to alter groundwater chemistry, typically lowering dissolved oxygen (DO) levels, redox potential (ORP), and pH due to microbial activities and degradation byproducts. Elevated electrical conductivity (EC) values have also been observed as being caused by degradation-induced mineral dissolution. These trends suggest that low-cost sensors could indicate PHC contamination in groundwater through sensor data fusion (SDF), a technique of integrating multiple sensor outputs for more informative estimates of target compounds based on the quantitative correlations between parameters. Despite these observable trends, there are no quantitative studies that have established clear correlations between these in-situ water quality parameters (iWQPs) and PHC concentrations.

This thesis aims to determine and assess the potential of iWQPs (i.e., pH, DO, EC, and ORP) for detecting and estimating PHC contamination in groundwater. The first step involved a literature review focused on the most hazardous fraction of PHCs:

benzene, toluene, ethylbenzene, and xylenes (BTEX). Field, laboratory, and modeling studies were synthesized in this literature review to understand the fate and transport of BTEX in groundwater, which are governed by physical, chemical, and biological processes. Important factors include aquifer heterogeneity, fluid properties, sorption mechanisms, redox conditions, and electron acceptor availability.

BTEX plume migration is driven by advection, dispersion, and diffusion, which are mainly influenced by groundwater flow and the aquifer hydraulic properties that control it. BTEX attenuation primarily occurs via terminal electron-accepting processes (e.g., respiration and fermentation) involving DO, NO3-, Fe3+, SO42-, and CH4. These processes can trigger reactions including sorption and both mineral dissolution and precipitation, which in turn affect groundwater chemistry and the iWQPs.

To overcome the lack of high-resolution field data to establish quantitative relationships between iWQPs and BTEX, a physics-based simulation approach was adopted using reactive transport modeling (RTM). RTMs couple physical transport processes with biogeochemical reactions such as degradation, aqueous speciation, and mineral precipitation/dissolution. This integration allows for realistic simulation of contaminant behavior and water quality evolution under controlled conditions. A synthetic aquifer was simulated incorporating key processes identified in the literature review. At the simulation start, the LNAPL was partitioned into multiple components: BEX compounds were grouped based on their shared migration and degradation properties, and toluene was assumed to degrade directly from the source zone.

Using the RTM-generated dataset, the temporal relationships between BEX concentrations and iWQPs were subsequently analyzed to support the development of the sensor data fusion-based ML framework, where multiple iWQP sensor signals were combined as input features. Time-series data from virtual observation wells downstream of the source zone, the long-term impacts of PHC degradation on groundwater chemistry over a 100-year duration were analyzed. Five distinct periods were observed, driven by shifts in dominant electron donors from BEX to non-volatile dissolved organic carbon (NVDOC). These transitions changed the conventional groundwater quality parameters. For instance, pH declined from 7.7 to ~6.5, while EC increased from 340 to ~950 µS/cm at a virtual observation well. Correlation analysis across the entire 100-year dataset revealed weak relationships between BEX and iWQPs due to overlapping degradation phases with varying dominant processes. However, when correlations were assessed with 5-year windows, stronger correlations were revealed. For instance, during early BEX-dominated periods, pH and EC correlated strongly with BEX (Spearman's $r \approx -0.8$ and 0.8, respectively), whereas later NVDOC-dominated periods showed stronger iWQP-NVDOC correlations and weaker iWQP-BEX correlations. This period-specific approach underscores the importance of temporal segmentation in understanding contaminant–water quality interactions, which would otherwise be masked in full timeseries analysis.

Scenario analysis also revealed that varying hydrogeologic and geochemical conditions altered the strength and direction of iWQP-BEX correlations, particularly during years between two different periods. Factors such as water table fluctuations, hydraulic conductivity, background EC, electron acceptor availability, and calcite presence influenced correlation patterns, with the absence of calcite and changes in terminal electron-accepting processes causing the most pronounced effects.

Following the correlation analysis, the next step was to translate these insights into a practical detection system. To achieve this, a machine learning (ML)-based SDF framework was developed, first focusing on early detection of BEX contamination before it reaches sensitive receptor zones such as drinking water wells. The framework was formulated as a binary classification task aimed at predicting and flagging contamination risk at virtual observation wells. To prioritize health protection, the U.S. Environmental Protection Agency's maximum benzene limit of 5 μg/L was adopted as a conservative threshold for the entire BEX group, even though ethylbenzene and xylenes have higher allowable limits. This ensured that any potential benzene presence would trigger an alarm, providing a safety margin for proactive intervention. The goal was to anticipate contamination and trigger alarms before the threshold breach.

The early detection framework focused on the initial five years of simulation, when BEX was the dominant electron donor and strong correlations with iWQPs (DO, pH, EC, and ORP) were observed. Spatial variability was captured by including the depth and horizontal distance of each virtual observation well from the contamination source in the input parameters. Training-testing split reflected the natural progression of contamination, avoiding biases from mixing data across locations and time periods. Models were trained on wells near the source, where BEX concentrations exceeded regulatory thresholds, and tested on uncontaminated downstream wells designated as warning wells.

As a proof-of-concept, five ML models, namely logistic regression (LR), random forest (RF), extreme gradient boosting (XGB), multi-layer perceptron (MLP), and support vector classifier (SVC), were trained and evaluated. Performance was assessed by comparing predicted alarms with RTM-derived arrival times at warning wells. Most models successfully triggered alarms before the 5 μg/L threshold breach, while SVC lagged. For example, at three wells, RF and XGB predicted contamination 25 to 4 days earlier than the actual threshold breach, while SVC triggered alarms 12 to 19 days late. Ensemble models (RF, XGB) performed best due to their ability to capture complex non-linear relationships. However, performance declined at far downstream wells because geochemical conditions differ significantly, leading to premature alarms. BEX arrived faster and at higher concentrations near the source, while farther wells experienced more pronounced geochemical changes (e.g., pH drop, EC rise) before contamination arrived.

Scenario simulations revealed that hydrogeochemical variability and sensor limitations strongly influence model performance. Introducing Gaussian noise and seasonal water table fluctuations degraded accuracy: at 100% combined noise and fluctuation, LR delayed alarms until BEX reached 14 μg/L (vs. 5 μg/L target), and XGB delayed alarms until 50 μg/L. To mitigate these effects, moving-average smoothing was applied to iWQPs, with a 5-day window yielding the best improvement. Despite these challenges, integrating affordable sensors with ML algorithms demonstrates strong potential for early warning systems. These results confirm that ML models could detect contamination based solely on iWQPs, even though the detected concentration deviated from 5 μg/L.

After establishing the feasibility of early detection, the next objective was to develop a continuous monitoring framework capable of estimating BEX concentrations at contaminated sites over extended periods using only time-series data of iWQPs. Therefore, the RTM-generated dataset of 100 years was used for training and testing. The input features included the same four iWQPs, DO, pH, EC, and ORP, along with well depth and distance from the source, as used in the early detection framework. Focus was given on sequence-learning models for estimating BEX concentration, specifically a long short-term memory (LSTM) network due to its ability to capture temporal dependencies in groundwater systems.

The LSTM network was trained on daily data at virtual observation wells from each five-year segment, tested on the subsequent year's data, then advanced by one year for retraining. This annual update was chosen to capture full hydrological cycles, while maintaining computational efficiency. Although more frequent retraining could improve performance in practice, the annual benchmark provided a conservative baseline for evaluating robustness across time periods. The five-year window length balanced stability and responsiveness, avoiding overfitting to short-term fluctuations while adapting to recent trends.

LSTM performance varied across test windows: while $R^2 \geq 0.80$ was not always achieved, the model showed strong practical potential. Over 70% of test windows at each observation well had a mean absolute percentage error below 5%, and 87% were below 10%, which is acceptable for general groundwater monitoring.

Recognizing the mechanistic link between groundwater flow and contaminant transport, hydraulic head was added as a predictor which considerably improved the LSTM performance. At the farthest well, windows with $R^2 \geq 0.80$ increased from 2 (~2%) to 35 (~37%) when hydraulic head was included. Overall, LSTM generally outperformed the classical models (MLR, SVR, RF, and XGB) when geochemical processes were more complex such as periods with changing dominant electron acceptor. However, the classical or hybrid approaches performed better during periods with strong instantaneous iWQP-BEX correlations.

Extreme hydrological changes, such as a tenfold increase in hydraulic gradient, degraded LSTM performance. While the LSTM failed to predict exact concentration magnitudes under these conditions, it correctly identified directional trends (increasing or decreasing) in six of eight virtual observation wells. This offers a practical anomaly detection capability for triggering manual sampling campaigns. Post-source removal scenarios further highlighted model limitations: the LSTM assumed steady-state conditions and failed to capture the rapid decline in BEX concentrations. A hybrid ML-Kalman filter approach was then developed to mitigate this limitation. The decreasing trend was successfully captured using the simple filter after incorporating periodic BEX measurements. These results suggest that combining ML with recursive algorithms such as the Kalman filter and expanding training datasets to include diverse hydrological and geochemical conditions are essential for continuous BEX estimation.

The findings from early detection to continuous BEX estimation led to the conclusion that iWQPs have strong potential for detecting and estimating PHC contamination in groundwater, with particular emphasis on BEX. Combining ML with domain-specific knowledge is essential to ensure these ML-based frameworks function effectively. To enhance the reliability and practicability of these findings, future work should focus on: 1) incorporating sorption and desorption processes into RTM simulations to better represent contaminant behavior; 2) extending applicability to other contaminants beyond PHCs; 3) developing affordable in-situ BTEX sensors to complement the ML-based approach 4) validating the ML-based frameworks under real field conditions; and 5) integrating environmental and anthropogenic factors as additional input into the ML models.

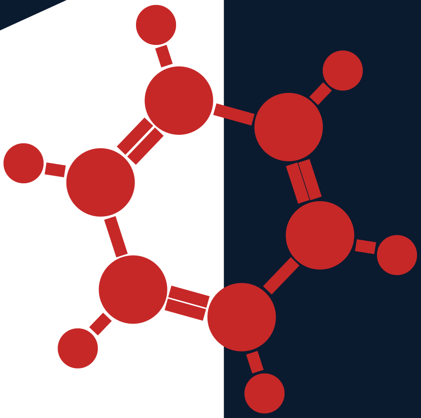
*It is not in the pursuit of dreams that we must perish.*

*In daring to dream, we choose to truly live.*

*And perhaps, that is the only kind of immortality worth chasing.*

# General Introduction

# 1



*“Oil creates the illusion of a completely changed life — life without work, life that comes free. Oil is a resource that anesthetises thought, blurs vision, corrupts. People from poor countries go around thinking, God, if only we had oil! The concept of oil expresses perfectly the eternal human dream of wealth achieved through lucky accident, through a kiss of fortune, and not through sweat and hard work. In this sense, oil is a fairy tale and, like every fairy tale, a bit of a lie.” – R. Kapuściński, 1982*

# 1. General Introduction

## 1.1 Petroleum as a Promise and a Curse

Often referred to as ‘black gold’, petroleum has fueled the advancement of the global industry for more than a century (Sorenson, 2020). Its derivatives are vital to modern infrastructure, from powering transportation systems (e.g., gasoline, diesel, jet fuel) to serving as essential feedstocks for manufacturing (e.g., plastics, chemicals). As of 2024, oil remains the world's primary energy source despite a growing shift toward renewables; oil accounts for approximately 34% of total global primary energy consumption (Energy Institute, 2025).

However, our reliance on petroleum, which is a major type of fossil fuel, has adverse environmental impacts. The extraction, refinement, and use of fossil fuels as energy source are leading drivers of climate change and contribute significantly to atmospheric, aquatic, and terrestrial pollution (Wuebbles and Jain, 2001; Nayak et al., 2019). While water use in energy production has traditionally focused on the volume required, the impacts on water quality cannot be overlooked (Allen et al., 2012). Among these impacts, groundwater contamination is particularly widespread and pervasive. There is considerable potential for accidental releases into the environment, especially since the global petroleum consumption was estimated at nearly 100 million barrels per day in 2022 (International Labour Organization, 2022).

Petroleum compounds enter the subsurface primarily through the release of light non-aqueous phase liquid (LNAPL), which consists of a mixture of petroleum hydrocarbons (PHC). LNAPL is introduced into the environment when oil leaks from anthropogenic sources such as underground storage tanks, especially at gasoline stations. Larger-scale infrastructure with extensive fuel handling operations also contributes significantly to oil leakage, including oil pipelines, decommissioned refineries (**Figure 1.1**), bulk storage terminals, and airports (Interstate Technology and Regulatory Council [ITRC], 2009). Once released, LNAPL migrates through the soil and accumulates within both the unsaturated (vadose) and saturated (phreatic) zones, forming an immiscible phase (Naval Facilities Engineering Command [NFEC], 2010). Due to its lower density relative to water, LNAPL tends to accumulate near the groundwater table, occurring as a zone of variable saturation that extends both above and below the water table.

This layer acts as a persistent and continuous source of dissolved PHCs and their toxic constituents, such as benzene, toluene, ethylbenzene, and xylene (BTEX), which exhibit dynamic behavior. Their transport and long-term persistence are governed by heterogeneous lithology, complex hydrogeology, and biogeochemical reactions. As a result, a PHC plume is formed, which can endure for decades or even centuries after the initial contamination event occurred (Eberhardt and Grathwohl, 2002; D'Affonseca et al., 2011).

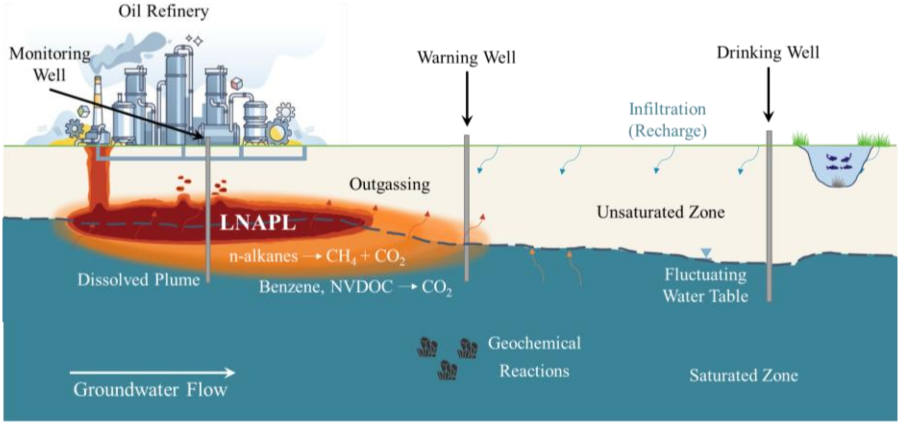


**Figure 1.1.** Schematic of a petroleum refinery illustrating a possible hydrocarbon release into groundwater and posing risks to downstream drinking water wells. Failures in corroded pipelines, faulty valves, and compromised connections during crude oil transfer and product storage can lead to groundwater contamination. Dissolved hydrocarbon plumes can trigger complex hydrogeochemical reactions.

## 1.2 LNAPL Management Strategies

Long-term persistence of PHC plumes makes groundwater contamination particularly challenging. The risk arises when dissolved-phase contaminant plumes, containing highly toxic constituents, migrate toward sensitive receptors, such as downstream drinking water wells. Benzene, a known carcinogen, has a drinking water limit of only 5 µg/L in the United States, meaning even trace concentrations pose significant health hazards (U.S. Environmental Protection Agency [US EPA], 2024). In standard practice, the initial response to an oil spill focuses on managing LNAPL and the contaminant plume generated by LNAPL constituents dissolving into groundwater. This involves characterizing the LNAPL and the contaminated site, identifying regulatory remediation requirements, and defining a management strategy including treatment methods (NFEC, 2010). Two treatment methods are generally considered: active treatments, such as excavation, pump and treat, and enhanced in-situ destruction or degradation of dissolved-phase LNAPL constituents; and passive treatments, including monitored natural attenuation (MNA) and permeable reactive barriers. Active methods are suitable when LNAPL poses an immediate risk, is migrating, or can be recovered cost-effectively (ITRC, 2009), whereas passive methods target the contaminant dissolved-phase plume from the LNAPL.

It is common to integrate multiple strategies when dealing with LNAPL. Active and passive treatment methods can be applied concurrently to address different groundwater contaminant pathways. These can also be applied sequentially by first removing the bulk contamination, followed by a final cleaning phase that removes the remaining LNAPL to achieve the remediation objectives (NFEC, 2010). Another approach is the direct degradation of LNAPL, a strategy that has gained considerable attention. Known as natural source zone depletion (NSZD), this process involves the partitioning of LNAPL compounds into other phases, followed by complete or partial mineralization into biodegradation products (Garg et al., 2017, Sookhak Lari et al., 2025). The goal is to reduce or eliminate the risks associated with LNAPL.

### 1.3 Monitored Natural Attenuation

MNA is the preferred long-term LNAPL and dissolved phase plume management strategy when active remediation becomes technically impracticable or prohibitively expensive to maintain. It is defined as the reliance on natural processes to achieve site-specific remediation objectives within a reasonable time frame. MNA specifically involves the reduction of mass, volume, concentration, and toxicity of the dissolved contaminant plume; these processes include dilution, dispersion, volatilization, sorption, and biodegradation (Declercq et al., 2012). Microbial degradation of PHCs involves the breakdown of contaminants into less harmful end products and is considered the primary attenuation mechanism (McGuire et al., 2018).

Evaluating the effectiveness of natural attenuation for LNAPL remediation is based on evidence-based frameworks. Primary evidence consists of long-term monitoring to demonstrate reductions in contaminant concentrations. This is typically collected quarterly over at least two years. Secondary evidence relies on geochemical and hydrogeological indicators to estimate the rates of active attenuation processes. Tertiary evidence involves site-specific studies such as isotope analyses or microcosm experiments to directly confirm these processes (Beck and Mann, 2010). As this framework becomes increasingly more detailed starting from primary to tertiary, regulatory guidelines generally require primary and secondary evidence; tertiary evidence data are used only when uncertainty still exists.

The main goal of the primary evidence is to show the temporal and spatial trends of contamination concentration. Thus, extensive and frequent data collection over several years is often necessary to account for seasonal variability or localization to accurately estimate degradation rates. This is especially important since the allowable benzene concentration of 5 µg/L in the United States is orders of magnitude lower than its maximum solubility of 1,780 mg/L in water at 25°C (Agency for Toxic Substances and Disease Registry, 2007). However, conventional monitoring for MNA relies on a costly and time-consuming process of manual groundwater sampling, followed by expensive off-site laboratory analysis for contaminants and geochemical indicators (Arambarri et al., 2004). This traditional approach provides only sparse, discrete time "snapshots" of subsurface conditions, inevitably missing critical temporal variations and contaminant fluxes. As a result, it cannot capture dynamic plume behavior or respond promptly to abrupt changes such as active source removal or shifts in hydraulic gradients. The expense of comprehensive, long-term monitoring becomes financially unsustainable for many sites.

Currently, sensor-based approaches for in-situ BTEX detection are mostly at the research stage, with only a few commercially available options. Experimental techniques include: gold nanoparticle chemiresistor sensor arrays, which provide a response pattern of the electrical resistance depending on the chemicals present in the water sample (Cooper et al., 2020); fiber-optic Raman dip-stick sensors that rely on the specific Raman spectra of excited BTEX molecules for calculating concentration (Stellman et al., 1998); quartz crystal microbalance sensors which selectively adsorb and subsequently measure the change in mass for BTEX quantification (Reichert et al., 2025), and surface acoustic wave sensors that detect BTEX compounds through selective interaction with coatings, inducing changes in the acoustic wave phase traveling across the sensor (Friedt et al., 2025). While these portable sensors enable field BTEX measurement, their reliance on costly materials, point-based sampling, and limited scalability hinders widespread adoption. An example is the commercially available spectrometry-based benzene probe that costs around €21,000 (ABB, n.d.).

Therefore, one of the most significant challenges in validating and optimizing MNA is the scarcity of dense, high-resolution spatial and temporal datasets. This data scarcity introduces uncertainty into site assessments, restricts regulatory confidence, and impedes the timely and effective management of contaminated groundwater sites. Thus, cost-effective and easily deployable monitoring tools are needed to enable large-scale implementation across multiple sites. That is, scalable solutions are needed that can be expanded economically without requiring extensive infrastructure or specialized expertise. Early detection involves identifying contamination soon after its onset, before migration to sensitive receptors, while continuous monitoring refers to tracking contaminant trends over time rather than relying on infrequent sampling events.

***Research Gap 1***: *There is an urgent need for scalable, real-time monitoring tools for early contamination detection and continuous monitoring of PHC in groundwater to support MNA.*

## 1.4 Sensor Data Fusion

In-situ measurements of parameters such as pH, dissolved oxygen (DO), oxidation–reduction potential (ORP), and electrical conductivity (EC) are hypothesized to serve as indirect indicators of BTEX contamination. Petroleum-contaminated groundwater typically exhibits anaerobic conditions especially in the initial years after contamination, reflected in low DO and ORP values due to oxygen consumption during biodegradation (McMahon et al., 2011; Ng et al., 2014); acidification, indicated by a decrease in pH from microbial activity and degradation byproducts (Appelo & Postma, 2005; Pandolfo et al., 2023); and elevated ionic content, which increases EC through mineral dissolution and reaction processes (Appelo and Postma, 2005). Despite these observable trends, no direct quantitative studies have established clear correlations between these parameters and BTEX concentrations.

The limitations of conventional monitoring methods have spurred interest in sensor data fusion (SDF) as a promising alternative for continuous and cost-effective water quality assessment. SDF is broadly defined as a set of techniques that integrate data from multiple sensors to produce a more informative measurement than any single source can provide (Mitchell, 2007). For example, this study used SDF to investigate whether combining readings from pH, DO, ORP, and EC sensors can indirectly detect or estimate PHCs in groundwater. The four main goals of using SDF according to Mitchell (2007) are: (1) to improve the representation of the system by deriving a more complete characterization of the measured parameter; (2) to increase accuracy by reducing random error and noise through the integration of complementary data; (3) to reduce uncertainty by increasing confidence in the resulting information; and (4) to achieve completeness by improving existing knowledge with new data or information.

SDF techniques range from classical analytics-based methods, developed in the mid-20th century, to learning-based computational methods that emerged with recent advances in data science (Koch, 2014; Hussain et al., 2024). Classical SDF techniques include Bayesian inference, which updates the probability of a hypothesis as new evidence becomes available (Zyphur and Oswald, 2015); Dempster-Shafer theory, a framework for combining evidence from different sources while accounting for varying degrees of belief and uncertainty (Shafer, 1990); and the Kalman filter (KF), a recursive algorithm for state estimation in dynamic systems (Simon, 2006). Computational SDF methods include machine learning (ML), a data-driven approach for identifying complex patterns to predict parameters of interest.

SDF has been applied in fields such as intelligent transportation systems (El Faouzi and Klein, 2016), management of smart civil infrastructures (Hakimi et al., 2023), and smart healthcare systems (Muhammad et al., 2021). In water quality monitoring, SDF has been used to enhance the monitoring accuracy and reliability of contamination warning systems by combining data from multiple sensors (Karami et al., 2012). Furthermore, it has been employed to detect contamination events in drinking water distribution systems by combining parameters like turbidity, chlorine, and flow (Gleeson et al., 2025). A data fusion model using Dempster-Shafer theory has also been designed to monitor water quality for aquaculture (Rupok et al., 2020). In addition, a KF framework was proposed to estimate uranium and tritium concentrations in groundwater in real-time using specific conductance and pH (F. Schmidt et al., 2018). However, there is a lack of continuous and real-time quantification of PHC concentrations in groundwater using data from affordable in-situ water quality sensors (e.g., pH, DO, EC, and ORP), especially under variable hydraulic and geochemical conditions typical of aquifers at contaminated sites.

***Research Gap 2*****:** *No systematic framework exists to translate affordable in-situ water quality measurements into reliable PHC estimates.*

## 1.5 Machine Learning for Water Quality Monitoring

Choosing the best SDF strategy from existing methods depends on the complexity of the problem, the characteristics of the data, and the goal of the approach. Classical SDF methods offer several advantages, including adaptability to dynamic environments, interpretability, and the capability to incorporate physical models and constraints (Liggins II et al., 2017). However, they also present limitations when applied to complex, nonlinear environmental systems such as groundwater contamination. PHC contamination of groundwater typically exhibit complex interactions between hydrochemical, biological, and geophysical processes that are often difficult to represent within an SDF framework using classical analytics-based algorithms. In such contexts, learning-based SDF methods such as ML provides a powerful alternative. ML models excel at capturing nonlinear relationships and subtle patterns in data that may be unnoticeable or unrepresentable in traditional SDF algorithms. This makes ML models especially suited for the high-frequency, multivariate datasets generated by modern sensor networks (Chen et al., 2024).

However, the generalization and transferability of ML models is strongest when sites share similar hydrogeochemical conditions; purely data-driven models can fail when applied outside the range of the training data (Baste et al., 2025). To mitigate this, hybrid approaches that integrate ML with process-based constraints are increasingly explored. For example, a common type of ML model called artificial neural network was trained on climate data from a basin and was successfully adapted to predict streamflow in a nearby ungauged basin (Besaw et al., 2010). This is particularly beneficial in groundwater monitoring, where site-specific calibration of traditional SDF algorithms can be resource-intensive and time-consuming. Besides, the computational capacity of modern systems allows ML models to process large-scale, high-dimensional data rapidly, making them suitable for real-time applications.

There are several ML models that can learn from data to make predictions or decisions without being explicitly programmed for every scenario. In water quality monitoring, common SDF applications using ML include: classification, to categorize water quality (e.g., contaminated vs. clean) based on sensor inputs; regression, to predict the concentration of a difficult-to-measure parameter from easy-to-measure parameters; and anomaly detection, to identify unusual patterns that may indicate a new contaminant release or a system malfunction. For instance, the concentration of a toxic groundwater contaminant called pentachlorophenol was predicted using linear and tree-based ML regressors (Rad et al., 2024). They found that the most significant input parameters for prediction were tetrachlorophenols, DOC and alkalinity. In addition, an ML model was developed to predict nitrate concentrations in groundwater (Arabgol et al., 2016). The prediction relied on several easily measurable groundwater quality parameters, including temperature, EC, depth to groundwater, total dissolved solids, DO, pH, land use, and seasonal changes.

The successful development and application of any SDF methodology are highly dependent on the availability of reliable, high-resolution datasets for calibration and validation. However, as illustrated in **Figure 1.2**, real-world field data from contaminated sites, such as the dataset provided by an oil company, remain spatially and temporally sparse. This data scarcity hinders the development of robust SDF models, since the full system dynamics are not adequately captured. One approach to solve the data scarcity challenge is to develop a physics-based model that simulates complex biogeochemical processes in groundwater.

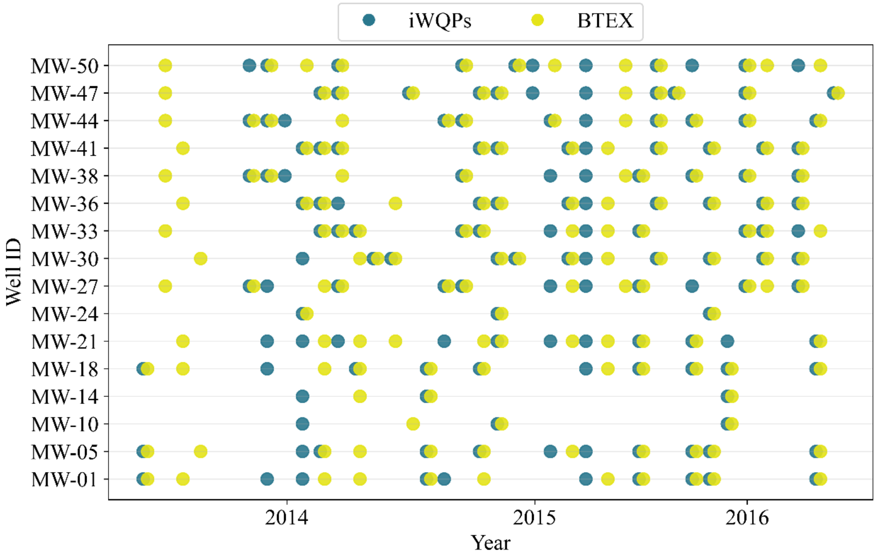


**Figure 1.2**. Available measurements of in-situ water quality parameters (iWQPs; i.e., pH, DO, EC, and ORP) and BTEX in monitoring wells from an oil contaminated site.

In principle, synthetic datasets that mimic realistic field conditions can be generated from this model. The generated datasets can then be used to explore hypothetical scenarios that are otherwise costly to test in the field and validate SDF algorithms before deploying in the field. For this purpose, a reactive transport model (RTM) can be developed to simulate the dissolution and degradation of PHC in groundwater, along with the associated physicochemical processes. Since aquifers vary significantly from one region to another in terms of geological structure, hydraulic properties, and geochemical composition, these differences strongly influence groundwater flow and contaminant transport. To capture this variability, the RTM can simulate diverse hydrogeochemical settings through scenario simulations, ensuring realistic representation of site-specific conditions.

***Research Gap 3.*** *Lack of tools for PHC-contaminated groundwater systems that can operate effectively under data-limited conditions, and limited availability of models* *capable of generating representative synthetic datasets to support analysis of PHC contamination across varying hydrogeochemical settings.*

## 1.6 Reactive Transport Modelling

Given the scarcity of high-resolution field data and the laborious, costly process required to collect it, a physics-based simulation is thus necessary to support SDF training and validation. RTMs are simulations that quantitatively couple the physical transport of fluids and solutes (advection, dispersion, diffusion) with biogeochemical reactions (e.g., microbial degradation, aqueous speciation, mineral precipitation/dissolution). For PHC contamination of groundwater, RTMs can simulate the complex fate and transport processes, including the dissolution of LNAPL, the aerobic and anaerobic biodegradation of BTEX and other organic compounds, and the resulting changes in geochemical conditions (Ng et al., 2015).

Advection, combined with mechanical dispersion and molecular diffusion, governs the migration of BTEX plume in groundwater (Picone et al., 2012; Cavelan et al., 2022). These processes are a function of the groundwater flow conditions (e.g. steady and transient), hydraulic properties of the aquifer (e.g. hydraulic conductivity, permeability, porosity, and dispersivity), fluid characteristics (e.g. viscosity, density and interfacial tension), and the distribution of contaminants in the aquifer (chemical gradient and available concentration), among others (Fetter et al., 2018; Gupta and Yadav, 2020).

The fate of BTEX compounds in groundwater is primarily influenced by terminal electron-accepting processes, such as biological respiration and fermentation. These involve electron acceptors like oxygen, $NO_3^-$, $Fe^{3+}$, $SO_4^{2-}$, and methane (McMahon et al., 2011). A range of reactions are also initiated, including homogeneous and heterogeneous chemical reactions (e.g., between acids, bases, and metals), sorption mechanisms (absorption, adsorption, ion exchange), and mineral dissolution and precipitation (Appelo and Postma, 2005). Their behavior is both influenced by and capable of affecting factors such as water quality parameters, mineral composition, background chemical concentrations, and the main electron donor present in the aquifer (Ng et al., 2015).

RTM offers a controlled environment for testing and validating the hypothesis that in-situ water quality parameters (iWQPs) can indirectly indicate BTEX contamination through SDF algorithms. Proof-of-concept studies using time-series data from virtual observation wells offer a practical foundation for assessing the feasibility of this framework. Analyzing data from virtual observation wells, rather than from every grid cell in the RTM simulation, is necessary to reflect what can realistically be measured in the field. These SDF algorithms rely on correlations between iWQPs and BTEX, which can be explored using data from the RTM. Moreover, RTM enables the simulation of diverse aquifer conditions and hydrological scenarios, supporting scenario analysis prior to field deployment.

## 1.7 Research Questions

The overall objective of the thesis is to determine and assess the potential of iWQPs for detecting and estimating PHC contamination in groundwater. This is achieved through scenario analysis using an RTM as a reference base with a focus on understanding degradation processes, evaluating relationship between PHC and iWQPs, and developing ML models to detect and estimate PHC concentration under dynamic aquifer conditions. Specifically, the following research questions were addressed in this dissertation:

1. How do iWQPs correlate with dissolved PHC concentration in groundwater, and how are these correlations influenced by spatiotemporal, hydrogeological, and geochemical characteristics of the aquifer? (Research Gap 3)
2. Can SDF-based ML models trained on iWQPs reliably provide early warning of PHC contamination in groundwater under varying aquifer conditions? (Research Gap 1 and 2)
3. Can SDF based-ML models accurately estimate toxic PHC concentration in groundwater using iWQPs, and maintain accuracy during abrupt changes (i.e., source removal, hydraulic gradient shifts)? (Research Gap 1 and 2)

## 1.8 Research Approach

Due to the lack of available field data, combined with the high cost and safety concerns associated with collecting high-resolution data from contaminated sites, a modeling approach was adopted. Specifically, an RTM was developed to serve as a reference base for generating synthetic datasets needed to train and validate SDF frameworks. To ensure the RTM is as realistic but practically simple as possible, the RTM incorporated essential processes governing groundwater flow, contaminant transport, and hydrogeochemical reactions.

Using this “virtual laboratory”, high-resolution synthetic datasets that describe the relationships between iWQPs and PHC concentration under controlled conditions were generated. This setup was used to understand the spatial and temporal variability of PHC plume and its relationship with iWQPs. Several aquifer scenarios were simulated by varying characteristics such as hydraulic conductivity, water table dynamics, and geochemical conditions to explore their influence on the iWQPs-PHC relationship.

Building on this foundation, several ML models were evaluated for two tasks: the classification of PHC contamination and the regression-based estimation of PHC concentrations. These models were tested using data from eight virtual observation wells, simulated in the reference base RTM, located at various depths and distances from the contaminant source. The overall research workflow is illustrated in **Figure 1.3**.

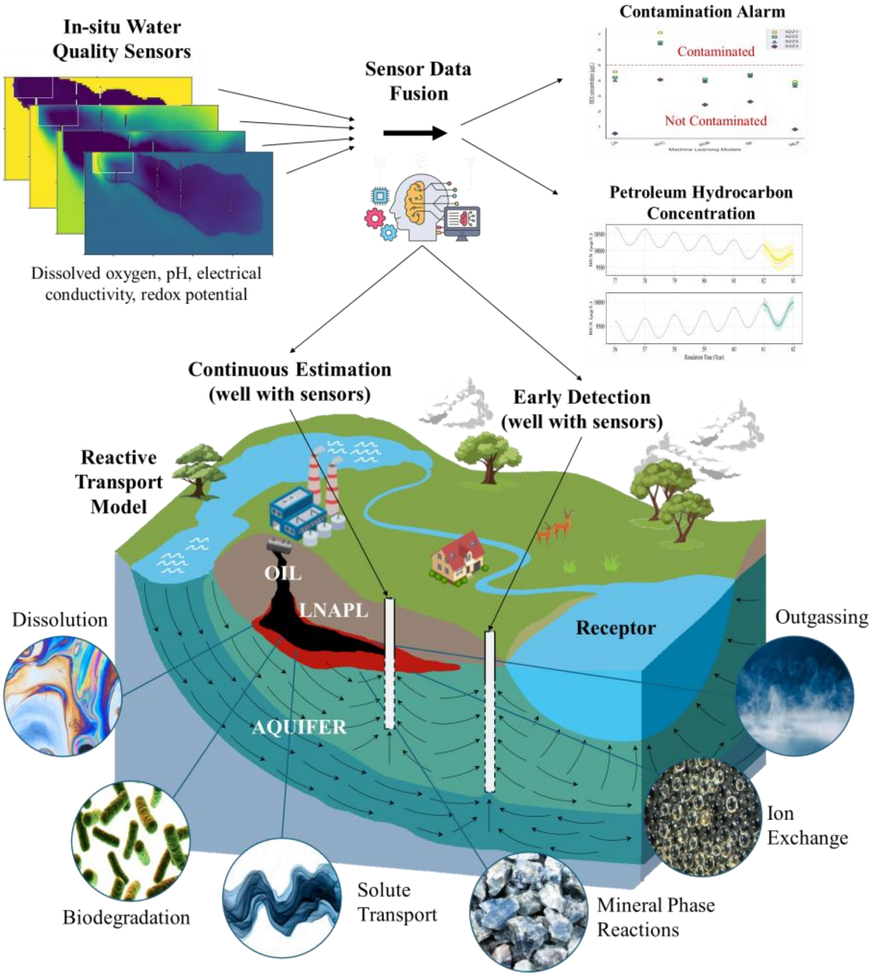


**Figure 1.3**. Schematic diagram showing the integration of iWQPs through SDF to detect and estimate PHC contamination in groundwater using synthetic data from an RTM. The iWQPs are taken from virtual observation wells in the RTM to represent observable data in the field.

## 1.9 Thesis Outline

A comprehensive synthesis of field, laboratory, and modeling studies is provided in **Chapter 2**. This aims to establish a foundational understanding of the hydrogeochemical and biological processes governing PHC contamination in groundwater, with a focus on dissolved BTEX compounds. The complex interaction of physical transport, multicomponent dissolution, biodegradation kinetics, sorption mechanisms, mineral-phase reactions, and gas dynamics, all of which influence plume behavior, is explored. By integrating insights from several studies, key processes that should be considered in reactive transport modeling are identified. This lays the groundwork for constructing realistic virtual aquifer environments for simulation and further analysis.

In **Chapter 3**, the primary source of synthetic data is introduced. An RTM was developed that simulates a stationary oil source zone continuously dissolved at the top of a heterogeneous and shallow sandy aquifer over a two-dimensional cross-section. The RTM is used to generate high-resolution synthetic datasets that illustrate the complex, spatiotemporal, and aquifer condition-dependent relationships between PHCs and iWQPs across a range of realistic aquifer scenarios, including heterogeneous hydraulic conductivity and transient water table conditions.

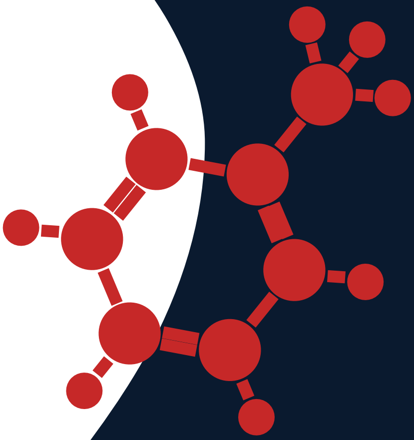
Using the virtual datasets from the developed RTM, the first component of the monitoring framework is built in **Chapter 4**: an affordable early-warning system through continuous groundwater monitoring. SDF based on classical ML classifiers are evaluated for their ability to detect the arrival of PHC contamination at monitoring wells downstream of the LNAPL source zone. The models are tested for robustness against noisy sensor data and across different aquifer conditions using the scenario simulations outlined in the previous chapter.

**Chapter 5** advances the monitoring framework from detection to quantification. The performance of various SDF based-ML regressors is assessed for accurately estimating PHC concentrations at different training-testing data windows. The focus is on a specific recurrent neural network termed as long-short term memory which handles sequential data for timeseries analysis. Furthermore, the use of a hybrid KF framework is demonstrated, showing how recursive updating can dynamically track concentration trends, particularly after a source removal event.

The thesis concludes in **Chapter 6** by synthesizing the key scientific findings and demonstrating how these findings address the research questions. Practical implications of these findings are also discussed for groundwater management. In addition, the chapter outlines future directions, including possible enhancements to RTM, steps required for validation and field-scale deployment, potential applications to other groundwater contaminants, and opportunities to advance BTEX monitoring through sensor innovation.

# Understanding the Behavior of Dissolved Petroleum Hydrocarbon Plumes in Aquifers: A Literature Review

# 2



# Abstract

Groundwater contamination by petroleum hydrocarbons, particularly benzene, toluene, ethylbenzene, and xylenes (BTEX), poses significant risks to human health and environmental safety. These compounds are commonly released from light non-aqueous phase liquids and migrate through the subsurface that form dissolved-phase plumes. The dissolution and migration of these toxic compounds are governed by complex hydrogeochemical and biological processes. In this literature review, findings from field studies, laboratory experiments, and numerical modeling were synthesized to establish a mechanistic understanding of BTEX fate and transport in groundwater. The important processes influencing plume behavior include multicomponent dissolution, sorption, biodegradation under aerobic and anaerobic conditions, mineral-phase reactions, and gas exchange. The role of aquifer heterogeneity, transient groundwater flow, and redox zonation in shaping contaminant migration were also established. Based on these insights, the essential processes for simulating BTEX behavior under realistic aquifer conditions were proposed. This serves as a foundation for constructing a reactive transport model for further analysis.

## 2.1 Introduction

The widespread use of petroleum products has caused the contamination of groundwater with hazardous organic compounds, threatening human health and ecological security upon exposure (Ololade et al., 2020; Li et al., 2021). The organic pollutants that are less dense than water such as diesel, gasoline, and other petroleum products are referred to as light non-aqueous phase liquids (LNAPLs) (Huntley and Beckett, 2002). The released hydrocarbons, e.g. from accidental leakage of underground storage tanks (**Figure 2.1**) or buried oil pipes, tend to migrate towards the water table, causing a portion of the LNAPL to be entrapped in the saturated zone (Peter, 2010; Rivett et al., 2011; Teramoto and Chang, 2020). LNAPLs in the subsurface could then generate dissolved phase LNAPL plumes in groundwater. These compounds are found to be carcinogenic and toxic for humans and animals (Veil et al., 2004; Mitra and Roy, 2011; Shores et al., 2017; Varjani et al., 2017). Thus, careful analysis of the fate and transport of LNAPL plume in groundwater is essential in managing and possibly preventing these compounds in reaching critical water resource zones.

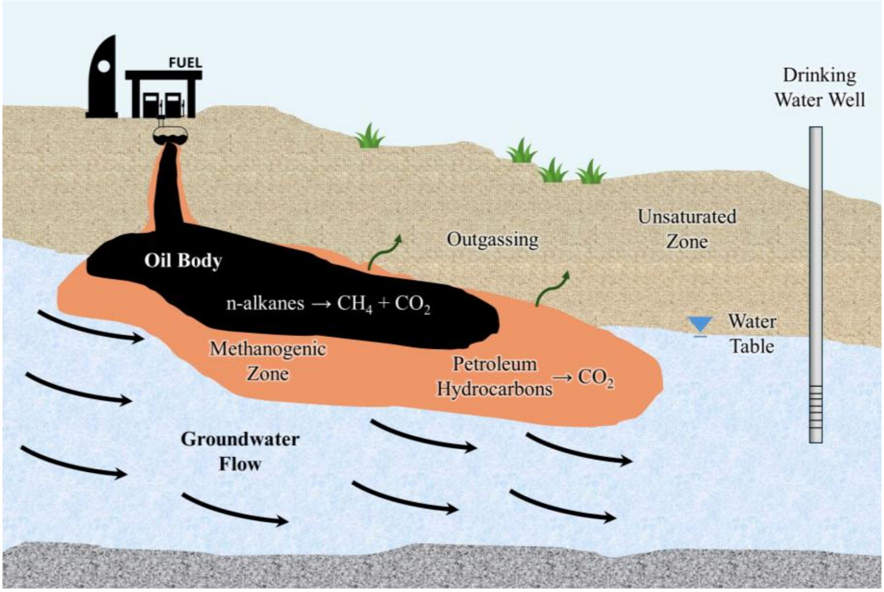


**Figure 2.1.** Major processes induced by hydrocarbon contamination in the subsurface.

The transport of LNAPL in the subsurface and its partitioning into different phases are governed by complex processes which are affected by several factors. The migration time in the vadose zone has been found to be greatly affected by: 1) the oil leakage rate and pore water saturation (Xu et al., 2015); 2) the geophysical characteristics of the media such as soil characteristics (Zanello et al., 2021) and capillary pressures (Sookhak Lari et al., 2016); and 3) the rate of water infiltration into the soil (Alazaiza et al., 2021). Furthermore, the solubility of the different LNAPL components dictates the distribution of these compounds in groundwater (Lekmine et al., 2014). The dissolution of LNAPL is influenced by groundwater flow regimes, where higher flow velocities tend to increase dissolution rates (Gupta and Yadav, 2020), as well as by water table fluctuations (Suthersan et al., 2015; Alazaiza et al., 2020), and various geophysical and biochemical properties (Essaid et al., 2003; Lekmine et al., 2017). A group of monocyclic aromatic hydrocarbons consisting of benzene, toluene, ethylbenzene, and the three forms of xylene (o-, m-, p-), termed as BTEX, are one of the most abundant LNAPLs in the environment (Astrahan, 2018). BTEX contamination in groundwater is heavily influenced by anthropogenic activities, especially in the oil refining industry (Duan and Li, 2017; Varjani and Upasani, 2017).

In this paper, the aim is to provide an overview of existing literature to understand the fate and migration of dissolved-phase BTEX plume in groundwater. The focus is on the biogeochemical and physical processes that influence the BTEX plume behavior. Insights were drawn from field studies, laboratory experiments, and numerical modelling studies. The goal is to identify the processes commonly incorporated in reactive transport models (RTMs), including microbial degradation, contaminant partitioning (e.g., dissolution and volatilization), degassing, advective-dispersive transport; mineral dissolution and precipitation; homogeneous and heterogeneous reactions with acids and bases; and sorption or surface complexation (Steefel et al., 2015; Li et al., 2017; Prommer et al., 2019).

## 2.2 Mass Transport Mechanism

Advection, combined with mechanical dispersion and molecular diffusion, governs the migration of BTEX plume in groundwater (Picone et al., 2012; Cavelan et al., 2022). These processes are a function of the groundwater flow conditions (e.g. steady and transient) and hydraulic properties of the aquifer (e.g. hydraulic conductivity, permeability, porosity, and dispersivity), among others (Fetter et al., 2018; Gupta and Yadav, 2020).

### 2.2.1. Groundwater Flow Conditions

Groundwater flow models often assume steady-state conditions because they are simpler to implement and provide acceptable approximations in many cases. For example, a modelling study at a crude oil spill site in Bemidji, USA compared steady-state and fully transient flow and found that steady-state simplifications yielded acceptable results (Essaid et al., 2003). Subsequent studies at the same site also considered a steady-state groundwater flow by imposing constant head boundary conditions and neglecting water table fluctuations and seasonal recharge variations (Molins et al., 2010; Ng et al., 2015; Lacey, 2021). These models have shown good approximation of field observations. Similar steady-state assumptions have been successfully applied to an alluvial aquifer at a contaminated oil refinery in Iran (Vaezihir, 2012), and other saturated porous systems (Valsala and Govindarajan, 2019; Verardo et al., 2021).

However, BTEX plume behavior can be influenced by transient groundwater flow conditions, which are usually caused by variable recharge, fluctuating water table, water extraction from wells, and artificial recharge (Prommer et al., 2002). These dynamics affect flow paths, travel times, and biodegradation processes due to changes in redox conditions, availability of electron donor, and nutrient infiltration (Vroblesky and Chapelle, 1994; Essaid et al., 2003). The importance of modeling transient groundwater flow to realistically simulate BTEX plume behavior under site-specific conditions has been emphasized in several studies. This is particularly relevant in dynamic aquifer environments, such as: a sand and gravel aquifer at a former gasworks site in Germany (Prommer et al., 2009; Salowsky et al., 2021); a sandy aquifer in a heavily polluted industrial site in Italy (Colombani et al., 2009); and an alluvial aquifer at a former coke and gas factory in Belgium (Batlle-Aguilar et al., 2009; Batlle-Aguilar et al., 2014).

Moreover, the contaminant flow path and plume spreading in a sand aquifer in a hydrocarbon-contaminated site in Australia have been found to be affected by the seasonally changing groundwater flow (Prommer et al., 2002). A field experiment in a sandy aquifer has revealed that the transport of BTEX in groundwater was highly influenced by the groundwater level fluctuations, and the seasonal recharge due to heavy rainfall (Lee et al., 2001). Another field experiment in a coastal aquifer in Brazil has shown that the changes in recharge caused a variation in the groundwater flow direction and velocity, ultimately leading to a greater BTEX plume spreading and migration (Rama et al., 2019). Several laboratory experiments have also been conducted to simulate the response of the BTEX plume under varying conditions. An experiment with porous media has shown that the dissolved LNAPL movements highly depend on varying groundwater flow conditions (Gupta and Yadav, 2020), and that the increase in water infiltration intensity could force the migration of LNAPL further below the saturated zone (Alazaiza et al., 2021).

### 2.2.2. Aquifer Hydraulic Properties

Aquifer heterogeneity plays an important role in groundwater flow and BTEX contaminant transport through groundwater. In the crude oil-spill site in Bemidji, USA, a heterogeneous aquifer permeability has been considered crucial for simulating the observed field conditions (Dillard et al., 1997; Ng et al., 2015; Lacey, 2021). The shape of the BTEX plume in a sandy aquifer near a petroleum oil and lubricant facility in Utah has also been found to be sensitive to the hydraulic conductivity and aquifer thickness (Lu et al., 1999). Additionally, the geological heterogeneity has been found to contribute to the enhancement of the dissolution of the contaminant source in a controlled gasohol spill site in Brazil, resulting to greater plume migration (Rama et al., 2019). Furthermore, it has been determined in a numerical model that the heterogeneity in hydraulic conductivity greatly affects the shape and size of the BTEX plume (Uçankuş and Ünlü, 2007). These studies have highlighted the effect of heterogeneous permeability and hydraulic conductivity of the aquifer to the characteristics of the BTEX plume.

BTEX contamination in fractured bedrock systems has been addressed in various studies. An analytical solution developed to simulate multicomponent dissolution and transport of non-aqueous phase liquids (NAPL) in porous media with discrete fractures revealed the influence of fracture spacing and organic carbon content in the matrix on contaminant plume behavior (Hansen and Keuper. 2014). In addition, numerical models simulating the transport of dissolved benzene (Renu and Kumar, 2016), and BTEX compounds (Valsala and Govindarajan, 2018a) in saturated fracture–matrix systems have highlighted the critical role of aquifer properties in controlling contaminant migration along fractures. The transport of hydrocarbon contaminants in peat soils and peatlands has also been characterized; the dual-porosity structure of peat and the determination of depth-specific hydraulic properties were identified as important considerations (Gharedaghloo and Price, 2017). These studies illustrate the importance of aquifer characteristics in determining the fate and migration of BTEX plumes.

## 2.3 BTEX Dissolution and Biodegradation

The fate of BTEX compounds in groundwater is governed by a range of physical and geochemical processes. Among the most significant are the terminal electron-accepting processes, including biological respiration and fermentation, involving oxygen, $NO_3^-$, $Fe^{3+}$, $SO_4^{2-}$, and $CO_2$, among others (McMahon et al., 2011). These redox reactions are often accompanied by geochemical processes, including homogeneous and heterogeneous chemical reactions (e.g., acid-base and metal interactions), sorption mechanisms (e.g., absorption, adsorption, and ion exchange), and mineral dissolution and precipitation (Appelo and Postma, 2005). These processes are both influenced by and capable of affecting environmental parameters, such as pH, dissolved oxygen (DO), and redox potential, mineral composition, background chemical concentrations, and the dominant electron donor present in the aquifer (Ng et al., 2015).

Typically, the fate of BTEX begins with the dissolution of these compounds from a source zone, followed by microbe-mediated degradation via respiration reactions. This degradation may occur either under aerobic or anaerobic conditions (Lueders, 2017). The sequence of electron acceptor utilization generally follows the order of decreasing energy yield per mole of oxidized organic carbon: DO, $NO_3^-$, $Mn^{4+}$, $Fe^{3+}$, $SO_4^{2-}$, and $CO_2$ (Appelo and Postma, 2005). The reaction products from these redox processes include $Mn^{2+}$ from manganese reduction, $Fe^{2+}$ from iron reduction, $H_2S$ from $SO_4^{2-}$ reduction, and $CH_4$ from methanogenesis. However, field and modeling studies have reported overlapping zones of different electron acceptors, indicating that these processes may not always occur in a strictly sequential manner (Brun et al., 2002; Vencelides et al, 2007; Miles et al., 2008; Teramoto et al., 2020).

### 2.3.1. BTEX Dissolution

The partitioning of BTEX constituents from the LNAPL into the aqueous phase, creating dissolved-phase BTEX plumes, is influenced by several factors. Among these, the distribution and saturation of the LNAPL, the geochemical properties of heterogeneous porous media, and the prevailing fluid flow conditions, have been found to be the governing factors (Vasudevan et al., 2014; Zanello et al., 2021). These include water table fluctuations, groundwater velocity, and temperature (Sulaymon and Gzar, 2011; Teramoto et al., 2020; Cavelan et al., 2022). The dissolution process can be modeled either as a non-equilibrium (rate-limited) kinetic reaction or as an equilibrium reaction (Molson et al., 2011). Kinetic dissolution is typically described using a non-linear dissolution rate coefficient, while equilibrium dissolution assumes ideal partitioning of BTEX into water, often based on Raoult's Law (Molson et al., 2011; Lekmine et al., 2014). Equilibrium conditions can be assumed when the dissolution rate coefficient is sufficiently high to justify a local equilibrium approximation (Lekmine 2017).

The BTEX dissolution rate is generally a function of the compound's aqueous solubility, the interfacial area between phases, and the mass transfer capacity (Njobuenwu et al., 2005; Lekmine et al., 2017). Laboratory experiments have been conducted to determine BTEX dissolution rates in porous media (Sulaymon and Gzar, 2011), while numerical modelling studies have explored dissolution under field-scale (Zanello et al., 2021) and pore-scale conditions (Gao et al., 2022).

In a controlled test in the Borden aquifer in Canada, NAPL dissolution was simulated as an equilibrium process from a residual source using effective solubility based on Raoult's law (Freitas et al., 2011). Similarly, equilibrium dissolution was found to adequately represent the residual gasoline dissolution in earlier studies (Rixey and Joshi, 2000). However, several studies have also reported non-equilibrium dissolution behavior. For instance, a conceptual model of BTEX biodegradation in a Brazilian aquifer represented dissolution as a rate-limited mass transfer process (Teramoto and Chang, 2019). The considered parameters were the compound concentration, mass transfer coefficients, and specific interface area between the phases. A dissolution rate coefficient has also been applied in inverse modeling (Essaid et al., 2003) and reactive transport modeling (Molins et al., 2010; Ng et al., 2015) at the Bemidji site, USA, where BTEX dissolution from the source zone was treated as a kinetic process. In another study at a coal tar-contaminated sandy aquifer beneath a former wood treatment plant in Germany, rate-limited kinetics have been used (D'affonseca et al., 2011). However, the mass transfer coefficient has been found to be high enough to approximate local equilibrium.

In addition to kinetic and equilibrium approaches, both single-component and multi-component dissolution models have been considered, since BTEX compounds exhibit different dissolution behaviors (Lekmine et al., 2014; Vasudevan et al., 2016). The most soluble compound in BTEX would be dissolved preferentially, resulting in differences in the spreading and transport of these compounds along the horizontal and vertical direction (Gupta and Yadav, 2020; Cavelan et al., 2022). During one study, it has been found that a single-component model produced different transport predictions compared to a multi-component NAPL dissolution model (Vasudevan et al., 2014). However, a single, uniform dissolution rate coefficient was deemed adequate in inverse modeling at the Bemidji site (Essaid et al., 2003). A subsequent study also grouped BEX compounds due to their similar tendency to migrate in dissolved form before degradation (Ng et al., 2015).

### 2.3.2. Terminal Electron Accepting Processes

The introduction of BTEX compounds into the aquifer system have been found to alter the aqueous concentrations of major ions and the physicochemical properties of groundwater. Shortly after a spill event, DO is rapidly depleted near the contamination source due to aerobic degradation of petroleum hydrocarbons (PHC), as observed at the Bemidji site (Ng et al., 2015). Aerobic BTEX degradation have been expressed using different kinetic approaches, including first order and Monod kinetics rates (Bregnard et al., 1996; Littlejohns and Daugulis, 2008; El-Naas et al., 2014). Laboratory experiments were also conducted to determine the aerobic degradation pathway of BTEX compounds (Liu et al., 2010).

Following oxygen depletion, anaerobic terminal electron-accepting processes become dominant. These include denitrification using $NO_3^-$, reduction of $Mn^{4+}$, $Fe^{3+}$, and $SO_4^{2-}$, and methanogenesis. Each process involves the oxidation of organic compounds such as BTEX, coupled with the reduction of these electron acceptors. The coupled redox reactions are important for BTEX biodegradation in anoxic environments (Appelo and Postma, 2005). Electron acceptors typically originate from dissolved species in groundwater and mineral phases within the aquifer matrix (McMahon et al., 2011).

RTMs of the Bemidji site considered Mn-oxides and Fe-oxydroxides as important electron acceptor sources. Methanogenesis has also been considered as essential for accurately simulating the field-observed pH (Essaid et al., 1995; Curtis, 2003; Ng et al., 2015). In a Brazilian aquifer, BTEX degradation has been found to favor dissimilatory iron reduction and concomitant methanogenesis due to the abundance of iron oxide-rich sediments (Teramoto and Chang, 2019). The authors also emphasized the importance of accurately representing the partial pressure of $CO_2$ ($PCO_2$) for simulating chemical speciation.

Iron reduction coupled with BTEX biodegradation has been modeled under field conditions in several studies (Prommer et al., 2002; Schreiber et al., 2004; Vencelides et al., 2007; Miles et al., 2008; Colombani et al., 2009). In a sand aquifer in Germany, natural oxidation with iron-oxyhydroxides and sulphate reduction were identified as the main degradation pathways (Wisotzky, 2000). In contrast, a contaminated sandy aquifer beneath a chemical plant in Italy have shown that sulphate reduction was the main BTEX degradation process, due to the high sulphur content of the NAPL source (Colombani et al., 2009; Mastrocicco et al., 2012). Similarly, in a chalk aquifer beneath a former retail filling station in England, sulphate reduction was dominant within 50 meters of the source zone. However, denitrification became the dominant degradation pathway further down-gradient (Spence et al., 2005).

Moreover, several studies have demonstrated that hydrocarbons can be mineralized via methanogenesis, facilitated by syntrophic microbial consortia composed of fermenters and methanogens (Lueders, 2017; Morris et al., 2013). Methanogenesis becomes a dominant degradation pathway when more energetically favorable electron acceptors are depleted. It has been identified as an important, or even dominant, process in hydrocarbon degradation in subsurface environments (Garg et al., 2017; Teramoto and Chang, 2019). These differences in dominant electron acceptors across sites highlight the aquifer-specific nature of BTEX degradation. Therefore, it is essential to carefully characterize the spatial distribution of electron acceptors in the aquifer for understanding BTEX plume behavior.

### 2.3.3. Biodegradation Kinetics

Accurately modeling BTEX degradation processes also requires a clear understanding of the underlying biodegradation kinetics (Littlejohns and Daugulis, 2008). BTEX degradation has been modeled using Monod kinetics, reversible mass action kinetics, irreversible rate-controlled reactions constrained by Gibbs free energy, and partial equilibrium approach (Curtis, 2003). First-order kinetics is commonly applied when microbial growth is negligible (Salanitro, 1993).

In contrast, the Monod kinetics is commonly used when microbial growth must be explicitly considered, allowing for a more detailed representation of biodegradation processes, especially of single toxic substrate mainly at low concentrations (Aronson et al., 1999; Trigueros et al., 2010). Monod kinetics has been applied in a modeling study of a petroleum-contaminated sandy aquifer in Monroe County, USA with microbial growth and decay. This included Haldane inhibition at high substrate concentrations (Schreiber and Bahr, 2002). Similarly, an RTM of the Bemidji site in USA incorporated multiple Monod kinetics with microbial growth to capture BTEX degradation (Essaid et al., 1995). However, subsequent inverse modeling has simplified the biodegradation reactions to first-order kinetics for computational efficiency (Essaid et al., 2003).

Another modeling study at the same site has adopted a partial equilibrium approach. This approach separates the oxidation (electron-donating) and reduction (electron-accepting) steps (Postma and Jakobsen, 1996). In this framework, the oxidation of organic carbon was modeled using first-order kinetics, while the corresponding reduction step was treated as an equilibrium reaction (Ng et al., 2015). A similar approach has been applied in a study of a contaminated sandy aquifer beneath a petrochemical plant in Italy, where the reduction step was modeled as an equilibrium reaction (Colombani et al., 2008). However, the oxidation step was modeled using a modified Monod-Haldane equation, assuming a stationary microbial population. Alternatively, a study on a coal tar-contaminated sandy aquifer beneath a former wood treatment plant in Germany applied first-order degradation kinetics under different conditions, incorporating inhibition factors (D'affonseca et al., 2011).

Moreover, laboratory experiments have shown that the various BTEX compounds exhibit different biodegradation rates, leading to sequential degradation (Goudar and Strevett, 1998). In situ microcosm experiments at the Bemidji site, USA, also revealed that toluene and o-xylene degraded first (within 12 days), followed by m-, p-xylene (around 400 days), benzene (around 700 days), and ethylbenzene (around 1600 days) (Godsy et al., 1999). Therefore, choosing the appropriate kinetic reaction type and the right kinetic coefficients depends on data availability, the composition of the microbial community, and the requirements for the model to reproduce observed field concentrations (You et al., 2018).

### 2.3.4. Inhibition and Induction Effects

The presence of BTEX compounds as a mixture in groundwater can result in competitive utilization of available electron acceptors (Barbaro et al., 1992; Bielefeldt and Stensel, 1999). This can also lead to an induction effect where the presence of one compound enhances the degradation of others. The outcome depends on the relative concentrations of individual BTEX compounds, and the specific microbial strains present in the aquifer (Wongbunmak et al., 2020).

Batch experiments have demonstrated that altering the concentration of individual BTEX components can either enhance or inhibit the overall BTEX removal rate (Jo et al., 2008). Additionally, non-interaction, competitive inhibition, and co-metabolism have been reported. The interaction mechanism depended on whether pure or mixed microbial cultures were involved, and on the concentration profiles of the BTEX compounds (Dou et al., 2008; Muneron Mello et al., 2019).

Aside from BTEX, other compounds introduced during oil spills can also affect biodegradation. For example, ethanol has been shown to inhibit BTEX degradation, contributing to plume persistence and elongation (Molson et al., 2002; Osterreicher-Cunha et al., 2009; Freitas et al., 2011; Rama et al., 2019). The biodegradation of ethanol has been found to release acetate, possibly reducing the thermodynamic viability of certain redox reactions, thereby inhibiting benzene degradation (Garg et al., 2017). The impact of various fuel alcohols on benzene plume behavior has also been modeled in a homogeneous sandy aquifer (Gomez and Alvarez, 2010).

## 2.4 Solid-phase Reaction

### 2.4.1. Mineral-phase Electron Acceptors

At neutral pH, solid-phase $Fe^{3+}$ and $Mn^{4+}$ minerals can act as natural sources of electron acceptors in subsurface environments. Iron commonly occurs in the form of oxides and oxyhydroxides such as hematite ($\alpha$-$Fe_2O_3$), goethite ($\alpha$-FeOOH), and ferrihydrite (a poorly crystalline hydrated ferric oxide). Manganese is typically found in oxide or carbonate minerals such as pyrolusite ($MnO_2$), birnessite [$(Na,Ca)Mn_7O_{14} \cdot 2.8\ H_2O$], and rhodochrosite (Richter et al., 2011). Amorphous phases of $Fe^{3+}$ and $Mn^{4+}$ may also occur as coatings on other minerals, such as smectite (Neumann et al., 2011). In addition, sulphate-bearing minerals such as gypsum and anhydrite can act as sources of $SO_4^{2-}$, another important elector acceptor (Plummer and Sprinkle, 2001; McMahon et al., 2011).

In a modelling study of the Bemidji site, pyrolusite ($MnO_2$) and amorphous ferric hydroxide [$Fe(OH_3)$] were included as solid-phase electron acceptors (Molins et al., 2010; Ng et al., 2015). Reductive dissolution of goethite was also considered in a BTEX biodegradation model for a contaminated aquifer in Brazil (Teramoto and Chang, 2020). Similarly, inverse geochemical modeling of a PHC-contaminated aquifer in the Czech Republic identified the importance of ferrihydrite and manganite dissolution in reproducing observed geochemical field values (Vencelides et al., 2007). The use of amorphous manganese and iron oxides, such as vernadite ($MnO_2$) and goethite (FeOOH), has also been tested in the laboratory as electron acceptors for BTEX biodegradation (Villatoro-Monzón et al., 2003).

### 2.4.2. Carbonate Chemistry

The background concentration of major ions in groundwater is usually controlled by the dissolution and precipitation of carbonate minerals. This is due to their natural abundance in relatively high solubility. Calcium carbonate minerals, such as calcite and dolomite, play a key role in controlling groundwater alkalinity, pH, and inorganic carbon concentrations (Maurer and Rittmann, 2004). At the Bemidji site, groundwater was modeled to be in equilibrium with calcite, while dolomite was assumed to be non-reactive (Ng et al., 2015). However, earlier studies have identified that the dissolution and precipitation of dolomite, along with other minerals such as quartz and feldspar, significantly influences the aquifer's geochemical background conditions (Bennet et al., 1993). In a contaminated aquifer in the Czech Republic, groundwater was found to be saturated with respect to calcite, but undersaturated with dolomite and rhodochrosite (Vencelides et al., 2007).

The oxidation of BTEX compounds can produce carbonic acid, which may further enhance the dissolution of carbonate minerals. This effect was observed at the Bemidji site, where elevated calcium and magnesium concentrations were attributed to increased acidity from BTEX oxidation (Bennet et al., 1993; Amos et al., 2011). A similar situation was reported in a calcareous aquifer in England, where BTEX degradation led to the production of carbonic acid and the subsequent rapid dissolution of calcium carbonate (Spence et al., 2005).

### 2.4.3. BTEX Biodegradation-induced Iron Cycle

The cycle of dissolution and precipitation of iron-bearing minerals induced by biodegradation of BTEX compounds could significantly alter the aquifer's mineralogical composition (Teramoto and Chang, 2019). During the early stages of BTEX oxidation, when oxygen and nitrate serve as electron acceptors, reductive dissolution of iron hydroxides occurs, releasing Fe(II) into the groundwater. As pH and alkalinity increase, Fe(II) may re-precipitate as secondary minerals such as siderite ($FeCO_3$) (D'Affonseca et al., 2011), rhodochrosite ($MnCO_3$) (Ng et al., 2015), or mackinawite (FeS) (Vencelides et al., 2007). Once iron hydroxides are depleted, $SO_4^{2-}$ reduction becomes the dominant terminal electron-accepting process. This leads to the production of sufide, which can react with dissolved Fe(II) to form iron sulfide minerals such as FeS or pyrite ($FeS_2$) (Prommer et al., 2009; D'Affonseca et al., 2011).

Further downgradient from the source zone, groundwater becomes re-oxygenated due to mixing with non-contaminated groundwater at the plume fringes (Bennet at al., 1993). In these areas, $Fe^{2+}$-bearing minerals including $Fe^{2+}$-oxides and $Fe^{2+}$-carbonates, can be oxidized back to $Fe^{3+}$, resulting in the precipitation of secondary iron minerals such as ferrihydrite, amorphous ferric hydroxide (Veneclides et al., 2007), or ferric oxyhydroxide (Eckert and Appelo, 2002). These phases may further transform into more crystalline forms such as goethite, completing the iron redox cycle.

## 2.5 Sorption

The mobility of contaminants and metals in groundwater is affected by sorption processes, including adsorption, absorption, and ion exchange (Appelo and Postma, 2005). The rate and extent of sorption are influenced by the surface area and charge of the solid phase, mineralogical composition, temperature, and pH.

### 2.5.1. BTEX Adsorption and Retardation

The migration of BTEX compounds in aquifers is often retarded by sorption onto mineral surfaces and organic matter (Powers et al., 2001; Noble and Morgan, 2002). Sorption is often described using linear, Langmuir, or Freundlich sorption isotherms. These models relate the sorbed concentration to the dissolved solute concentration at a given temperature and under equilibrium conditions (Appelo and Postma, 2005). Experimental studies have shown that BTEX sorption is highly dependent on the organic carbon content of the porous media, with the Freundlich isotherm often providing a good fit for describing this relationship (Zytner, 1994). Other studies have similarly emphasized the role of organic carbon fraction in controlling BTEX retardation, with individual components having different adsorption coefficients (Molson et al., 2002; Noble and Morgan, 2002).

In a numerical modelling study, the influence of different sorption isotherms and kinetic sorption rate on BTEX plume behavior has been compared (Valsala and Govindarajan, 2018b). It was found that the Langmuir isotherm tended to overestimate BTEX mobility compared to the Freundlich model. The kinetic sorption was also found to affect the total mass of dissolved BTEX but had limited impact on plume extent. Other models have assumed linear and reversible processes (Molson et al., 2002; Prommer et al., 2002; Schreiber et al., 2004). Additionally, a study on peat soils demonstrated that adsorption of benzene and toluene varied with depth and organic carbon content and was effectively modeled using linear isotherms (Gharedaghloo and Price, 2021). Organic carbon has been found to be the major factor controlling BTEX adsorption. This study also reviewed previous research on BTEX adsorption in peat using both linear and Freundlich models.

### 2.5.2. Cation-exchange

BTEX degradation can indirectly trigger sorption processes by mobilizing metal ions such as $Fe^{2+}$ and $Mn^{2+}$ through reductive dissolution of iron and manganese minerals. Similarly, $Ca^{2+}$ and $Mg^{2+}$ can be mobilized through calcite dissolution (Eckert and Appelo, 2002). Organic matter and iron oxide minerals have been shown to possess a high sorption capacity for these metal ions, particularly $Fe^{2+}$ (van Groeningen et al., 2020).

At the Bemidji site, $Fe^{2+}$ and $Mn^{2+}$ sorption was modeled using a simplified cation exchange approach; these metal ions were immobilized as $FeX_2$ and $MnX_2$, with subsequent release of $H^+$ ions (Ng et al., 2015). This approach was supported by field observations of low pH and abundant $Fe^{2+}$ concentrations bound to sediments. Previous studies have also considered adsorption onto clays and oxides as an immobilization process for $Fe^{2+}$ ions (Miles et al., 2008); linear sorption models were used to represent $Fe^{2+}$ retention in reactive transport simulations (Schreiber et al., 2002; Miles et al., 2008).

## 2.6 Dissolution and Outgassing

Gas exchange processes play an important role in the redox dynamics of BTEX-contaminated aquifers. Air entrapment due to water table fluctuations is one of the main mechanisms for introducing oxygen into an otherwise anoxic groundwater system (Williams and Oostrom, 2000). Other sources of oxygen include diffusive flux from soil air, infiltration from surface water bodies, and percolating rainwater (Kohfahl et al., 2009; Menz, 2016).

At the Bemidji site, modeling studies have shown that air entrapment can enhance BTEX degradation in shallow unconfined aquifers by supplying oxygen for aerobic methane oxidation near the water table (Amos et al., 2005; Amos et al., 2011). Isotopic analyses of a jet fuel-contaminated shallow aquifer in Brazil also revealed the coexistence of both aerobic and anaerobic zones within the source zone, attributed to oxygen release induced by water table fluctuations (Teramoto et al., 2020).

When more energetically favourable electron acceptors such as $Mn^{4+}$, $Fe^{3+}$, and $SO_4^{2-}$ are depleted, methanogenesis can become the dominant BTEX degradation pathway. Methanogenic hydrocarbon degradation begins with fermentation, where microbes convert hydrocarbons into acetate and hydrogen. Methanogens then produce methane through three main pathways: hydrogenotrophic ($CO_2$ reduction), acetotrophic (acetate cleavage), and methylotrophic, each distinguished by its carbon source and isotopic signature (Teramoto et al., 2020). Several studies have identified methanogenesis as a critical process in hydrocarbon degradation in various contaminated aquifers (Cozzarelli et al., 2001; Bekins et al., 2005; Garg et al., 2017).

To account for gas production, a degassing model was developed for the Bemidji site. This model incorporated the production and outgassing of $CH_4$ and $CO_2$ due to methanogenesis and was found to be essential for accurately simulating contaminant degradation (Amos et al., 2005). Mass transfer calculations further confirmed that $CO_2$ and $CH_4$ outgassing is a major process in anoxic groundwater (Baedecker et al., 1993). Surface efflux modeling of $CO_2$ and $CH_4$ emissions from the source zone was also included in a successive model, which was driven by biodegradation of BTEX and other organic compounds (Sihota, 2012; Ng et al., 2015).

## 2.7 Implications for Reactive Transport Modeling

In this literature review, a wide range of field, laboratory, and modeling studies were synthesized to provide a comprehensive understanding of the fate and transport of dissolved BTEX compounds in groundwater. The complexity of BTEX behavior has been shown to arise from a combination of physical, chemical, and biological processes. Important factors affecting the dynamics of BTEX plume include aquifer heterogeneity, fluid properties, sorption mechanisms, redox conditions, and the availability and distribution of electron acceptors. Interactions between BTEX degradation and geochemical processes have also been found to be important, such as carbonate dissolution and iron cycling. Gas dynamics, particularly oxygen input via water table fluctuations and outgassing of methane and carbon dioxide, further contribute to the complexity of dissolved BTEX behavior in groundwater.

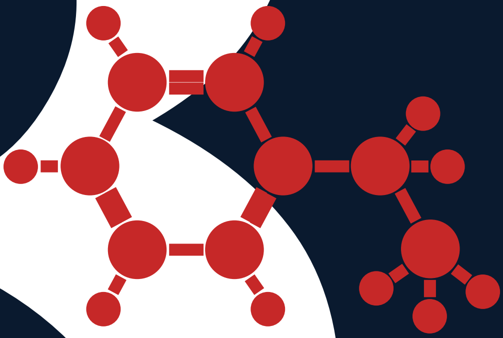
To simulate a hypothetical yet realistic contaminated aquifer, a base RTM can incorporate the following processes: 1) multicomponent dissolution of LNAPL, representing the partitioning of PHC into dissolved phases based on Raoult's Law; 2) aerobic and anaerobic biodegradation, modeled using a partial equilibrium approach, where oxidation reactions follow first-order kinetics and reduction reactions are treated as equilibrium processes; 3) mineral-phase reactions, including dissolution and precipitation of carbonate and iron-bearing minerals; 4) cation exchange, involving major cations such as $Ca^{2+}$, $Mg^{2+}$, $K^{+}$, $Na^{+}$, and $Fe^{2+}$, to simulate ion exchange dynamics in a generic aquifer matrix; and 5) gas production and outgassing, accounting for the release of $CO_2$, $CH_4$, and $N_2$ as byproducts of biodegradation and methanogenesis.

Seasonally fluctuating water table could also be considered, as it affects plume migration and contaminant dispersion. A heterogeneous hydraulic conductivity field can be used to represent the spatial variability of a porous media aquifer, which is essential for simulating realistic flow and transport conditions. For modeling simplification, the LNAPL source zone can be assumed stationary, with its composition grouped into representative fractions: volatile BTEX compounds, precursors to non-volatile dissolved organic carbon (pre-NVDOC), and long- and short-chain alkanes. A residual oil fraction that neither dissolves nor degrades may be treated as unreactive.

Numerical implementation of the RTM can be achieved using a finite-difference solution to the groundwater flow equation, coupled with reactive transport equations that account for multicomponent solubility, biodegradation kinetics, sorption, and gas exchange. Including these processes provides a starting point for simulating BTEX behavior in contaminated aquifers and can be adapted to site-specific conditions as necessary. These processes were incorporated into the RTM of the Bemidji site (Ng et al., 2015) which provided the basis for the RTM discussed in **Chapter 3**, with description of both similarities and differences.

# Reactive Transport Modeling for Exploring the Potential of Water Quality Sensors to Estimate Hydrocarbon Levels in Groundwater

## 3



## Abstract

Petroleum products have contaminated groundwater with harmful organic compounds, such as benzene, toluene, ethylbenzene, and xylenes (BTEX). Collecting and analyzing polluted groundwater samples is expensive and undertaken infrequently. However, quick remedial action in case of unexpected events requires continuous monitoring. In-situ water quality sensors (pH, EC, DO, ORP) may show correlations with the components of dissolved petroleum hydrocarbon (PHC) such as aromatics and non-volatile mobile fractions. Correlations are prerequisite to ultimately develop real-time prediction models. Since suitable field datasets are limited, the fate of hydrocarbons in groundwater under various realistic conditions was simulated using a reactive transport model as novel approach to explore when, where, and why correlations occur. A stationary oil source zone continuously dissolved at the top of a heterogeneous and shallow sandy aquifer over a two-dimensional cross-section. The RTM considered transient conditions (fluctuating water table) and spatially uniform hydrogeochemical composition. A strong correlation between PHCs and water quality sensors (rolling Spearman's correlation $> |0.8|$) was observed at varying periods. These correlations are strongly affected by the location of observation wells, the aquifer's hydraulic conductivity, and the availability of calcite and oxide minerals, and other electron acceptors. DO and ORP are significant for the early detection of hydrocarbon contamination, whereas pH and EC are important features for the long-term monitoring of hydrocarbons. These findings lay the foundation for the subsequent development of a data analysis model to detect and estimate in real time PHC levels in groundwater using in-situ water quality sensors.

This chapter has been published as:

Wu, C. L. R., Wagterveld, R. M., & van Breukelen, B. M. (2024). Reactive transport modeling for exploring the potential of water quality sensors to estimate hydrocarbon levels in groundwater. Water Resources Research, 60, e2023WR036644. https://doi.org/10.1029/2023WR036644

The repository associated with this chapter can be accessed at:

https://doi.org/10.4121/f7742f02-ee3a-4a84-adf1-625b4a9fd703

## 3.1 Introduction

The widespread use of petroleum products has caused groundwater contamination with dissolved petroleum hydrocarbons (PHC). These compounds are detrimental to both human health and the environment, especially the benzene, toluene, ethylbenzene, and xylenes (BTEX) (Li et al., 2021; Ololade et al., 2021). Leakages from underground storage tanks (Hadley et al., 2015) or burst buried oil pipes release hydrocarbons that tend to migrate towards the water table and further into the groundwater (Peter, 2010; Teramoto & Chang, 2019). Several processes are then triggered by the generated groundwater pollution plume, including degradation. These processes have been found to change the groundwater composition. Although the contamination of drinking water by PHC poses a threat to public health, groundwater is only monitored for possible pollutants from one to six times per year (Ayscough et al., 2002; Ross, 2013). This may lead to the late detection of any contamination, a delayed response time after a leakage, and an increased cost in remedial actions.

In the United States alone, there are approximately 590,000 underground storage tanks for petroleum and other hazardous substances, which could potentially leak (Wilson et al., 2014). The current practice of manually collecting groundwater samples is expensive and has associated health risks (US EPA, 2003; Wu et al., 2017; Zanello et al., 2021). As a result, a reliable, low maintenance, and low-cost technology that enables the early-detection and real-time monitoring of contaminants in groundwater is needed. Various efforts have been made to develop this kind of technology. Some studies applied sensors to estimate the level of hydrocarbons directly and rapidly in groundwater (Buerck et al., 2001; US EPA, 2003; Bender et al., 2012; Erickson & Lear, 2014). Still, these sensors are expensive, demand high maintenance, and require large construction projects.

In general, the challenges associated with PHC monitoring in groundwater originate from low-resolution data from manual sampling, and the high cost and maintenance of existing PHC sensors. Virtual sensors are among the recent technological solutions used to address these challenges. In principle, virtual sensors are developed by combining the signals received from physical sensors typically through sensor data fusion. These signals are embedded into complex functions or analytics algorithms to measure variables or conditions that might not be easily measurable physically (Kabadayi et al. 2006; Martin et al., 2021). Virtual sensors were recently applied to estimate uranium and tritium in groundwater over time using specific conductance and pH sensors (F. Schmidt et al., 2018). This real-time monitoring system was able to observe the evolution of the radionuclide plume. However, no attempts have yet been made to use virtual sensors to detect and estimate PHC concentration in groundwater.

The development of a virtual sensor involves identifying the relevant parameters and establishing their relationships. These parameters are then combined, typically by using analytics algorithms, to measure the variable of concern (Martin et al., 2021). The current paper is solely focused on the initial stage of virtual sensor development. The goal of this paper, therefore, is to explore the potential of water quality sensors, such as pH, dissolved oxygen (DO), redox level (ORP), and electrical conductivity (EC), to predict and monitor the level of PHC in groundwater.

Since the available data on groundwater contaminated sites are insufficient for a robust statistical analysis, a reactive transport model (RTM) to simulate the fate and transport of PHC compounds in the aquifer can be developed as data source. RTMs have shown that hydrocarbons in groundwater can trigger geochemical and biological reactions. Previous RTM studies have simulated the migration of PHC plumes in groundwater by combining the effects of advection, mechanical dispersion, and molecular diffusion (Picone, 2012; Cavelan et al., 2022), equilibrium and kinetic dissolution of PHC compounds (Molson and Eng, 2011; Lekmine et al., 2014), aerobic and anaerobic PHC biodegradation (Prommer et al., 2002; Schreiber et al., 2004; Vencelides et al., 2007; Miles et al., 2008; Colombani et al., 2009; Ng et al., 2015), dissolution and precipitation of carbonate minerals (Maurer & Rittmann, 2004; Spence et al., 2005), sorption of dissolved PHC to organic matter in sediments (Valsala & Govindarajan, 2018b; Gharedaghloo and Price, 2021), and (de)sorption of mobilized ions (Miles et al., 2008; Cozzarelli et al., 2016; Ziegler et al., 2017).

These processes have been found to change the physical and chemical properties of the aquifer including pH, DO, and ORP (Appelo and Postma, 2005; Ng et al., 2014). Thus, the RTM can provide mechanistic insights into the parameter correlations and help elucidate the effect of site-specific factors on these correlations. Furthermore, the RTM has no limitations on virtual datasets that can be produced.

Therefore, a generic RTM that simulated the fate and transport of hydrocarbons in groundwater and the associated biogeochemical processes under presentative conditions was developed. Then, the correlation between in-situ water quality parameters (iWQP) and PHC concentration was quantified, specifically BEX and the non-volatile dissolved organic carbon, NVDOC). Finally, the effect of different factors in terms of spatiotemporal, hydrogeological, and geochemical properties of the aquifer to the identified correlations was assessed.

## 3.2 Materials and Methods

An RTM simulating groundwater flow and reactions under hypothetical but realistic aquifer conditions were developed to provide temporal and spatial datasets needed for correlation analysis under various field conditions. The model domain and parameters were based on existing studies that examine PHC contamination. This study was specifically inspired by existing studies on the crude oil spill site in Bemidji, Minnesota, which is one of the most studied oil contaminated sites in the world (Ng et al., 2015).

### 3.2.1. Base model development

Our reactive transport model was set-up in Jupyter notebook using FloPy, an open-source Python-based modelling tool for creating, running, and post-processing groundwater models (Bakker et al., 2016). MODFLOW 2005 (Harbaugh, 2005) was used to solve the groundwater flow equation utilizing a finite-difference numerical solution, and MT3DMS (Zheng et al., 2012) to simulate the contaminant transport in the aquifer by activating the advection and dispersion packages. PHT3D (Prommer and Post, 2010) was then used to manage the coupling of flow and transport with geochemical reactions from PHREEQC-2 (Parkhurst and Appelo, 2000). Multicomponent dissolution of LNAPL (Molins et al., 2010; Ng et al., 2015), first-order aerobic and anaerobic biodegradation (Salanitro, 1993, D'Affonseca et al., 2011, Ng et al., 2015), mineral-phase reactions (Molins et al., 2010; Ng et al., 2015, Teramoto and Chang, 2019), cation exchange (Ng et al., 2015), and outgassing of methane, CO2, and N2 (Amos et al., 2005; Ng et al., 2015) were included. A fluctuating water table as a transient driver and a heterogeneous hydraulic conductivity field were also considered in the RTM.

The synthetic aquifer represented a saturated porous media that is 260 m in length and 7 m in thickness, as used by Ng et al. (2015) to simulate the Bemidji aquifer. An aerobic aquifer with 0.25 mmol/L of $O_2$, 0.36 mmol/L of $NO_3^-$, and 0.52 mmol/L of $SO_4^{2-}$ was considered. The pH value is 7.7 and the background EC is 339 µS/cm representing a freshwater aquifer. To prevent any boundary effects on the simulations, the computational grid was extended to a total length of 300 m and a total thickness of 10 m. The model domain has a discretization of 150 grid cells in the horizontal (x) direction and 50 grid cells in the vertical (z) direction (grid size of 2 m by 0.2 m). This is comparable with the previous Bemidji model with a grid size of 4.3 m by 0.47 m (Ng et al., 2015). For this model, a mean homogeneous porosity value of 0.38 (Dillard et al., 1997) and a longitudinal and transverse dispersivity of 1 m and 0.04 m, respectively (Ng et al., 2015) were used. For the aquifer storage coefficient, 0.01 was chosen, while 0.20 for the storage yield, both within the values for an unconfined aquifer (Lohman, 1972). **Table A1** summarizes the parameters used for the flow and physical transport model.

### 3.2.2. Heterogeneous hydraulic conductivity

A heterogeneous sandy aquifer was represented using a heterogeneous hydraulic conductivity (HK) field (**Figure 3.1**) generated using the field generator package of PMWIN (Chiang & Kinzelbach, 2003). The HK field has a mean value of 6.2 m/d, resembling the findings (homogeneous value of 6.1 m/d) of an inverse modelling study which was previously conducted at the Bemidji site (Essaid et al., 2003). Using 0.3 as the common log standard deviation value, the minimum and maximum HK values were 0.53 m/day and 51.74 m/day, respectively. This represents an aquifer consisting of fine and coarse sand according to the typical HK values of geological units (Domenico & Schwartz, 1998). Furthermore, a correlation length/field width along rows and columns of 0.1 was assumed. The PMWIN field generator is based on Mejia's algorithm and was used in different studies (Zammouri and Ribeiro, 2017; Thouement & Van Breukelen, 2020). Based on the commonly reported vertical-to-horizontal anisotropy ratio, 10% of the saturated HK was used as the vertical HK (e.g., 0.61 m/d for the mean HK value) (Todd, 1980).

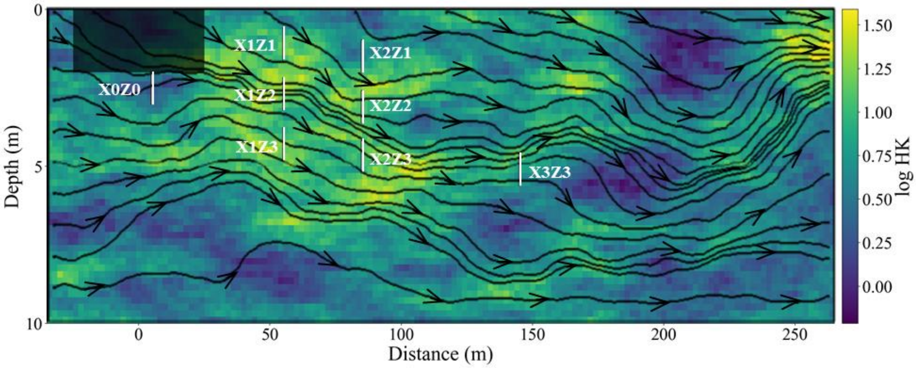


**Figure 3.1.** Logarithmic hydraulic conductivity field (log of m/day) and the steady-state flow paths with the oil body location (shaded area) and observation wells. Well names are based on the distance from the oil body (X1 nearest, X3 farthest), and the depth (Z1 shallowest, Z3 deepest). X0Z0 is located below the oil body.

### 3.2.3. Simulation of transient flow conditions

Seasonally changing groundwater flow can affect the contaminant flow directions and plume spreading (Prommer et al., 2002). Although this was not considered in the previous modelling study conducted at Bemidji (Ng et al., 2015), the present study aimed to simulate the periodic changes in the water table. Generally, fluctuations could be caused by factors including seasonal variability in recharge, tidal actions, or dam operations, among others (Bakker and Post, 2022). Additionally, extreme events such as increased precipitation due to climate change can affect the water table by influencing groundwater recharge. Although previous studies have explored how climate change affects contaminant transport (Libera et al., 2019; Xu et al., 2022), the focus of this study did not extend to simulating the potential impacts of climate change.

To simulate the transient groundwater flow, a steady-state flow field was established at the beginning. An average water table gradient of 0.0035 m/m was simulated for day 1, similar to the previous studies conducted at the Bemidji site (Essaid et al., 2003; Ng et al., 2015). This was achieved by imposing a constant head boundary condition at the left and right model boundaries, and a uniform recharge flux of 178 mm/year (Ng et al., 2015) at the upper layer of the model domain. The model bottom was considered as a no flow boundary, assuming that an impermeable clay layer is present.

Transient groundwater flow was then simulated by imposing a general head boundary (GHB) condition with a monthly changing head at the left model boundary (Anderson et al., 2015). The GHB condition represented a monthly water table fluctuation (WF) caused by open water in direct contact with the aquifer at the upstream part of the model. A commonly observed pattern in groundwater level variation is a cyclic water-level fluctuation (Healy & Cook, 2002; Park & Parker, 2008; Neto et al., 2016; Jiang et al., 2017). Thus, a simple sinusoidal function (Bakker 2019; Wu et al., 2021) with an amplitude of 0.25 and a period of 12 months was used to approximate the monthly variations in groundwater level at the left boundary. This amplitude corresponds to the maximum observed change in water level (0.5 m) at the Bemidji site (Essaid et al., 2003). The WF at the right boundary, however, was based on a larger model with a length of 5,200 m. Detailed description of this larger model can be found in **Appendix A** (**Text A1**, **Figure A1**). In this larger model, a fluctuating water table on the left boundary and a constant head on the right boundary were imposed. The water table response in an observation well 300 m from the left boundary was then recorded and used as the right boundary condition in the final model (**Figure A2**). Note that at the right boundary, the amplitude was dampened, and the peak arrived later.

Moreover, since the change in the water table was much smaller than the considered saturated thickness, a constant transmissivity for the unconfined synthetic aquifer was assumed (Bakker and Post, 2022). The total simulation time was 100 years, based on studies showing that contamination of petroleum hydrocarbons in aquifers can last for decades to over a century (Kulkarni et al., 2020; Pishgar et al., 2022). There were 1201 stress periods with 30 days per stress period (monthly stress period for 100 years plus 1 day of steady state at the beginning to establish a steady state flow field). Daily timesteps with sufficiently small grid cells was used to prevent any convergence issues caused by numerical oscillations and dispersions.

### 3.2.4. Initial hydrogeochemical conditions

For the initial background concentration of ions and physical properties of the aquifer, the values from the Bemidji site studies (Bennet et al., 1993, Ng et al., 2015) was adopted. The model does not include the Mn element for simplification since $Mn^{2+}$ was assumed to behave similarly to $Fe^{2+}$ because of similar geochemical properties in aquatic environments (Giblin, 2009; Rusydi et al., 2021). The recharge composition was similar with the native groundwater. Using PHREEQC, the background concentrations were first charge balanced and equilibrated with calcite, resulting in the parameter values shown in **Table A2**.

The included dissolvable calcite in the model is 131.6 mmol/kg because of its contribution to the aquifer's pH-buffering and carbonate chemistry, with the saturation index set to 0.05 (Ng et al., 2015). Also included is 7.6 mmol/kg of the mineral phase electron acceptor amorphous ferric hydroxide ($Fe(OH)_3$) that exists at the Bemidji site (Ng et al., 2015). The electron donor siderite ($FeCO_3$) was configured to precipitate in response to increasing dissolved ferrous iron levels because of iron-oxide reduction. Pyrite was also included in the RTM in this research but not in Ng et al. (2015) because $SO_4^{2-}$ was not detected at the Bemidji site, USA. Pyrite was allowed to precipitate as a sink for $Fe^{2+}$ and $H_2S$. **Table A3** lists the minerals included in the model, along with the Log K values.

### 3.2.5. Fate and transport component

#### Simulation of LNAPL dissolution

The contamination of the aquifer by PHCs starts with the dissolution of the light non-aqueous phase liquid (LNAPL), referred to in this study as the source zone. Dimension of the source zone varies at different sites, ranging from less than 5 m in length and less than 1 m in thickness (Freitas et al., 2011; Xu et al., 2015; Prommer et al., 2009; Gomez & Alvarez, 2010) to more than 50 m in length (D'Affonseca et al., Miles, 2008) and more than 1 m in thickness (Suarez & Rifai, 2004; Ng et al., 2015). In this study, the dimensions of the oil body were assumed to be 50 m in length and 2 m in thickness situated at the uppermost layer and 10 meters downstream from the left boundary. The location of the left boundary, x is -35 m with reference to the center of the source zone (**Figure 3.1**). Following the Bemidji studies (Essaid et al., 2003; Ng et al., 2015), the slow movement of the oil phase were ignored, and the source zone was assumed to be stationary for modelling simplification (Dillard et al., 1997).

Oil saturation can vary depending on factors such as the spill size, the LNAPL composition and thickness, and the aquifer characteristics (Lenhard et al., 2004; Lenhard et al., 2018; Mao et al., 2019). Previous reports have shown that LNAPL saturations may range from about 2% to 60% of the pore volume (Huntley et al., 1994; Beckett and Lundegard, 1997). A homogeneous oil saturation of 15% was thus assumed, comparable to the oil saturation values at the Bemidji area (Ng et al., 2014; 2015).

At the beginning of simulations, the LNAPL was partitioned into 4.3% BEX ($C_6H_6$) for combined benzene, ethylbenzene, and xylene; 1.3% toluene ($C_7H_8$); 39.2% preNVDOC ($C_{19}H_{24}O_6$) for "oil phase precursors to non-volatile dissolved organic carbon (NVDOC)" (Thorn and Aiken, 1998); 16.1% long-chain n-alkanes ($C_{15}H_{32}$); and 16% short-chain n-alkanes ($C_{11}H_{25}$). The remaining oil pool (23.1%), which does not dissolve nor degrade from the LNAPL, was considered unreactive (Baedecker et al., 2011; Ng et al., 2015). The partitioning and representative stoichiometries in the parentheses were adopted from Ng et al. (2015). These include short-chain n-alkane compounds representing all $C_6$-$C_{12}$ n-alkanes, and long-chain n-alkane compounds representing all C13-C32 n-alkanes. BEX compounds were grouped based on their similarities in migration and degradation (for more details, see Ng et al., 2014; 2015).

The multicomponent solubility of organic carbons following Raoult's law was also adopted from previous modelling studies (Ng, et al., 2015; Essaid et al., 2003). This includes the simplification that toluene does not dissolve in groundwater but degrades directly from the source zone. The multicomponent dissolution kinetics from a stationary oil body adequately represented the observations at the Bemidji site (Essaid et al., 2003; Ng, et al., 2015).

### Hydrocarbon degradation

Aerobic and anaerobic biodegradation of PHCs were simulated using the partial equilibrium approach (Jakobsen and Postma, 1999). This was derived from the modelling study previously conducted at the Bemidji site (Ng et al., 2015). In general, this approach assumes that the kinetically controlled oxidation of natural organic carbon ($CH_2O$):

$$CH_2O + 2\,H_2O \rightarrow HCO_3^- + 5\,H^+ + 4\,e^-, \tag{3.1}$$

is the rate-limiting step, while the reduction of the various terminal electron acceptors:

$$O_2 + 4\,H^+ + 4\,e^- \leftrightharpoons 2\,H_2O, \tag{3.2}$$

$$2\,NO_3^- + 12\,H^+ + 10\,e^- + \leftrightharpoons N_2 + 6\,H_2O, \tag{3.3}$$

$$Fe(OH)_3 + 3\ H^+ + e^- \leftrightharpoons Fe^{2+} + 3\ H_2O, \tag{3.4}$$

$$SO_4^{2-} + 9\ H^+ + 8\ e^- \leftrightharpoons HS^- + 4\ H_2O, \tag{3.5}$$

$$HCO_3^- + 9\ H^+ + 8\ e^- \leftrightharpoons CH_4 + 3\ H_2O, \tag{3.6}$$

is in equilibrium, thus governed by Gibbs free energy. The first-order biodegradation kinetics used by Ng et al. (2015) was also used, which was temporally and spatially uniform within the aquifer, and does not depend on the redox condition (Essaid et al., 2003). Although more complex biodegradation kinetics were previously used, such as varying degradation rate constants depending on redox conditions (Höyng et al., 2015) and Monod kinetics (Aronson et al., 1999; Essaid et al., 1995; Schreiber and Bahr, 2002), the first-order degradation model was found to be adequate in simulating the field conditions at the Bemidji site (Ng et al., 2015). Individual degradation parameters for separate compounds that comprised the oil mixture was implemented, as a previous simulation study considered this essential (D'Affonseca et al., 2011). Ng et al. (2014; 2015) provide more details on the organic carbon degradation pathways and the rate constants used.

### Cation exchange and sorption

Like the previous model at the Bemidji site (Ng et al., 2015), basic cation exchange was simulated. However, instead of including only $Fe^{2+}$, $Mn^{2+}$, and H+ in ion exchange (Ng et al., 2015), the four major cations $Ca^{2+}$, $Mg^{2+}$, $K^+$, $Na^+$ (Appelo and Postma, 2005) were also considered as exchangeable cations aside from $Fe^{2+}$ to simulate a more generic aquifer. This is also in line with other RTM studies (van Breukelen et al., 2004). Based on Ng et al. (2015), a cation exchange capacity (CEC) of 10.94 mol/kg was also assumed for the aquifer. Then, the distribution of exchangeable cations was computed by equilibrating the cation exchanger with the initial groundwater composition using PHREEQC-2. Assuming that the fraction of organic carbon in the aquifer is low, the retardation of PHCs due to sorption on sedimentary organic matter was neglected.

### Degassing from solution

Degassing of $CO_2$, $CH_4$, and $N_2$ were simulated in this model to represent the carbon loss through outgassing. Following the implementation of Ng et al. (2015), outgassing was simulated in two parts: direct outgassing of $CO_2$ and $CH_4$ from the source zone, mainly from poorly soluble short and long-chain n-alkanes, and outgassing of gas species from aqueous solution, including the ones produced from biodegradation of BEX, toluene, and NVDOC, once the sum of partial pressures of $CO_2$, $CH_4$, and $N_2$ reaches the average hydrostatic pressure (Van Breukelen et al., 2004; Amos et al., 2005). A depth gradient pressure threshold corresponding to the yearly annual average hydrostatic pressure was applied for each depth in the aquifer.

For simplicity, the pressure increased from 1 atm at the water table to about 1.97 atm at 10 m depth, disregarding the effect of WF. Each gas species was allowed to outgas at a rate proportional to the fraction of partial pressure to the total pressure. A detailed description of the outgassing mechanism can be found in Amos et al. (2005), and Ng et al. (2015).

### 3.2.6. Scenario simulations

Contaminated sites have unique hydrogeologic conditions, hydraulic properties, and geochemical characteristics. Hence, various scenarios were therefore simulated to test the effect of aquifer conditions on the correlation between the concentration of PHCs and the iWQPs. First, the amplitude of the WF in the general head boundary condition was changed from 0.25 to 0.50 (WF+) and 0.75 (WF++) to consider the effects of this transient driver. Then, the mean HK was changed from 6.2 to 3.1 m/day (HK-) and 9.3 m/day (HK+) to consider the uncertainties in aquifer heterogeneities. The same heterogeneous field was used but the values were divided or multiplied to decrease or increase the mean HK. This kept the HK values within the reported range of HK for fine and medium sand (Domenico and Schwartz, 1990). The initial background concentration of ions was also varied from fresh groundwater (EC of 339 µS/cm) to water with specific conductance of about 1078 µS/cm (EC+) and 2039 µS/cm (EC++) to determine the effect of background salinity on the correlation coefficient. This was based on reports regarding released hydrocarbons in coastal aquifers (Robinson et al., 2009; Geng et al., 2017). The ion compositions were determined by mixing a percentage of seawater (Nordstrom et al., 1979) with freshwater (initial concentration) using PHREEQC to reach the desired specific conductance.

PHC contaminated aquifers have varying dominant electron acceptors as reported for several sites (Eckert and Appelo, 2002; Spence et al., 2005; Colombani et al., 2009; Mastrocicco et al., 2012). Hence, the dominant electron acceptor was varied by removing the dissolved oxygen (DO0) from the initial level of 7.87 mg/L and removing the nitrate ($NO_30$) and sulfate ($SO_40$) from the initial levels of 5 mg/L and 50 mg/L, respectively. The nitrate level was also increased to 50 mg/L ($NO_3$+) (Fraters et al., 2008; European Union, 2020), and the sulfate level to 250 mg/L ($SO_4$+) (European Union, 2020).

Carbonate minerals usually control the background concentration of ions in an aquifer because of their abundance in nature and high solubility in groundwater. The dissolution or precipitation of calcium carbonates in an aquifer matrix could affect the alkalinity and pH of groundwater, as well as the inorganic carbon concentration (Maurer and Rittmann, 2004). Hence, the absence of calcite (Cal0), reduced $Fe(OH)_3$ from 7.59 mmol/kg to 0.76 mmol/kg and increased (CEC+) and decreased (CEC-)

cation exchange capacities from 10.94 mol/kg to 2.19 mol/kg and 54.70 mol/kg were also considered. This is to evaluate the effect of minerals and the exchangeable cations on the correlation coefficients. The simulated scenarios are listed in **Table A4**.

### 3.2.7. Virtual sampling of groundwater

The concentration time series of simulated iWQPs were obtained from eight virtual observation wells with screens at multiple depths (**Figure 3.1**). These wells, located at different depths and distances from the source zone, consisted of eight vertical grid cells from the top layer (1 m total length) and one horizontal grid cell (2 m diameter, minimum grid cell size).

The concentration at each virtual groundwater sample, $C_{\text{sample}}$, was calculated as the average of pumping flow-rate-weighted mean concentration at all the well grid cells to represent the well screen surroundings (Höyng et al., 2015). Assuming that the pumping flow rate is proportional to the hydraulic conductivity (Thouement & Van Breukelen, 2020), $C_{\text{sample}}$ was computed as:

$$C_{\text{sample}} = \frac{\sum C_i K_i}{\sum K_i}, \tag{3.7}$$

where $C_i$ is the species concentration and $K_i$ is the hydraulic conductivity of the grid cell $i$. This method considered the influence of hydraulic conductivity on the groundwater flux towards the well screen during pumping (Thouement and Van Breukelen, 2020).

### 3.2.8. Statistical data analysis

The relationships between iWQPs (DO, EC, ORP, and pH) and the PHCs (i.e., BEX and NVDOC) were analyzed. PhreeqPython, a Python implementation of PHREEQC (Vitens, 2021), was used to calculate the electrical conductivity based on the chemical water composition during post-processing. Histograms with kernel density estimation (KDE) and probability density function (PDF) plots were then created for a descriptive overview of the distribution and frequency of the parameters.

Using Spearman’s rank-order correlation, the association between the iWQPs and the PHCs at all observation wells was quantified. Spearman's correlation is a non-parametric measure that assesses the monotonic relationship between two variables (Sheskin, 2020). A “moving window” analysis (Worrall et al., 2003), termed here as rolling Spearman’s correlation analysis with a 5-year window, was then conducted to capture temporal dynamics. Repeated calculations of correlations allow for the identification of time-dependent patterns and shifts in correlation coefficients. The 5-year window was primarily chosen for illustrating the temporal dynamics inherent to the system. Finally, the effect of different simulated scenarios on the rolling

Spearman's correlation was assessed by visualizing correlation coefficients between iWQPs and PHCs under various scenarios using heatmaps.

## 3.3 Results and Discussion

### 3.3.1. Hydrogeochemistry of the groundwater pollution plume

The leakage of petroleum hydrocarbons in the aquifer influenced the concentrations of aqueous phase constituents and solid phase minerals over time and space. The dissolution of petroleum hydrocarbons from a source zone was simulated, which triggered various terminal electron-accepting processes (TEAPs). The total simulation period is 100 years, but the cross-sectional plumes was presented after 30 years to approximately match the observation period at the Bemidji site (Ng et al., 2015). The plumes at three distinct time steps (5, 10, and 30 years) are presented in **Figure A3** to **A5** to visualize the plume progression.

#### Aqueous phase simulation results

**Figure 3.2** shows the cross-sectional plume simulations of major aqueous electron donors (BEX and NVDOC), a tracer representing BEX in the case of no degradation, dissolved inorganic carbon (DIC) produced by all TEAPs, and groundwater pH. The plumes of various electron acceptors and the redox zones development due to hydrocarbon degradation are also illustrated in **Figure 3.2**, along with the solid phase minerals and the cation exchangers (X represents a sorption site). In this figure, pH, DO, and ORP are the parameters measurable with water quality sensors. The location of the observation wells is displayed as vertical white lines, and the source zone is delineated by white rectangles. The model effectively simulated hydrocarbon dissolution, advection, and dispersion, as evident from BEX and NVDOC plume spreading. The results align with the previous study on the Bemidji site (Ng et al., 2015), although it has differences such as plume extent and concentration. Several factors contribute to the variability between the simulation results and those of the Bemidji plume such as differences in initial oil saturation, mineral composition, hydraulic conductivity field, transient groundwater flow, and the presence of $NO_3^-$ and $SO_4^{2-}$. Note that the purpose is to simulate a general hydrocarbon pollution plume inspired by the Bemidji case, but not to represent the exact situation as it was at the Bemidji site.

The findings of a report on a groundwater-contaminated site at a former gasworks plant in Düsseldorf, Germany (Eckert et al., 2000) demonstrate the feasibility of the simulated BEX concentration of more than 0.2 mmol/L (15.6 mg/L) and plume length of more than 100 m. The study by Eckert et al. (2000) reported PHC concentrations of up to 100 mg/L, with the plume extending to 600 meters over a period of more than 50 years of natural attenuation. Comparing the extent of the BEX plume with that of the much larger tracer (BEX without decay) plume demonstrates hydrocarbon degradation. The degradation of all PHCs led to significant DIC and pH plumes stretching far beyond the hydrocarbon plumes (Ng et al., 2014; 2015). Note that the NVDOC levels increased considerably in the source zone over 30 years, whereas BEX levels remained largely similar.

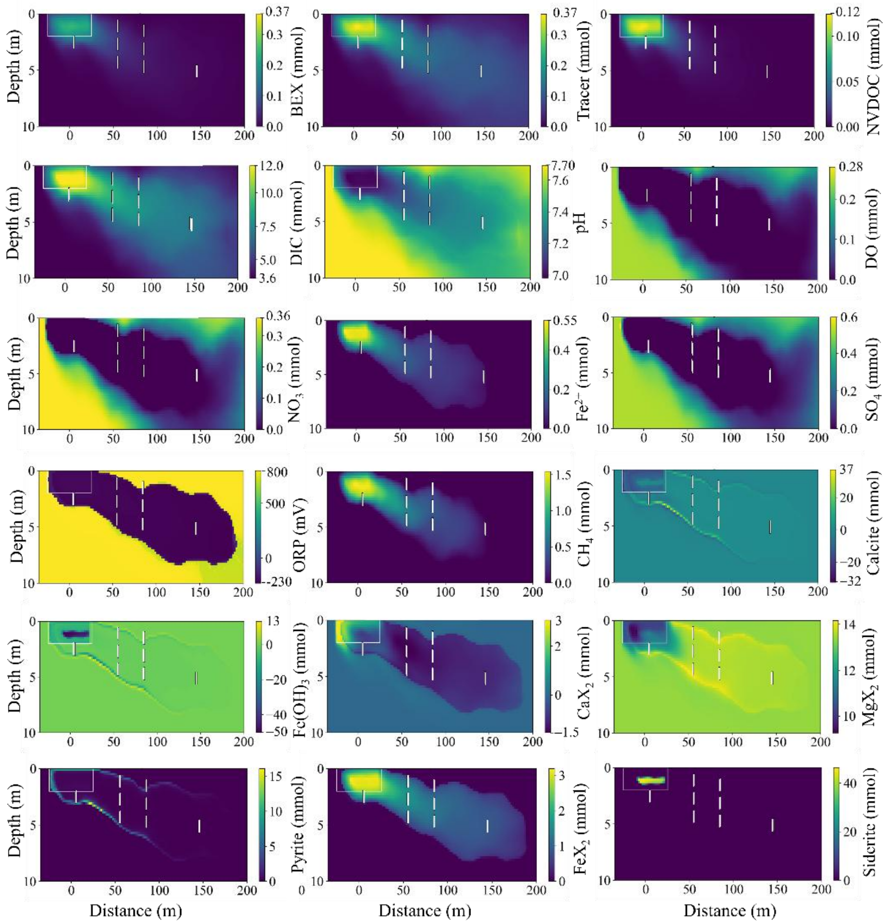


**Figure 3.2.** Model simulation results for major electron donors (BEX and NVDOC), the tracer representing BEX without degradation, DIC produced by all TEAPs, groundwater pH, depleted electron acceptors with redox zones development (ORP plume), solid phase minerals, and cation exchangers, after 30 simulation years. The solid phase plumes are depicted as the difference between the final and initial values. White rectangles delineate the source zone location.

The hydraulic conductivity of the aquifer significantly affects the characteristics of the hydrocarbon plume. A region of relatively high hydraulic conductivity (lighter color in **Figure 3.1**) caused the plumes to initially migrate downward and then almost horizontal. Similarly, the shape of the PHC plume in a sandy aquifer near a petroleum oil and lubricant facility in Utah was also found to be sensitive to the hydraulic conductivity and aquifer thickness (Lu et al., 1999). Additionally, the geological heterogeneity enhanced the contaminant source dissolution in a controlled gasohol spill site in Brazil, resulting in greater plume migration (Rama et al., 2019). Furthermore, it was determined in a numerical model that the heterogeneity in hydraulic conductivity greatly affects the shape and size of the PHC plume (Uçankuş & Ünlü, 2008).

In aquifers, the release of carbonic acids from hydrocarbon degradation tends to decrease the pH of groundwater (Appelo and Postma, 2005; Pandolfo et al., 2023). This was simulated in the model by the oxidation of organic carbon through the partial equilibrium approach, where $H^+$ is produced in the TEAP. Examples are the aerobic degradation of benzene and NVDOC, respectively:

$$C_6H_6 + 3\ H_2O + 7.5\ O_2 \rightarrow 6\ HCO_3^- + 6\ H^+, \tag{3.8}$$

$$C_{19}H_{24}O_6 + 7\ H_2O + 22\ O_2 \rightarrow 19\ HCO_3^- + 19\ H^+, \tag{3.9}$$

which produce $H^+$ and $HCO_3^-$ resulting in lower pH and higher DIC concentration, as observed in **Figure 3.2**. The possibility to detect PHC using pH and EC because of the increase in ions is based on this principle. Additionally, a fraction of the produced $CO_2$ from n-alkane degradation dissolves, providing an additional source of $H^+$. The simulation captures a low pH plume originating from the core of the source zone, reflecting the early and expanding low pH plume similarly observed at the Bemidji site (Ng et al., 2015). However, the previous study reported lower simulated pH values (below 7) due to an additional source of $H^+$ following their cation-exchange model.

The depletion of DO, $NO_3$, and $SO_4$, along with the release of $Fe^{2+}$, also shown in **Figure 3.2** demonstrates the conversion of organic carbon into inorganic carbon through kinetically controlled biodegradation. This process progressively consumes the oxidation capacity of both aqueous and mineral electron acceptors (Prommer and Post, 2010). Near the contamination source, DO was depleted shortly after a spill incident caused by the aerobic degradation of organic matter at the Bemidji site (Ng et al., 2015). This shows that DO sensors can be used to detect the presence of PHC. Note that the $SO_4$ depletion zones are smaller than those of $O_2$ and $NO_3$ because $SO_4$ is a less preferred electron-acceptor.

Previous studies also proved that PHC degradation in contaminated aquifers has been coupled with $NO_3$ and $SO_4$ reduction (Eckert and Appelo, 2002; Colombani et al., 2009), and iron reduction (Miles et al., 2008; Mastrocicco et al., 2012). When the more favorable electron acceptors are completely utilized, fermentation of the PHCs can still occur. This leads to the partial oxidation of the carbon atoms to $HCO_3^-$ and partial reduction of the carbon atoms to $CH_4$, as reported at contaminated aquifers (Morris et al., 2013; Lueders, 2017; Teramoto et al., 2020). The development in redox zones was simulated in the model as demonstrated by the change in redox potential (ORP) shown in **Figure 3.3**. Thus, ORP sensors could potentially detect PHC in groundwater especially when degradation affects the redox zones.

Following the model of Ng et al. (2015), concomitant $Fe^{3+}$-reduction and methanogenesis was simulated by using low $Fe(OH)_3$ solubility to produce $CH_4$ early as observed at the Bemidji site. This leads to higher $CH_4$ concentration at later times, but the partial equilibrium model addressed this by representing anaerobic oxidation of methane coupled with iron reduction (Amos et al., 2011; Ng et al., 2015). In general, long-distance transportation of reaction products such as methane through the aquifer is possible (Appelo and Postma, 2005), leading to an extensive $CH_4$ plume as shown **Figure 3.3**. The mineralization of PHCs also resulted in accumulation of $CH_4$, producing a larger $CH_4$ plume with higher concentration than BEX and NVDOC. However, $CH_4$ oxidizes at the fringes when $CH_4$-containing groundwater is mixed with oxygenated water.

### Solid phase simulation results

Near the source zone, the degradation of PHCs dissolves calcite because of proton formation (**Figure 3.2**); denitrification dissolves less calcite because of a decrease in proton formation in the redox reaction. Calcite dissolution (simulated as equilibrium process) increases Ca and alkalinity and thus acts as a buffer for groundwater acidity and increases the EC. As oxygen and nitrate become depleted, the subsequent reductive dissolution of iron hydroxides released $Fe^{2+}$ into the groundwater (Vencelides et al., 2007; Teramoto and Chang, 2019). Calcite also precipitated since iron reduction consumes protons, as illustrated by calcite formation at areas with low $Fe(OH)_3$.

The increased $HCO_3^-$ and $Fe^{2+}$ from PHC degradation coupled with reductive $Fe(OH)_3$ dissolution triggered siderite precipitation. Siderite can stabilize when the dissolved sulfide is depleted, allowing the $Fe^{2+}$ concentration to increase sufficiently (Appelo and Postma, 2005). This process continuously removed $Fe^{2+}$ from the solution and lowered the pH, further promoting $Fe(OH)_3$ dissolution and reduction (Ng et al., 2015).

Within the source zone, the cation exchange sites adsorbed the calcium ions (increased $CaX_2$ concentration), mobilizing $Mg^{2+}$, $K^+$, and $Na^+$ (**Figure A6**). However, the produced $Fe^{2+}$ from iron-reduction was partial sorbed to the CEC (increased $FeX_2$) and mostly exchanged with $CaX_2$, releasing more $Ca^{2+}$ to the water. $CaX_2$ is thus depleted in most of the plume, while $FeX_2$ is enriched. As observed at the Bemidji site, PHC oxidation increased groundwater acidity. This in turn increased the levels of calcium and magnesium in groundwater in response to calcite dissolution (Bennet et al., 1993; Amos et al., 2011). The resulting mineral depletion/enrichment patterns, however, differed from the study at the Bemidji site (Ng et al., 2015). Ca sorption was included in the model which was omitted in the sorption model for the Bemidji site, USA. The extensive elevated $Ca^{2+}$ plume reported at the Bemidji site was also simulated, highlighting the significance of calcite dissolution in controlling groundwater chemistry.

### Effect of hydrocarbon degradation on in-situ water quality

**Figure 3.3** presents the PHC concentration and the iWQP at three observation wells located directly below (X0Z0) the source zone and at horizontal distances of 85 m (X1Z3) and 145 m (X3Z3) away from the source zone, following the general direction of groundwater flow. The data from all observation wells were analyzed, but for illustrative purposes, the data from these three specific wells were chosen for visualization. BEX levels were also multiplied by five only for visualization since BEX has much lower levels than NVDOC (less than 10x). It must be noted that the accuracy and resolution of sensors must be considered when interpreting the practical implications of these findings. There are sensors with accuracy and resolution of ±0.1 and ±0.01 for pH, ±5 mV and ±0.1 mV for ORP, ±0.5% (of reading) and ±0.1 µS/cm for EC, and ±0.1 mg/L and ±0.01 mg/L for DO (In-Situ, 2024). Thus, small variations in sensor data might be difficult to detect in real-life applications.

The time series data in **Figure 3.3** reveals five distinct periods. The first period, spanning the initial 2.5 simulation years for X0Z0 (**Figure A7**), consists of aerobic conditions with BEX as the primary electron donor. This is consistent with the approximate travel time of particles from the source zone to X0Z0 which is 2 to 4 years (**Figure A8**). During this period, DO was completely consumed, leading to a stepwise decrease in ORP from +800 mV (associated with the $O_2/H_2O$ redox couple) to around −125 mV (associated with the $CO_2/CH_4$ redox couple) (McMahon et al., 2011). From the PHC degradation, $H^+$ and $HCO_3^-$ were produced, leading to a slight increase in EC from around 340 µS/cm to 440 µS/cm and a decrease of pH from about 7.7 to 7.2.

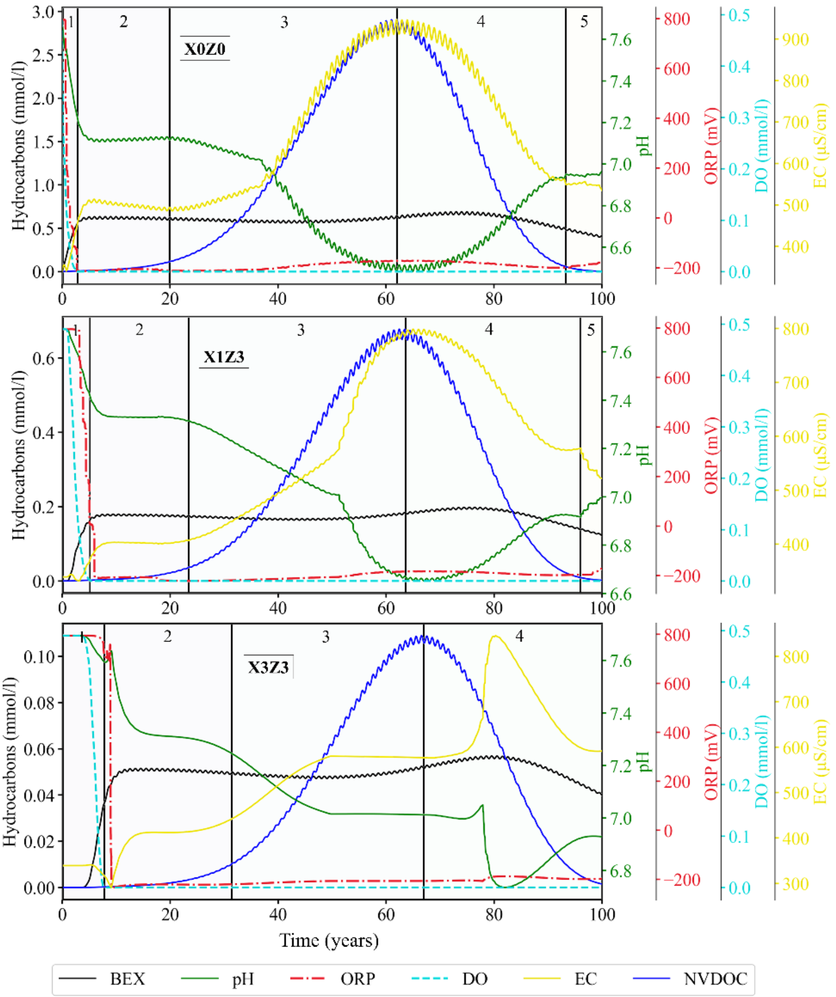


**Figure 3.3.** Simulation results of BEX, NVDOC and in-situ water quality parameters at three observation wells X0Z0 (top), X1Z3 (middle), and X3Z3 (bottom) for 100 years, divided into identified periods. BEX levels are multiplied by five for visualization.

The second period is characterized by BEX still acting as the main electron donor, but under anaerobic conditions. In this period, the pH further decreased to 7.1, while the EC further increased to about 510 µS/cm. After approximately 20 years, the third period started which marks the transition from BEX to NVDOC as the primary electron donor in the source zone which aligns with observations at the Bemidji site (Ng et al., 2015). The NVDOC concentration peaked at 61 years proving that petroleum contamination can accumulate and persist in groundwater for decades (Pishgar et al., 2022). As the degradation process progressed, groundwater further acidified with pH reaching the minimum value of 6.5. This is because NVDOC degradation produces more $H^+$ than BEX degradation per molecule (**eqn. 3.8** and **3.9**).

This caused the calcite to dissolve further, increasing the EC to the maximum value of about 950 µS/cm. An important observation is that the extreme EC and pH values were reached in about 1 and 3 years, respectively after the peak concentration of NVDOC was observed. A possible reason for this is the delay between the migration of NVDOC and its degradation by-products. Variation in reaction kinetics of the different hydrocarbons can also contribute to the lag time before the impact on water quality can be observed. For example, the first-order degradation coefficient of BEX is 0.31 (1/year) while for NVDOC it is 0.46 (1/year). The delay between the peak NVDOC concentration and the maximum values of EC and pH differs among observation wells, with a longer delay at the farthest well.

In the fourth period starting in year 61, the NVDOC concentration starts to decline from its peak, indicating the depletion of the NVDOC source, during which the pH increased to about 6.9 while the EC decreased to 560 µS/cm. ORP increased slightly during periods 3 and 4, possibly due to the mixing of water with different redox states in the observation well. The fifth and final period starts at approximately 91.5 simulation years until the end of simulation period. During this period, hydrocarbons are almost depleted leading to stabilization in iWQP, as illustrated at the end of the time series. However, the iWQPs have not returned to their initial levels, suggesting that a century is insufficient to completely remove the impact of PHC contamination. These time periods provide valuable insights into the temporal variations of hydrocarbon degradation and their impact on iWQP. Although it was known that PHC degradation can influence these iWQPs, the changes to water quality parameters have not been quantified and used to estimate the concentration of PHC in groundwater. Hence, these findings could serve as the core logic for virtual sensor development.

The starting and ending times of these characteristic periods depend on the position of the observation well with respect to the source zone, and more specifically the groundwater travel times between the source zone and the wells. These differences can be observed from the data at three observation wells shown in **Figure 3.3** and the approximate travel times based on particle tracking in **Figure A8** (Pollock, 2017).

Period 1 ends at around year 5 for X1Z3, which is related to the approximate travel time of 4 to 6 years from the source zone to the well, and around year 8 for X3Z3 with an approximate travel time of 8 to 10 years. Furthermore, data from observation well X3Z3 show that the presence of BEX was detected 2.5 years after the start of simulation, as compared to X0Z0 where BEX was detected within the first simulation year. The DO was completely consumed and the ORP dropped to -225 mV after about 6.8 years at X3Z3 while only about 2.5 at X0Z0. The change from BEX to NVDOC as the main electron donor was observed after approximately 31 years at X3Z3, although only after 20 years at X0Z0. The NVDOC concentration peaked at a simulation time of around 67 years, approximately 6 years later than at observation well X0Z0.

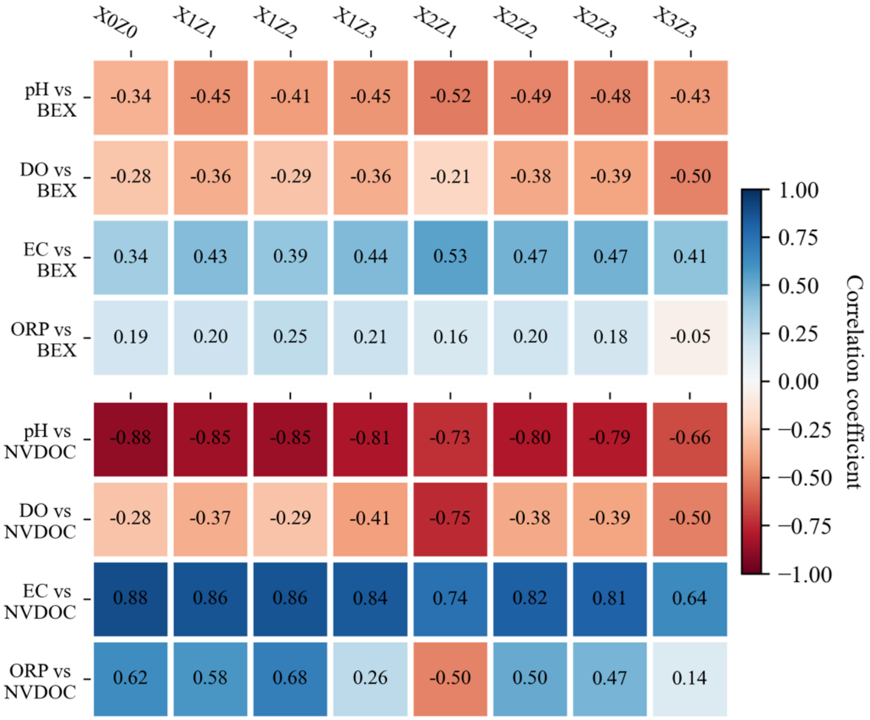
The travel time of PHCs from the source zone to the observation wells can be influenced by the aquifer's hydraulic and hydrologic properties (Fetter et al., 2018; Gupta and Yadav, 2020) as shown in **Figure A8**. At the farthest observation well X3Z3 (horizontal distance of 145 m from the source zone's center), a lag time of about 14 years was observed between the peak NVDOC concentration and the maximum EC value of about 850 µS/cm; the lowest observed pH value of 6.7 had a longer lag time of approximately 15.5 years.

Aside from the differences in travel times, it was also found that the signal from the water table fluctuation was dampened as the analysis moved from well X0Z0 closest to the left boundary, to the farthest well (X3Z3). This dampening is especially noticeable for pH and EC (**Figure 3.3**) and is attributed to the aquifer's heterogeneity that attenuates the signal as the water moves through the system. This is commonly observed in amplitude attenuation of tide-induced groundwater level fluctuations with the rate controlled by the aquifer's hydrogeologic properties (Rotzoll and El-Kadi, 2008; Huang et al., 2015; Housego et al., 2021). For clear illustration, the signal attenuation for the BEX tracer at different wells is also presented in **Figure A9**. This shows that the location of the observation well plays an important role in observing the behavior of hydrocarbon contaminants and their effects on the iWQPs.

### 3.3.2. Exploratory data analysis of simulation results

Simulated NVDOC, pH, and EC values at observation well X0Z0 have a multimodal distribution. BEX is negatively skewed while NVDOC is positively skewed (**Figure A10**). Therefore, Spearman's rank correlation was used instead of Pearson's correlation for data analysis. Spearman's correlation determines the direction and strength of the monotonic relationship between variables, rather than the linear relationship measured by Pearson's correlation (Sheskin, 2020).

At eight observation wells, 100 years of PHC and iWQP simulation data were analyzed (**Figure 3.4**). As anticipated, different observation wells displayed varying correlation coefficients due to the factors influencing the fate and transport of contaminants, as discussed earlier. The correlations between the iWQP and NVDOC were generally stronger than those with BEX, except for DO which exhibited similar correlations. At 95% significance level, pH demonstrated a strong to very strong negative correlation with NVDOC (-0.66 to -0.89) and a moderate negative correlation with BEX (-0.34 to -0.52). Conversely, ORP displayed a moderate positive correlation with NVDOC, except at wells X2Z1 and X3Z3 where the correlation weakened. ORP also had a weak positive correlation with BEX, except at well X3Z3 where the correlation was negligible (0.08). The early consumption of DO results in weak negative correlations with both BEX and NVDOC. The EC, however, exhibited weak to moderate positive correlations with BEX and strong to very strong positive correlations with NVDOC. These correlation coefficients support the previous discussions, confirming that the degradation of PHCs significantly influences groundwater pH and EC.



| | X0Z0 | X1Z1 | X1Z2 | X1Z3 | X2Z1 | X2Z2 | X2Z3 | X3Z3 |
|---|---|---|---|---|---|---|---|---|
| pH vs BEX | -0.34 | -0.45 | -0.41 | -0.45 | -0.52 | -0.49 | -0.48 | -0.43 |
| DO vs BEX | -0.28 | -0.36 | -0.29 | -0.36 | -0.21 | -0.38 | -0.39 | -0.50 |
| EC vs BEX | 0.34 | 0.43 | 0.39 | 0.44 | 0.53 | 0.47 | 0.47 | 0.41 |
| ORP vs BEX | 0.19 | 0.20 | 0.25 | 0.21 | 0.16 | 0.20 | 0.18 | -0.05 |
| pH vs NVDOC | -0.88 | -0.85 | -0.85 | -0.81 | -0.73 | -0.80 | -0.79 | -0.66 |
| DO vs NVDOC | -0.28 | -0.37 | -0.29 | -0.41 | -0.75 | -0.38 | -0.39 | -0.50 |
| EC vs NVDOC | 0.88 | 0.86 | 0.86 | 0.84 | 0.74 | 0.82 | 0.81 | 0.64 |
| ORP vs NVDOC | 0.62 | 0.58 | 0.68 | 0.26 | -0.50 | 0.50 | 0.47 | 0.14 |

**Figure 3.4.** Correlation Coefficients between In-situ Water Quality Parameters and PHCs (BEX and NVDOC) at all Observation Wells with Well Locations for the 100 Years of Simulation Results). All correlations are statistically significant, p-value < 0.05.

**Figure 3.5** illustrates the five different phases observed at well X0Z0. During period 1, strong negative BEX-pH and NVDOC-pH correlations were observed, while EC exhibited a strong positive correlation. In period 2, pH and EC started to correspond with BEX (Spearman's r of about -0.8 and 0.8, respectively) more than NVDOC (Spearman's r of around 0.3 and -0.6, respectively). This shift can be attributed to BEX being the main electron donor in this period. Additionally, the positive NVDOC-pH correlation may suggest a slight pH buffering effect due to calcite dissolution.

The transition of the main electron donor from BEX to NVDOC at period 3 (**Figure 3.5**) notably reduced the BEX-pH and BEX-EC correlations and substantially increased the NVDOC-pH and NVDOC-EC correlations. This is because the peak NVDOC concentration was around 2.9 mmol/L, as compared to the BEX concentration of 0.14 mmol/L. The decline of NVDOC in period 4 caused a point of inflection in the correlation coefficients at around 60–65-year range (**Figure 3.5**) but still maintained a strong correlation with pH and EC until the beginning of period 5. From period 3 to period 5, the NVDOC-pH and total PHC-pH correlations behaved similarly, as well as with EC. It is worth noting that although period 3 was delineated based on NVDOC being the dominant PHC, the effect of increasing NVDOC concentration can already be observed at the end of period 2 where the BEX-pH and BEX-EC correlation coefficients started to decrease. These findings were also observed in the other wells. This includes the transition in the primary electron donor from BEX to NVDOC, and the switch between positive and negative correlations during certain periods.

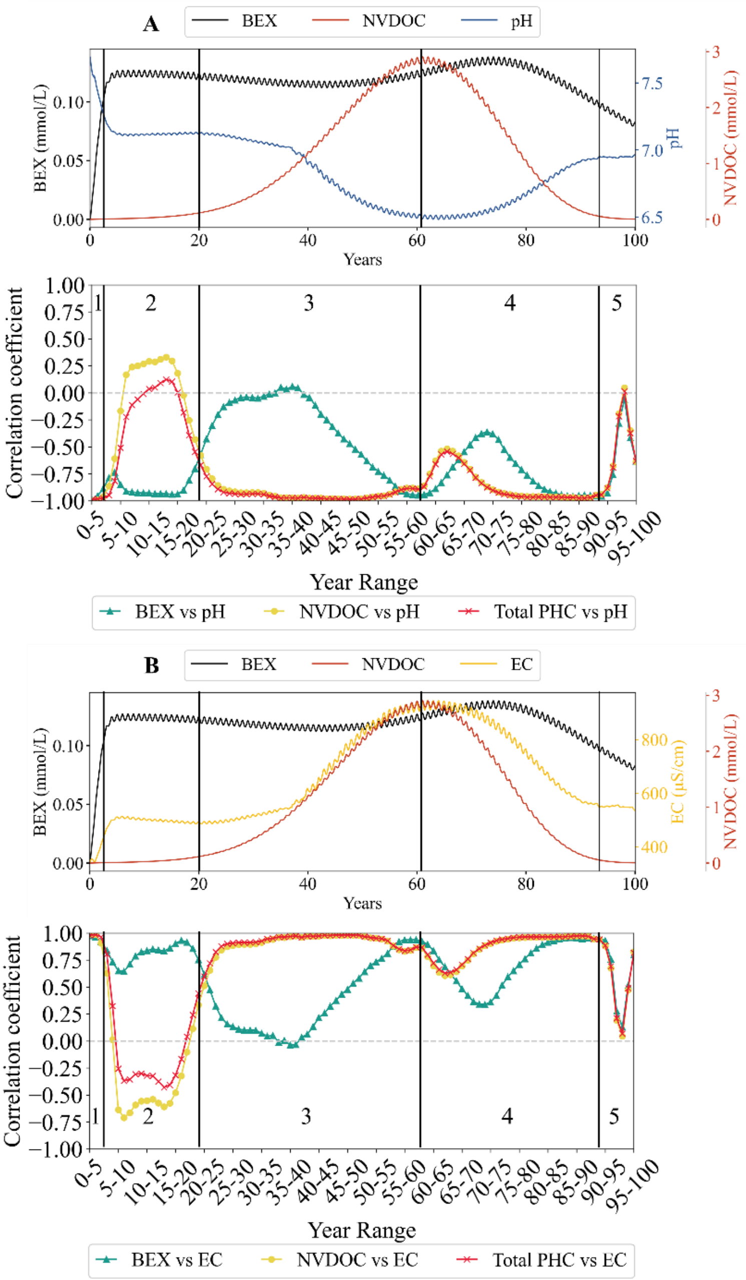


**Figure 3.5.** Spearman's rolling correlation with 5-year window of PHCs (BEX, NVDOC, and combined) vs pH (panel A) and EC (panel B) at observation well X0Z0 below the LNAPL source.

### 3.3.3. Influence of different scenarios on the correlation coefficients

The possible effects of various conditions on the correlation between PHC and iWQP were investigated. These factors have been known to influence the transport of ions and contaminants in groundwater, as well as the TEAPs and pH buffering capacity of the aquifer, among others. Spearman's rolling correlation coefficients of PHCs (BEX and NVDOC) versus pH and EC at observation well X0Z0 for different scenarios are presented in **Figure 3.6**. As can be seen, focus was given to pH and EC since after 2.5 years, DO was depleted resulting in a reduced aquifer condition. Identified periods from the base case scenario are visually delineated by black vertical lines within the figure for comparison with other scenarios.

The elevated background EC (EC+) slightly decreased the BEX-EC correlation during period 1 but increased this correlation during period 5. The correlation in periods 1 and 2 further decreased with a much higher background EC (EC++), which also affected the BEX-pH correlation, more noticeably during period 1. Higher background EC values reduced the noticeability of the signal from BEX contamination, causing this decrease in correlation (lower signal-to-noise ratio). NVDOC-pH and NVDOC-EC correlations exhibited the opposite effect, resulting in an increased correlation in the EC++ scenario mostly at period 2. This shows that the change in EC caused by NVDOC is significant even at high background EC.

Since oxygen, $NO_3^-$, and $SO_4^{2-}$ are important mainly during the initial periods (until these are exhausted), the BEX-pH and BEX-EC correlations remained relatively stable when DO and nitrate were removed (DO0 and $NO_3$0, respectively), with a slight decrease when sulfate was removed ($SO_4$0). However, increasing the nitrate and sulfate levels ($NO_3$+ and $SO_4$+, respectively) decreases the BEX-pH correlation significantly at period 2 while increasing the NVDOC-pH and NVDOC-EC correlations; $SO_4$+ also altered the direction of BEX-EC correlation at period 2, end of period 4, and period 5. By varying the available electron acceptors, the timing of the periods could also change since each period is affected by the presence of the main electron acceptor.

When the system has no calcite (Cal0), the BEX-pH and BEX-EC correlations significantly decreased causing the periods to be indistinguishable. Since pH was not buffered because of the absence of calcite, pH correlated more with NVDOC. Without calcite, the pH dropped more than 0.7 as compared to the base case while the EC increased more than 350 µS/cm at certain periods. Degradation of NVDOC (**eqn. 3.9**) produces more $H^+$ than BEX (**eqn. 3.8**), thus without buffering, NVDOC controls the groundwater pH. Decreasing the cation exchange capacity (CEC-) did not significantly impact the correlations, except for NVDOC-pH, which showed a slight increase in correlation during period 2. Increasing the cation exchange capacity (CEC+) reduced both the BEX-pH and BEX-EC correlations. For the NVDOC-pH and NVDOC-EC correlations, these were either increased during period 2, or period 1 became longer.

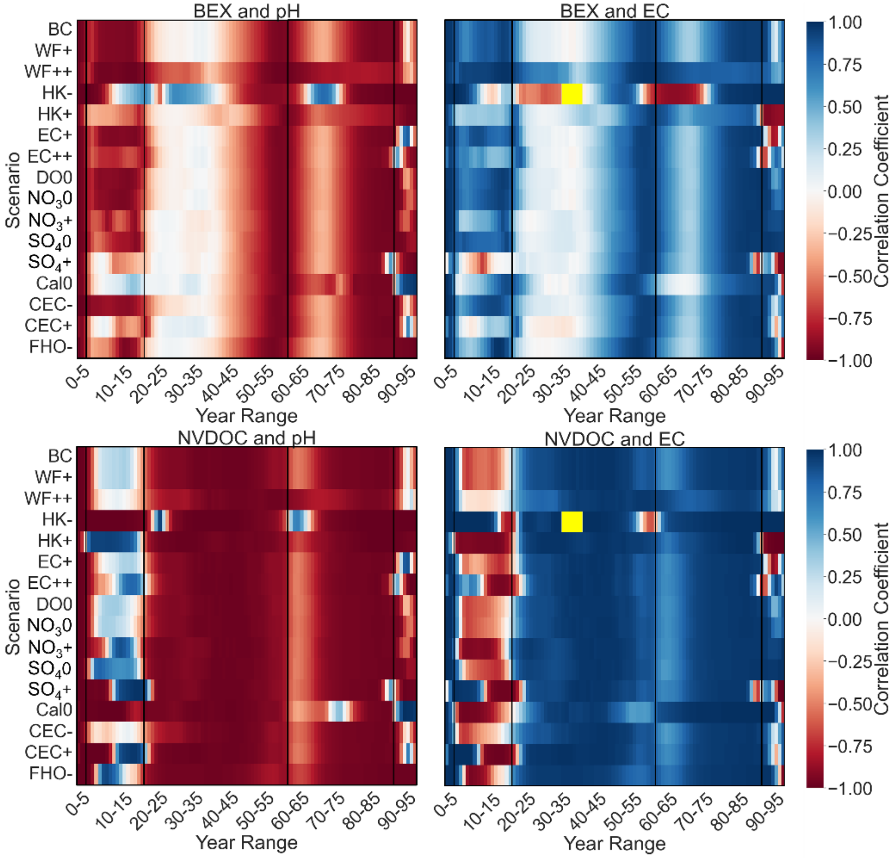


**Figure 3.6.** Spearman's rolling correlation with 5-year window of PHCs (BEX and NVDOC) vs. iWQP (pH and EC) at observation well X0Z0 for different scenarios (BC: base case, WF: water table fluctuation, HK: hydraulic conductivity, EC: electrical conductivity, DO: dissolved oxygen, $NO_3$: nitrate, $SO_4$: sulfate, Cal: calcite, CEC: cation exchange capacity, FHO: amorphous ferric hydroxide). Parameters were changed relative to BC (+ for increased, - for decreased, 0 for removed). Yellow rectangles mark the period with no EC data. The changed values are presented in **Table A4**.

The elevated background EC (EC+) slightly decreased the BEX-EC correlation during period 1 but increased this correlation during period 5. The correlation in periods 1 and 2 further decreased with a much higher background EC (EC++), which also affected the BEX-pH correlation, more noticeably during period 1. Higher background EC values reduced the noticeability of the signal from BEX contamination, causing this decrease in correlation (lower signal-to-noise ratio). NVDOC-pH and NVDOC-EC correlations exhibited the opposite effect, resulting in an increased correlation in the EC++ scenario mostly at period 2. This shows that the change in EC caused by NVDOC is significant even at high background EC.

Since oxygen, $NO_3^-$, and $SO_4^{2-}$ are important mainly during the initial periods (until these are exhausted), the BEX-pH and BEX-EC correlations remained relatively stable when DO and nitrate were removed (DO0 and $NO_30$, respectively), with a slight decrease when sulfate was removed ($SO_40$). However, increasing the nitrate and sulfate levels ($NO_3$+ and $SO_4$+, respectively) decreases the BEX-pH correlation significantly at period 2 while increasing the NVDOC-pH and NVDOC-EC correlations; $SO_4$+ also altered the direction of BEX-EC correlation at period 2, end of period 4, and period 5. By varying the available electron acceptors, the timing of the periods could also change since each period is affected by the presence of the main electron acceptor.

When the system has no calcite (Cal0), the BEX-pH and BEX-EC correlations significantly decreased causing the periods to be indistinguishable. Since pH was not buffered because of the absence of calcite, pH correlated more with NVDOC. Without calcite, the pH dropped more than 0.7 as compared to the base case while the EC increased more than 350 µS/cm at certain periods. Degradation of NVDOC (**eqn. 3.9**) produces more $H^+$ than BEX (**eqn. 3.8**), thus without buffering, NVDOC controls the groundwater pH. Decreasing the cation exchange capacity (CEC-) did not significantly impact the correlations, except for NVDOC-pH, which showed a slight increase in correlation during period 2. Increasing the cation exchange capacity (CEC+) reduced both the BEX-pH and BEX-EC correlations. For the NVDOC-pH and NVDOC-EC correlations, these were either increased during period 2, or period 1 became longer.

A reduction in amorphous iron oxide (FHO-) had a significant effect on the correlations during period 2 and period 5. It decreased both BEX-pH and BEX-EC correlations in period 2 but increased the BEX-pH, BEX-EC, NVDOC-pH, and NVDOC-EC correlations in both periods. As previously discussed, iron oxides play a role in reductive mineral dissolution, precipitation, and cation exchange. Thus, reducing its availability impacted these processes leading to different correlation coefficients. This shows the importance of iron oxides in characterizing groundwater quality with PHC contamination (Jiang et al., 2019). In general, increasing the electron acceptors can significantly affect the correlations between PHCs and iWQPs. The correlation coefficients were notably affected in period 2 and period 5. These periods correspond to a phase before a transition between electron donors occurs, highlighting the significance of TEAPs in characterizing these relationships. In contrast, the correlation coefficients in period 4 remained relatively stable except for major changes during scenarios HK- and Cal0. Furthermore, a decrease in BEX-pH or BEX-EC correlations is typically accompanied by an increase in NVDOC-pH or NVDOC-EC correlations, and vice versa.

### 3.3.4. Implications for virtual sensor development

Virtual sensors use models to process and transform physical sensor data into signals of interest (Brunello et al., 2021; Martin et al., 2021), such as PHC concentration. In contrast to RTM where the input parameters include the aquifer chemistry and the PHC, the input parameters for the hydrocarbon virtual sensor will only be the iWQPs; the virtual sensor was expected to detect and estimate the concentration of PHC in groundwater by combining these iWQPs. The correlation between the physical sensor data and the PHC concentration must therefore be determined to identify the most suitable process for virtual sensor development. Since the correlation coefficients have been found to have temporal dependencies causing varying lag times between PHC contamination and changes in water quality, at least two different virtual sensor systems can be conceptualized. These systems will vary in terms of purpose and the processing of sensor data; these will be the focus of future research and are not within the scope of this paper.

First is an early warning system utilizing the initial years following an oil spill (period 1). During the period when the condition is aerobic, PHC concentration has a perfect negative correlation with DO and ORP, making these parameters invaluable for early PHC detection. This system can be used to detect PHC contamination in real-time, allowing for a quick remediation response and thus preventing the PHC plume from spreading. Second is a system for long-term monitoring of PHC-contaminated aquifer, considering periods after period 1. EC and pH play important roles during these periods as these provide insights into the crucial chemical processes in groundwater. As the correlation coefficients differ per period due to variations in available electron acceptor and elector donors, among others, this system must have a mechanism for selectively updating the model based on the period that is monitored. This low-cost, low-maintenance, and continuous monitoring of PHC concentration in contaminated areas can be useful for developing effective and efficient remediation strategies across the lifespan of the plume. These findings serve as an essential intellectual footing for developing a robust hydrocarbon virtual sensor.

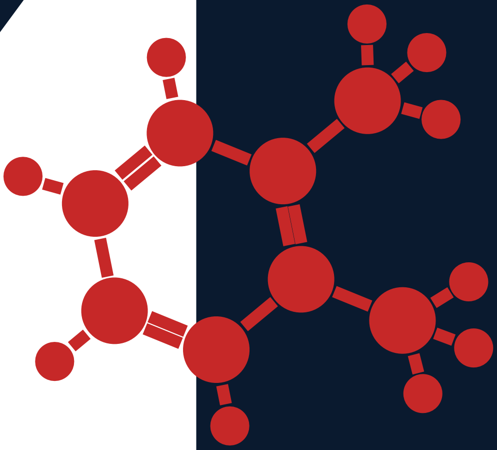
Hydrogeological properties of the aquifer were also found to affect the correlation coefficients. **Figure 3.6** shows that the coefficients change in the identified periods, and that the timing and duration of each period change. Thus, aquifer characteristics and geochemical conditions such as hydraulic conductivity and available minerals are important considerations when developing the virtual sensors. Other factors which were beyond the scope of this study might also affect the correlation coefficients and thus must be further studied. These factors include the source zone dimensions, heterogeneity in background aquifer chemistry, and hydrocarbon sorption on sedimentary organic matter. Field applications must also account for noises present in the data due to possible natural variations in groundwater conditions. Moreover, sensor accuracy, precision, and stability should be considered; small fluctuations in simulated water quality parameters were considered in the statistical analysis but may not be within the sensors' precision which can potentially result in weaker correlation coefficients.

## 3.4 Conclusion

This study investigated the potential of in-situ water quality sensors for estimating petroleum hydrocarbons (PHC) in groundwater. A reactive transport model was developed to generate synthetic datasets, which were essential towards creating a virtual sensor to detect PHC in groundwater. Exploratory and statistical data analysis reveals that several factors influence the correlation between in-situ water quality parameters (iWQPs) and PHCs. These correlation coefficients exhibit spatial and temporal variability among observation wells and among the five identified periods in the base case scenario. Scenario simulations revealed the impact of various factors, including water table fluctuations, hydraulic conductivity, aqueous and mineral phase electron acceptors, and other geochemical properties of the aquifer, on the correlation between PHCs and iWQPs. These findings demonstrate the feasibility of developing two virtual sensor systems: an early warning system for rapid responses to oil leakages affecting groundwater; and a continuous, real-time, long-term monitoring system aimed at devising effective remediation strategies for PHC-contaminated aquifers.

# Machine Learning-based In-situ Detection of Toxic Petroleum Hydrocarbons in Groundwater

# 4



## Abstract

Monitored natural attenuation is commonly used to manage petroleum hydrocarbon-contaminated groundwater. However, it requires periodic, costly grab sampling. A cost-effective, real-time groundwater monitoring proof-of-concept machine learning (ML) framework was proposed using in-situ sensors—pH, dissolved oxygen, electrical conductivity, and redox potential—to detect benzene, ethylbenzene, and xylenes (BEX). This study built upon the established correlations between hydrocarbon concentrations and in-situ water quality parameters (iWQPs). Due to limited field data, the framework was validated using datasets at virtual wells within a simulated aquifer from the previously developed reactive transport model. In this application, the spreading of pollution downstream of the established pollution plume was detected. The used framework is a binary classification system that flags contamination at virtual downstream wells. Five ML classifiers, i.e. Logistic Regression, Random Forest, XGBoost, Multi-layer Perceptron, and Support Vector Classifier, were compared for early warning when BEX reached or exceeded the regulatory threshold of 5 µg/L. The models were trained on virtual wells at and near the source zone and predicted contamination before BEX reached the threshold at downstream virtual wells. This reflects the spatial variability in flow and reaction dynamics that altered BEX-iWQP relationships. Scenario analyses revealed the ML models' sensitivity to aquifer properties, i.e., hydraulic conductivity, electrical conductivity, and electron acceptor availability. The impact of sensor noise and seasonal fluctuations on iWQPs were also assessed. Even moderate levels of noise (10–20%) can significantly affect model accuracy, particularly when the noise was introduced into the test data. Therefore, it is recommended to combine hardware stabilization with adaptive smoothing techniques. With these approaches, the proposed framework remains promising for providing early warnings of plume migration toward sensitive receptors.

This chapter has been published as:

Wu, C.L.R., Wagterveld, R.M., Rietveld, L.C., van Breukelen, B.M. (2026). Machine learning-based in-situ detection of toxic petroleum hydrocarbons in groundwater. Journal of Contaminant Hydrology. 104771, ISSN 0169-7722. https://doi.org/10.1016/j.jconhyd.2025.104771

The repository associated with this chapter can be accessed at:

https://doi.org/10.4121/9dea333c-7439-4c39-9f74-b7904d6d4a83

## 4.1 Introduction

Petroleum hydrocarbons (PHCs) are widely used across various sectors, including residential, agricultural, and transportation industries. Among these compounds, aromatic hydrocarbons such as benzene, toluene, ethylbenzene, and xylenes (BTEX), are particularly hazardous due to their toxicity and persistence in groundwater (Li et al., 2021). These contaminants often infiltrate groundwater through landfill leachate, leaks from underground storage tanks, and industrial discharges (Haider et al., 2021). To safeguard water resources, agencies such as the U.S. Environmental Protection Agency established strict guidelines for permissible levels of organic contaminants in water systems (US EPA, 2024).

Conventional PHC monitoring in groundwater uses manual grab sampling and laboratory analysis. While accurate, this approach is costly, infrequent, and lacks the capability to provide real-time data. These limitations are especially crucial when Monitored Natural Attenuation (MNA) is the remediation strategy. MNA relies on naturally occurring physical, chemical, and biological processes within the soil and groundwater to reduce contamination levels, offering a cost-effective and environment-friendly alternative to immediate full-scale remediation (Beck and Mann, 2010).

Under MNA, contaminant plumes are expected to shrink over time, resulting in relatively stable conditions with respect to receptor exposure. Contaminant concentrations should not exceed the compliance limits at designated warning wells, located between the contaminant source and the receptor areas. Receptors encompass both human populations and ecological systems dependent on groundwater resources (McKnight et al., 2010). Exceeding contaminant thresholds at warning wells may signal an unaccounted pollution source, requiring early warning systems and prompt remediation to protect receptors.

Consequently, there is a need for cost-effective, continuous monitoring systems capable of real-time PHC detection in groundwater. Advances in sensor technology and machine learning (ML) offer promising solutions. Emerging portable sensors are being developed for field-based detection of PHCs (Cova et al., 2022). But these remain costly, require complex sample preparation, and have slow response times.

ML models and sensor fusion have predicted contamination in both surface and groundwater using environmental and hydrogeological variables. LightGBM and XGBoost (XGB) effectively predicted *E. coli* levels on three Lake Erie beach sites (L. Li et al., 2022). Logistic regression (LR), random forest (RF), and support vector machines have estimated contamination probabilities of specific pollutants in groundwater. These include using static well measurements to predict nitrate and arsenic contamination (Singh et al., 2021). Linear and RF regressors were also used to predict and identify wells with high PFAS levels in California groundwater using co-contaminant fingerprints, hydrological and soil properties, proximity to pollution sources, and geospatial data (George and Dixit, 2021). Additionally, continuous sensor data integrated with Kalman filter had estimated tritium and uranium concentrations, enabling real-time plume tracking (F. Schmidt et al., 2018).

Despite these advancements, a significant gap remains in using continuous data from conventional, low-cost sensors to detect PHCs at warning wells. Although Li et al. (2017) monitored groundwater for potential oil and gas contamination using sensor data, PHCs were not directly detected; instead, anomalies were identified by comparing new data to historical records, without converting data to PHC levels. Furthermore, integrating sensor data with ML models for real-time PHC detection has not been explored. For MNA, early warning of contaminant migration toward the receptors is essential for triggering active remediation measures.

Indirect detection of PHCs is possible by monitoring their degradation processes, which influence groundwater quality through various terminal electron-accepting processes (**Chapter 3**). Studies have shown that following an inland oil spill, these degradation reactions drive measurable changes in geochemical conditions. To better understand these processes at contaminated sites, reactive transport models (RTMs) have been developed (e.g., Ng et al., 2015).

In an earlier study, an RTM was developed to examine how PHC degradation affects groundwater quality under hypothetical yet representative conditions. A stationary oil source dissolving at the top of a heterogeneous, shallow sandy aquifer was simulated in two dimensions. The model showed the spatiotemporal evolution of dissolved PHCs under various realistic conditions and revealed correlations between PHC concentrations and in-situ water quality parameters (iWQPs), namely pH, dissolved oxygen (DO), electrical conductivity (EC), and redox potential (ORP) (**Chapter 3**). ORP in water measures the tendency of a chemical species to gain (reduction) or lose (oxidation) electrons (Copeland and Lytle, 2014). These findings suggested that conventional water quality sensors could detect PHCs in groundwater.

Qiao et al. (2025) have recently presented a study that closely parallels the methods and frameworks introduced in the earlier work (**Chapter 3**). They made minor modifications to the original Python-based RTM to generate the datasets required for the ML-based integration of iWQPs in estimating PHC concentrations in groundwater. While their stated objective was to "predict the spatiotemporal distribution of dissolved-phase NAPL plumes," this study focused on the early detection of benzene, ethylbenzene, and xylenes (BEX) compounds at warning wells across a range of aquifer conditions. Toluene was excluded in the BTEX group due to its tendency to degrade directly from the pollution source without dissolving (Ng et al., 2015).

A proof-of-concept ML framework that integrates data from affordable in-situ sensors was presented to detect BEX in groundwater. BEX compounds were grouped based on their shared migration and degradation properties (Ng et al., 2015). The sensors include pH, DO, EC, and ORP. Due to limited field data, the framework was validated using RTM-generated datasets from virtual observation wells. The task was framed as a binary classification problem, determining whether a virtual well is contaminated (≥5 µg/L BEX, based on U.S. EPA standards) or uncontaminated. Based on current knowledge, this is the first study to explore the use of low-cost sensor data in combination with ML to provide early warnings of PHC contamination at downstream wells. This approach differs from prior work by directly predicting exceedances of regulatory thresholds, rather than mapping plumes or identifying anomalies (Li et al., 2017). In essence, this ML framework was trained to detect contamination at warning wells and provide a timely warning of contaminant migration before it reaches sensitive receptor areas.

Additionally, the ML models were evaluated across diverse hydrogeological conditions by training and testing separate models for each RTM scenario detailed in **Chapter 3**. In the scenario simulations, the amplitude of the water table fluctuations, hydraulic conductivity, and background salinity were varied. Electron acceptor availability was also modified by removing or increasing DO, $NO_3^-$, and $SO_4^{2-}$ concentrations. Mineralogical controls were also adjusted via calcite presence, $Fe(OH)_3$ reduction, and cation exchange capacity.

To further evaluate the ML models under realistic conditions, noise was introduced to iWQPs from the previous RTM (**Chapter 3**). Incorporated in this study are (1) Gaussian sensor noise based on commercial specifications, and (2) sinusoidal fluctuations to simulate seasonal groundwater variations. These were introduced only to the RTM outputs. The models' performance was then evaluated on these noisy datasets by comparing BEX arrival times from both the RTM and ML predictions at virtual warning wells. To improve detection accuracy under noisy conditions, a moving-average filter with varying window sizes was implemented.

## 4.2 Methods

### 4.2.1. Reactive Transport Modeling as a Data Source

The previously developed RTM in **Chapter 3** simulates the dissolution, advection, dispersion, and biodegradation of BEX and non-volatile dissolved organic carbon (NVDOC) from a light non-aqueous phase liquid (LNAPL) source in a porous media aquifer. Furthermore, the model accounts for the direct degradation of other components within the source zone. In addition to organic processes, the RTM incorporates inorganic reactions such as mineral dissolution and precipitation, cation exchange, and redox-related outgassing. The model's domain and parameter values were based on existing studies on PHC contamination, particularly the extensively studied crude oil spill site in Bemidji, Minnesota (Ng et al., 2015). In the previous study, the correlation between BEX and various in-situ water quality parameters (iWQPs) were examined, specifically pH, DO, ORP, and EC. Based on these findings, ML models were developed to detect the arrival of dissolved BEX using iWQPs at virtual warning wells located at various depths and distances from the oil source.

While synthetic datasets generated by RTMs offer a controlled and informative environment for proof-of-concept development, they cannot fully capture the complexity and variability of real-world aquifers including microbial variability, contamination source changes, and sensor fouling. As such, the applicability of this ML framework to field conditions remains to be validated. Still, the use of RTM facilitates generalizability by allowing testing with diverse hydrogeochemical scenarios without the constraints of cost, safety, or other physical limitations typically associated with field studies.

### 4.2.2. Machine Learning Model Development

In this study, focus was given on binary classification problem: determining whether a virtual observation well is contaminated or uncontaminated. Classification problems are a type of supervised learning task where each data point is associated with a class label (Sen et al., 2020). The goal is to develop a function or a classifier that can accurately categorize data points into the correct classes based on their features. To be effective, this classifier must demonstrate both high predictive accuracy—meaning it correctly classifies most of the data it is tested on—and a low generalization error, which refers to its ability to maintain that accuracy when applied to new, unseen data (Wang and Shen, 2006). Rather than estimating exact BEX concentrations, the aim is to demonstrate how an early-warning system based on inexpensive water quality sensors can support timely field investigations during contamination events. For stakeholders, a binary output indicating the presence or absence of contamination offers a more practical and interpretable first step in risk management. This simplification, in which 5 µg/L and 9.9 mg/L are treated as the same class, was designed to prioritize rapid risk identification over accurate concentration estimation.

Several of the most widely used ML classifiers were implemented, which have demonstrated their effectiveness across various domains (Ahsan et al., 2021; Li et al., 2022). These include LR, Support Vector Classifier (SVC), RF, XGB, and Multi-layer Perceptron (MLP) as described in **Appendix B** (**Text B1**). Classification ML models were chosen over regression to align with the practical goal of providing contamination alarms. By framing the problem as a binary classification task, the model outputs can be directly interpreted as the likelihood of contamination for decision making. Classification also provides a more direct and interpretable framework for defining decision thresholds for early detection system. To classify contamination levels, the U.S. EPA's maximum benzene concentration limit of 5 µg/L in drinking water was used as a threshold during ML model development; benzene is recognized as the most toxic constituent within the BEX group. In comparison, ethylbenzene and xylenes have higher allowable concentrations in drinking water, which are 700 µg/L and 10,000 µg/L, respectively (US EPA, 2024).

At the threshold of 5 µg/L total BEX concentration, benzene concentration is below its 5 µg/L limit since the value represents the sum of all constituent concentrations. However, for the purpose of developing a sensitive classification model, the maximum limit for benzene was applied as a conservative, health-protective threshold for the total BEX group. This approach prioritizes the detection of any potential benzene presence, ensuring a safety margin that flags samples for further investigation even if the immediate cumulative risk is low.

The ML model development process starts with creating a database that captures the relationship between BEX concentration and the iWQPs. Due to limited high-resolution temporal data from contaminated sites, including the well-studied Bemidji site, the RTM was used to generate a comprehensive virtual dataset for model training and evaluation. In this study, the focus is on the initial 5 years of simulation, when BEX was the primary electron donor. **Figure 4.1** illustrates the simulated cross-sectional BEX plume after 5 years (**Chapter 3**). DO, pH, EC, and ORP were selected as input variables based on strong correlations with PHCs (**Figure B1**). To capture spatial variability, the depth and horizontal distance of each virtual observation well from the contamination source (**Table B1**) were also included in both training and test datasets.

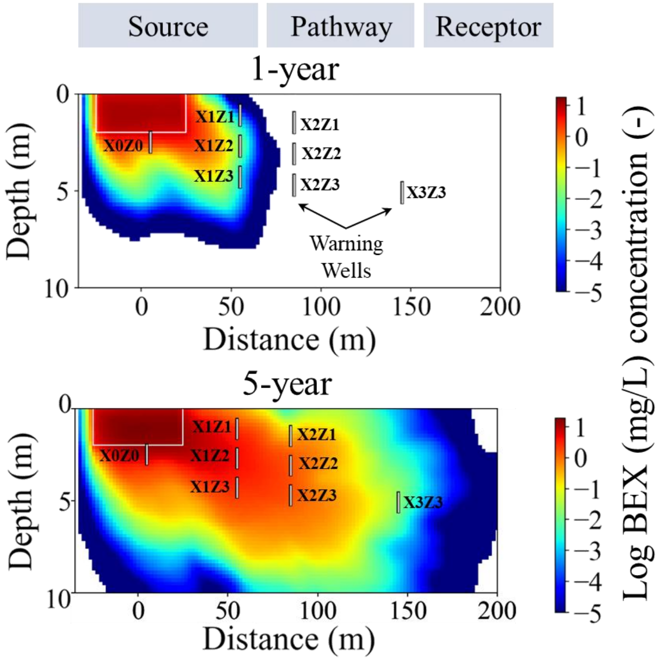


**Figure 4.1.** BEX plume distribution at 1 and 5 years of simulation. BEX concentration is converted to base 10 log form with 1 mg/L as reference. The white square indicates the oil source zone, and the eight vertical white lines represent virtual observation wells. BEX concentration reached a 5 µg/L threshold at the farthest virtual warning well approximately 4 years into simulation.

Model training and evaluation followed a two-layer approach combining Tree-Structured Parzen Estimator (Watanabe, 2023) for hyperparameter optimization and leave-one-group-out cross-validation across four virtual training wells. In each iteration, one well was used for validation while the remaining three were used for training. A custom scoring function was used to prioritize timely detection of BEX contamination, balancing prediction accuracy (F1 score) and delay in alarm timing. While the F1 score reflects classification performance, it does not capture how quickly contamination events are detected. Therefore, a delay metric was introduced, defined as the time difference between the true arrival of contamination based on RTM and the first instance of two consecutive contamination predictions from the ML model. This combined metric was designed to reflect both predictive accuracy and the practical requirement of early warning in groundwater monitoring.

Although techniques were not applied explicitly to address class imbalance, it was determined for the scope of this study that the minority class remained sufficiently represented to support stable model training and evaluation. However, the potential impact of class imbalance can be explored in future work. Additionally, penalty weighting for missed contamination events were incorporated. A requirement of at least two consecutive detections was also adopted to trigger an alarm, reducing the impact of isolated false positives. Details of the model training and hyperparameter tuning process are provided in **Text B2**.

To evaluate the real-world applicability, sensitivity analyses was conducted using a realistic base case and additional simulations reflecting varied but plausible hydrogeological conditions. The base case, based on typical field settings (**Chapter 3**), served as a benchmark. Meanwhile, scenario simulations assessed the model performance under different boundary conditions, aquifer characteristics, and groundwater chemistries. ML models were trained and tested across these diverse conditions to assess their robustness and adaptability to field-relevant scenarios.

### 4.2.3. Data Preprocessing

A fundamental concept in ML is the separation of training and testing data to ensure unbiased performance evaluation (Xu and Goodacre, 2018). The training dataset is used to calibrate the models, allowing it to learn the relationships and patterns inherent in the data. The testing dataset is then used to evaluate the models and assess the model's generalization capabilities and predictive accuracy on unseen data. This process is crucial in preventing overfitting, a phenomenon where the model performs exceptionally well on the training data but poorly on new data.

Unlike random shuffling approaches that mix all available data (Zhang, 2025), a spatially structured training-testing split that reflects the natural time progression of contamination (i.e., plume migrates from upstream to downstream) was implemented. This method avoids potential biases that arise from mixing data across different locations and time periods. The models were trained on 5-year data from virtual wells near the source zone (X0Z0, X1Z1, X1Z2, X1Z3) where BEX concentrations exceed regulatory thresholds. The models were then tested on downstream virtual wells—which were identified as warning wells (X2Z1, X2Z2, X2Z3, X3Z3)—to simulate early detection in uncontaminated areas. The receptor boundary was defined beyond the furthest virtual well (X3Z3). This approach achieves the practical objective of predicting contamination migration using in-situ sensor data.

Since the iWQPs vary in scale and units, the data were standardized using the StandardScaler function from the scikit-learn library, transforming each parameter to have a mean of 0 and a standard deviation of 1. This ensures that all parameters contribute equally to model training, preventing any single feature from disproportionately influencing the results (Ahsan et al., 2021). While tree-based models (e.g., RF and XGB) are generally invariant to feature scaling, standardization was applied uniformly across all models to maintain a consistent preprocessing method. This transformation is not expected to significantly affect the performance of tree-based models.

Data were categorized into two groups: non-contaminated (assigned a value of 0) for all data with BEX levels of 0 to below 5 µg/L and contaminated (assigned a value of 1) for all data with BEX levels of 5 µg/L and above. This binary labeling approach (assigning a label of 0 or 1) is widely used in binary classification tasks (Li and Tong, 2020). The highest BEX level in the dataset from the base case is 9.9 mg/L, observed at the virtual observation well X0Z0.

### 4.2.4. Model Evaluation

As a proof-of-concept, the performance of the ML models was assessed based on their ability to detect when BEX reached the downstream virtual observation wells (X2Z1, X2Z2, X2Z3, and X3Z3) using the virtual iWQPs. The performance was quantified by comparing the BEX arrival time predicted by the ML models with the actual arrival time determined by the RTM. A key criterion in the evaluation was whether the ML model could trigger an alarm when BEX contamination occurred in a virtual observation well, specifically when it reached the threshold of 5 µg/L.

When the ML model triggered an alarm, its prediction to the RTM-generated data was compared. A threshold breach was defined as the moment when BEX concentrations reached or exceeded the regulatory limit of 5 µg/L. If the RTM showed BEX concentrations above 0 µg/L (not detectable or ND) but still below 5 µg/L, the event was identified as a premature alarm, meaning the ML model predicted a potential risk before the threshold was breached. In these cases, the number of days the alarm occurred before the threshold was breached was recorded. If the RTM showed BEX concentrations as ND, the ML alarm was considered a false alarm, meaning the ML model predicted contamination when none was present. If BEX concentrations from the RTM reached or exceeded 5 µg/L but the ML model did not trigger an alarm on the same day, it was labeled a missed alarm. This means that contamination was present, but the ML model failed to detect it. Missed alarms were evaluated by calculating the delay in the model's response, defined as the number of days between the actual threshold breach and the ML model's alarm trigger.

### 4.2.5. Introduction of Sensor Noise and Seasonal Fluctuations in Groundwater Chemistry

Water quality sensors are inherently prone to measurement errors (de Winter et al., 2019). Therefore, measurement noise was integrated into the virtual data to accurately represent real-world conditions. Zero-mean Gaussian noise was added to the RTM output data using fixed standard deviations that match the accuracy of commercial sensors (**Table B2**). This measurement noise was introduced only at the RTM output stage, not during the RTM simulation. This approach for adding noise follows the method outlined by Zhu and Wu (2004), in which noise is injected into the training data, test data, or both to simulate measurement uncertainty (**Figure B2**). The noise levels were increased in a stepwise manner, using 10% increments, where 100% noise corresponds to the standard deviation specified by the sensor's accuracy.

Changes in iWQPs induced by BEX degradation are the signal of interest which were observed. While changes in iWQPs may also influence BEX degradation, first-order biodegradation kinetics were assumed in the RTM that are temporally and spatially uniform and independent of redox conditions (**Chapter 3**). This simplification allowed isolating the effect of BEX degradation on iWQPs. It was determined that the aquifer's initial baseline iWQP values do not significantly influence the magnitude of this signal, as the observed changes are governed primarily by BEX degradation kinetics. Therefore, it was assumed that the noise augmentation approach—introducing noise specifically to iWQPs—adequately captures the real-world noisy measurements.

Whereas the RTM assumed that the groundwater flowing into the pollution zone had a constant chemistry, groundwater composition can vary spatially and temporally due to anthropogenic exploitation, natural mixing processes, and varying recharge conditions (Yao et al., 2024). While bulk groundwater chemistry responds slowly to these changes, depending on aquifer characteristics (Khatri and Tyagi, 2015), iWQPs are generally more sensitive to environmental fluctuations. Additionally, iWQPs are measured using sensors characteristically sensitive to (seasonally changing) temperature. While automatic temperature compensation (ATC) is typically applied, temperature fluctuations can still introduce variability in sensor readings. To simulate the effect of seasonal variations in the composition of clean upstream groundwater, sinusoidal annual fluctuation was superimposed on the RTM-derived iWQPs. Fluctuations were introduced exclusively at the RTM output stage while maintaining noise-free RTM simulations. These fluctuations were introduced solely to the training data, in contrast to sensor noise, which was added to both training and testing datasets.

To introduce sinusoidal fluctuations into the iWQPs, it was first needed to determine the signal amplitude ($A$) to generate the synthetic sine wave for each parameter. Standard deviation of field measurements ($\sigma_{measurement}$) was first calculated using continuous measurement data obtained from the U.S. Geological Survey (n.d.) and from a published study (Lyons et al., 2023) (**Table B3**). For a pure sine wave with zero mean, $A$ is related to the standard deviation of the signal ($\sigma_{signal}$) through the root-mean-square relationship (Irvine, n.d.): $\sigma_{signal} = 1/\sqrt{2}A$ .

Since $\sigma_{measurement}$ includes sensor noise, $\sigma_{signal}$ was estimated by removing the noise component using the following expression: $\sigma_{signal} = \sqrt{\sigma_{measurement}{}^2 - \sigma_{sensor}{}^2}$, where the standard deviation of the sensor, $\sigma_{sensor}$ corresponds to the technical accuracy of the sensor. Because the actual accuracy of the USGS sensors is unknown, the variance based on the specifications of sensors used in the noise simulations (**Table B2**) was subtracted. For EC, the sensor accuracy is expressed as a percentage of the measured value, which complicates amplitude calculation because the resulting $\sigma_{sensor}$ varies with each reading. To simplify, the mean EC value from the USGS data (approximately 520 µS/cm) was used. Following the same procedure used for sensor noise, the amplitude of the sinusoidal fluctuation was increased in 10% increments, where 100% corresponds to the calculated amplitude $A$.

Thus, the performance of the ML models was evaluated under three different scenarios: 1) data with added sensor noise, 2) data with added sinusoidal annual fluctuations in clean groundwater composition, and 3) data combining both factors.

Finally, various methods for filtering the data with added noise and fluctuations were explored to enhance the predictive performance of the ML models. These included applying smoothing techniques using 3-day, 5-day, and 10-day averages. **Figures B3a**, **B3b**, and **B3c** illustrate the iWQPs at virtual training well X0Z0 under conditions of 10%, 50%, and 100% added sensor noise and sinusoidal fluctuations, along with the corresponding 5-day moving average smoothing.

## 4.3 Results and Discussion

In this section, the performance evaluation of the ML models was present through four key assessments. First, the baseline accuracy was established by testing the models using noise-free RTM-generated base case data. The goal is to trigger a contamination alarm when BEX concentration reached 5 µg/L. While detection earlier than this threshold is generally acceptable, the models were trained to prioritize timely rather than excessively premature alarms to avoid unnecessary field responses. For example, an alarm triggered 5 days before the threshold breach is considered more desirable than one triggered 30 days early. However, there is no standardized optimal time for premature alarms, as alarms triggered far in advance may resemble false positives and can be costly in terms of operational impact.

Next, the ML model’s adaptability to varying conditions were assessed using data from scenario simulations (**Chapter 3**). Their robustness was then tested by introducing artificial noise at multiple intensity levels to the base case data and quantifying the corresponding performance degradation. Finally, the effectiveness of noise-reduction techniques in restoring the timely detection accuracy of the ML models was evaluated.

To reflect the practical goal of early warning, model performance was evaluated using a custom scoring function that prioritizes timely detection of BEX contamination. The score penalizes premature alarms and missed contamination events more heavily than overall classification accuracy. Details of this scoring approach are provided in **Text B2**. These assessments demonstrate the capabilities and limitations of the ML models, and the implications for real-world groundwater monitoring application.

### 4.3.1. Baseline Performance Evaluation

The five ML models trained to detect BEX performed differently in triggering the contamination alarm when tested on virtual warning wells. LR, RF, XGB, and MLP models triggered the alarms in all virtual warning wells before BEX concentrations reached the 5 µg/L threshold (**Figure 4.2**). In virtual warning wells X2Z1, X2Z2, X2Z3 (located 85 m from the source zone’s center), these concentrations ranged from 3.4 µg/L to 4.7 µg/L. These values correspond to predictions that predated the actual concentration threshold breach by 25 to 4 days. In contrast, SVC triggered alarms at the X2 virtual warning wells when BEX concentrations exceeded the threshold (6.4 µg/L to 7.1 µg/L), resulting in a delayed alarm trigger of 12 to 19 days.

At the farthest virtual warning well, X3Z3 (located 145 m from the source zone's center), most models triggered alarms earlier than expected. The LR model triggered the alarm 189 days earlier at a concentration of 0.6 µg/L, while the MLP model triggered the alarm 126 days earlier at 1.2 µg/L. The RF and XGB models showed moderately better performance, triggering alarms 54 days (at 2.6 µg/L) and 61 days (at 2.4 µg/L) earlier, respectively. However, the SVC performed best at this virtual warning well, triggering the alarm 18 days earlier at a concentration of 4.1 µg/L.

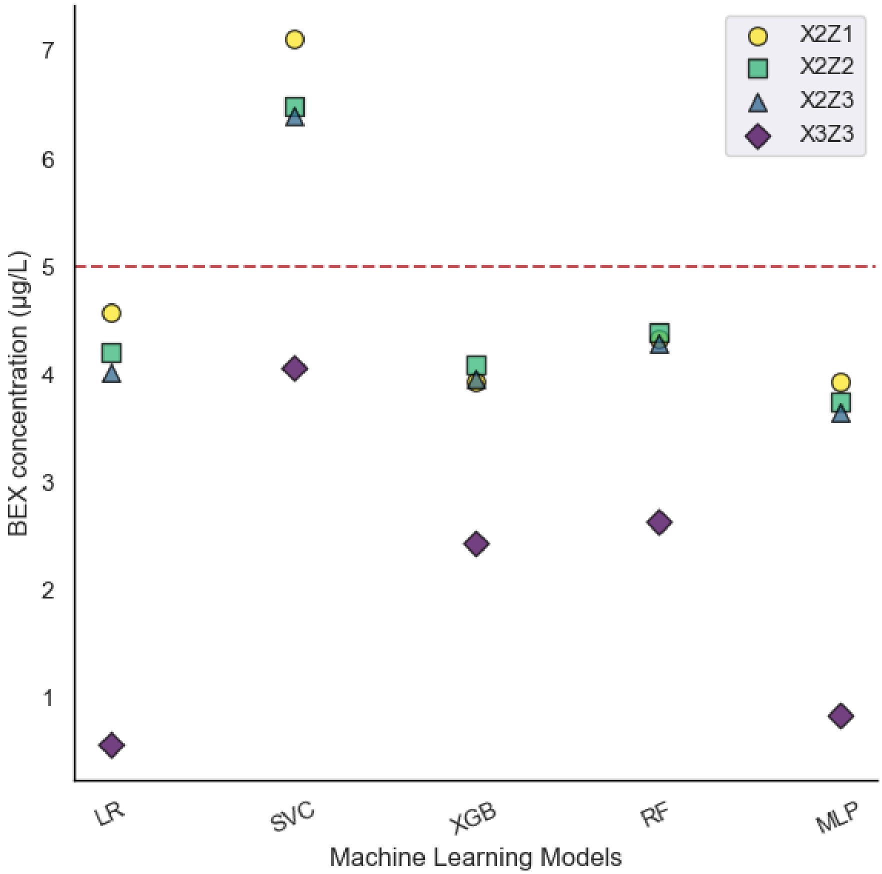


**Figure 4.2.** Concentration of BEX when the different ML models triggered the contamination alarm at the virtual warning wells. Red dashed line represents the US EPA threshold of 5 µg/L.

Most ML models predicted contamination prior to the actual event. This discrepancy likely stems from differences between the training and test datasets: the models were trained on data from source-zone virtual wells (X0Z0) and nearby virtual wells (55 m from the source center) but tested on downstream virtual wells farther from the source (warning wells). Due to varying flow and reaction dynamics, the relationships between BEX and iWQPs were not consistent across distances. Specifically, BEX arrived more quickly at closer virtual wells with higher concentrations, whereas geochemical reactions had more time to influence the iQWPs at farther virtual wells, leading to different parameter relationships (**Figure 4.3**). This mismatch in geochemical conditions across distances affected model generalization, as seen in the premature alarms at X3Z3. The results suggest that the initial ML framework was less robust at distant wells, and that either warning wells should be placed closer to the source zone or additional training data from farther locations is needed to improve performance.

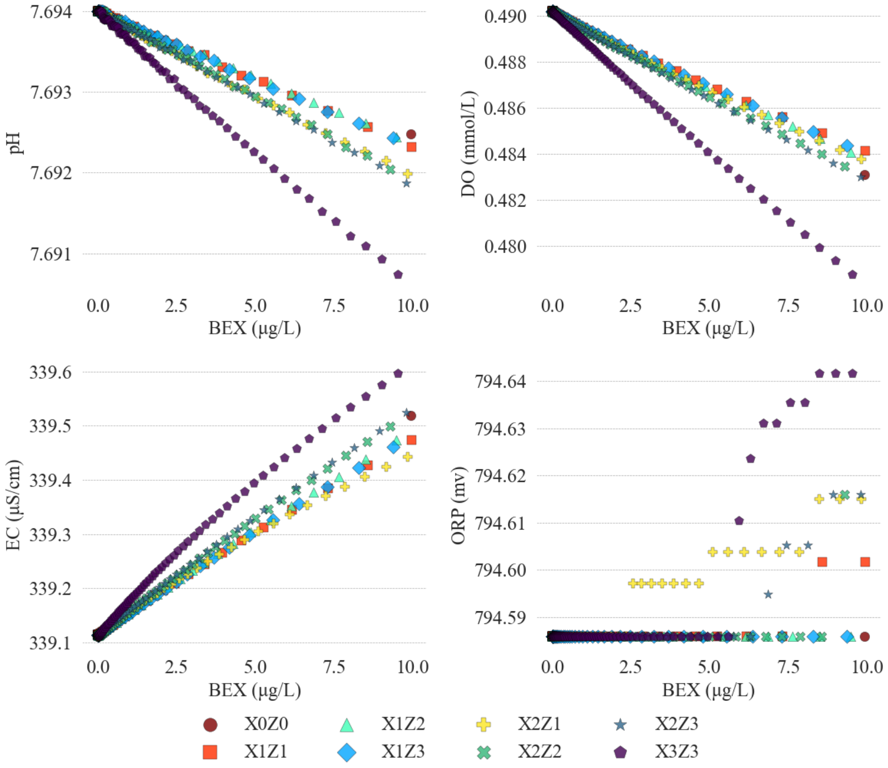


**Figure 4.3.** Scatter plots of four in-situ water quality parameters versus BEX concentration (µg/L) based on RTM results. Data points are plotted at 5-point intervals to reduce visual clutter while preserving trend resolution. Note that variations in sensor parameter values were minimal across the ND–10 µg/L range.

Furthermore, the iWQPs were responsive to changes in BEX due to degradation rather than to its absolute concentration. As a result, the models still tended to trigger contamination alarms earlier than the set thresholds even when the training threshold was increased from 5 µg/L to 10, 50, 100, and 500 µg/L (**Figure B4**).

RF and XGB performed best at all virtual warning wells. This can be attributed to their ensemble nature, which enables them to effectively capture complex, non-linear relationships in the data (Lin et al., 2022). Ensemble methods can also improve predictive performance and robustness against overfitting by combining multiple models or learners, allowing them to generalize better across different datasets.

Although LR and SVC are both linear models (with a linear kernel used for SVC), SVC was more conservative in its predictions, requiring a higher BEX concentration and thus greater certainty in exceeding the threshold before it triggered the alarm. The LR models the probability of a sample belonging to a class (i.e., contaminated vs. uncontaminated). However, it assumes a linear relationship between the independent variables and the log odds of the outcome, which may not be valid in all scenarios (Bisong, 2019). The training data may have caused the decision boundary to be close to the threshold, leading to more samples being classified as contaminated, even if their concentrations were below but near the threshold. On the other hand, SVC aims to find the hyperplane that maximizes the margin between the two classes (Vapnik, 1982). SVC focuses on samples closest to the decision boundary (support vectors), making it less sensitive to the overall data distribution, which results in the SVC being more conservative when classifying samples as contaminated.

While MLP can capture complex patterns due to its multi-layer architecture, it is also prone to overfitting (Rynkiewicz, 2019). MLP triggered an alarm at a BEX concentration of 1.2 µg/L at the virtual well X3Z3, suggesting that it did not generalize well to unseen data or underperformed with data that were outside the range of the training data. This lack of generalization is often a consequence of the model's reliance on specific training data characteristics that may not be representative of the unseen conditions.

Additionally, due to the stochastic initialization of weights in MLP, as in other neural networks (Narkhede et al., 2022), the BEX concentration that triggered the alarm varied between runs. To account for this variability, the prediction was simulated 1000 times, and the results were recorded. **Figure B5** shows the histogram of the BEX concentration when MLP triggered the alarm at the virtual warning wells. Most of the alarms were triggered at BEX concentration of approximately 3.3 µg/L to 3.8 µg/L for all virtual warning wells, except for the farthest virtual warning well X3Z3. At this virtual well, MLP was prone to false alarms, triggering alarms even when the BEX concentration was ND.

It is important to note that small variations in sensor data may be difficult to detect in real-life applications, and that the technical limitations of available sensors must be considered when interpreting the practical implications of these findings. Sensors typically have accuracies and resolutions of ±0.1 and ±0.01 for pH, ±5 mV and ±0.1 mV for ORP, ±0.5% (of reading) and ±0.1 µS/cm for EC, and ±0.1 mg/L and ±0.01 mg/L for DO (**Chapter 3**). Despite these limitations, results suggest that the combined response of multiple sensor parameters—even when individual deviations are small—still supported the proof-of-concept framework for an early warning system.

### 4.3.2. Scenario Simulations Performances

In this section, the results of the scenario simulations were presented, which compare the performance of ML classifiers under different aquifer conditions (**Figure 4.4**). Each scenario involves a specific modification to the aquifer's properties in the previous RTM in **Chapter 3**, including hydraulic conductivity, background EC, $NO_3^-$ and $SO_4^{2-}$ concentrations, presence/absence of calcite, and cation exchange capacity (CEC). As in the base case, the ML models were trained using data from virtual wells X0Z0, X1Z1, X1Z2, and X1Z3 within a given scenario and tested them on downstream virtual warning wells X2Z1, X2Z2, X2Z3, and X3Z3 under the same conditions.

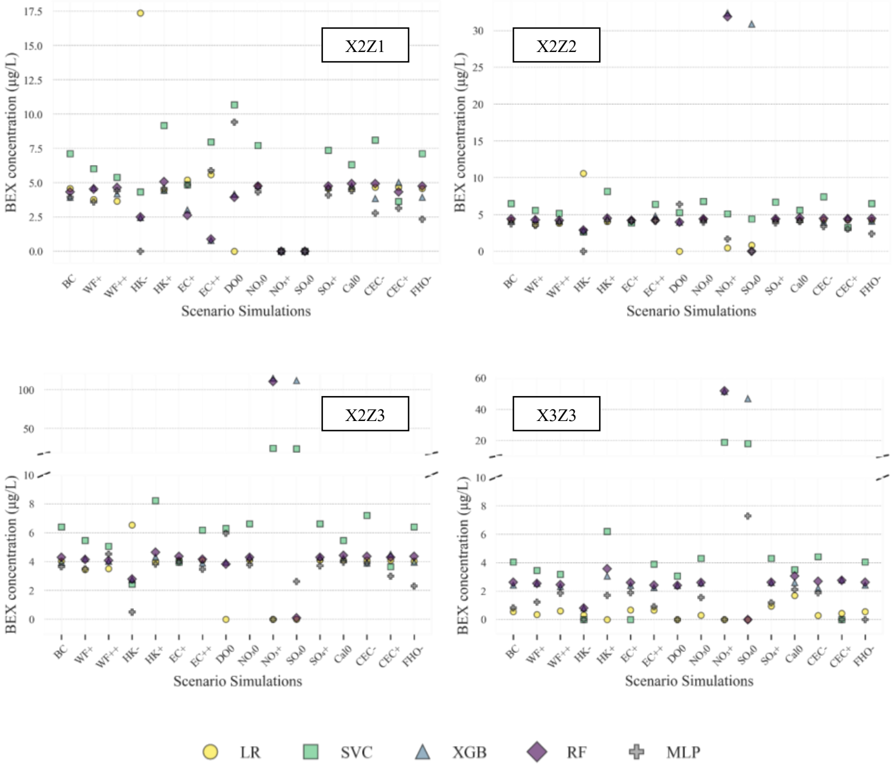


**Figure 4.4** Model prediction results showing the BEX concentration at which the contamination alarm was triggered at four virtual monitoring wells (i.e., X2Z1, X2Z2, X2Z3, X3Z3) under different scenario simulations. Broken y-axis was used to clearly display BEX concentrations between ND and 10 µg/L at X2Z3 and X3Z3.

Scenario simulations revealed consistent patterns in model performance relative to the base case. Most models triggered detections earlier than the actual 5 µg/L threshold, except for the SVC model, which showed a more conservative behavior. However, these general trends of the models' performances were significantly affected by three hydrogeochemical changes: reduced hydraulic conductivity, elevated background salinity, and the availability of electron acceptors.

Reduced hydraulic conductivity slowed down advective transport, increasing residence time and enhancing redox-driven degradation of BEX. This condition altered the geochemical gradients of the aquifer, weakening the iWQP-BEX correlations on which the ML models relied for detection. The LR model became particularly conservative under these conditions, triggering alarm at elevated concentration of 17.4 µg/L at virtual warning well X2Z1 (**Figure 4.4**). In contrast, the SVC model triggered the alarm at 4.3 µg/L in virtual warning well X2Z1 and falsely detecting contamination in virtual warning well X3Z3 (ND). These responses reflect how reduced flow velocities fundamentally shift chemical relationships in the aquifer, making parameters like pH and EC more responsive to NVDOC than to BEX compounds (**Chapter 3**).

Changes in background water chemistry also affected the performance of the ML models. The elevated EC, induced by increased salinity, reduced the signal-to-noise ratio of BEX contamination (**Chapter 3**). High baseline EC masked the subtle changes caused by BEX degradation, resulting in SVC to generate false positives even in uncontaminated virtual warning wells (ND at X3Z3).

Scenarios with elevated nitrate or absent oxygen/$SO_4^{2-}$ disrupted the expected BEX degradation pathways, causing the ML models to trigger false or delayed contamination alarms. For instance, LR incorrectly flagged contamination at all virtual warning wells (ND) when dissolved oxygen was absent at the RTM simulation. Similarly, both RF and XGB triggered alarms during the simulation scenario with elevated nitrate concentrations at BEX levels above 30 µg/L at the virtual warning wells X2Z2, X2Z3, and X3Z3 (**Figure 4.4**). This scenario likely shifted the redox balance, delayed BEX degradation, and thus delayed the alarm until the concentration was significantly higher than the 5 µg/L threshold.

The kernel density estimates in **Figure B6** illustrate these overall trends, showing that most model predictions clustered below the 5 µg/L threshold (i.e., at the peak of the curves). These KDE plots provide a smoothed visualization of the data distribution, offering intuitive insights into prediction patterns while masking exact frequencies (Węglarczyk, 2018).

While scenario simulations explored hydrogeochemical variability, they did not yet account for potential measurement uncertainty in real-world field deployments. To address this limitation and assess the robustness of the ML models under more practical conditions, Gaussian noise and sinusoidal fluctuations were introduced to the iWQPs.

### 4.3.3. Sensitivity to Noise and Groundwater Variability

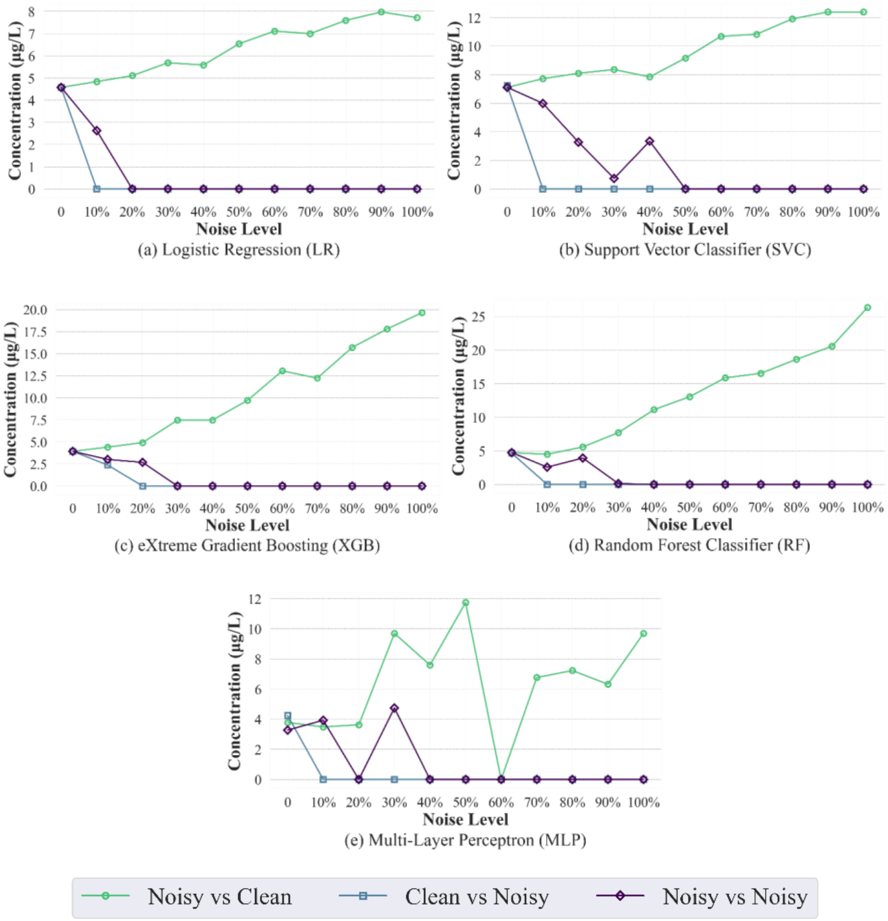
#### Impact of Sensor Noise on Model Performance

Adding Gaussian noise to the test data significantly degraded the performance of all ML models. Even at relatively low noise levels (10–20%), false alarms were consistently triggered, particularly at virtual warning well X2Z1 (**Figure 4.5**). This is presumably because typical sensor noise levels were comparable to the subtle signal variations of iWQPS at BEX concentration threshold of 5 µg/L. As the threshold increased, the impact of sensor noise diminished, since higher BEX levels were associated with stronger signals (**Figure B7**). False alarms occurred at 10–30% noise levels for the 10 µg/L threshold, while premature alarms—approximately 30 to 100 µg/L lower than the set threshold—were observed at 80–100% noise levels for thresholds at 50 to 500 µg/L (**Figure B8**). Among the tested models, MLP was the most sensitive, leading to inconsistent alarm triggering. This suggests a loss of generalization and stability under noisy conditions, highlighting the need for enhanced regularization techniques (Dey et al, 2018).

When noise was added to the training data, its effect on model performance was more gradual compared to test data noise. As noise levels increased from 10% to 100%, the BEX concentration at which the models triggered alarms also increased (**Figures B9a to B9c**), except for MLP. This trend aligns with findings from other studies, which have shown that the type and level of noise significantly affect classifier accuracy; for example, experiments with Gaussian noise demonstrated a more gradual decline in accuracy, particularly when noise was added to the training data (Schooltink, 2020).

However, the addition of training noise caused LR and SVC to generate false positives at the farthest virtual warning well X3Z3 (**Figure B9c**). This points to a possible interaction between noise and signal strength at the fringe of the plume, where BEX concentrations are already low. Under these conditions, even small fluctuations in input features were enough to push the model toward incorrect classifications. Furthermore, as noted by L. Schmidt et al. (2018), even small input perturbations that are often imperceptible to humans can cause state-of-the-art classifiers to make incorrect predictions with high confidence.

Compared to other models, LR demonstrated greater robustness. While noise introduced a slight delay in alarm triggering (from 5 µg/L to about 8 µg/L) at virtual warning well X2Z1 (**Figure 4.5**), LR maintained relatively stable performance across noise levels. This can be attributed to its linear decision boundary, which is inherently less sensitive to small perturbations in the input data, allowing it to better filter out minor fluctuations (Hasan and Chu, 2022).



(a) Logistic Regression (LR)

(b) Support Vector Classifier (SVC)

(c) eXtreme Gradient Boosting (XGB)

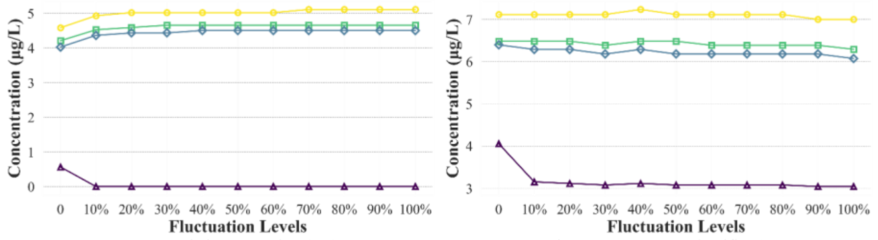
(d) Random Forest Classifier (RF)

(e) Multi-Layer Perceptron (MLP)

**Figure 4.5**. BEX concentrations (µg/L) at which the five ML models triggered the contamination alarm under varying sensor noise levels at virtual warning well X2Z1: Noisy vs Clean (noisy training, clean test data), Clean vs Noisy (clean training, noisy test data), and Noisy vs Noisy (noisy training and test data).

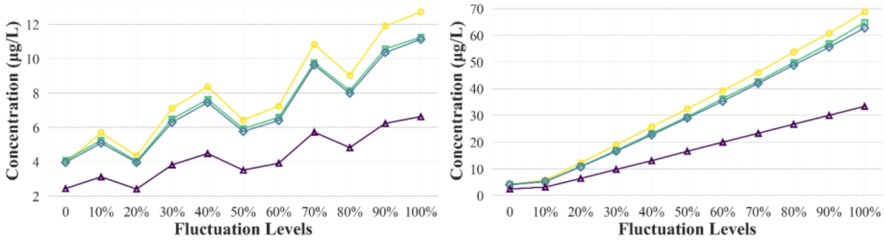
### Impact of Seasonal Fluctuation on Model Performance

Sinusoidal fluctuations in the training data had a limited impact on the performance of linear models such as LR and SVC, especially when compared to sensor noise **(Figure 4.6)**. These simpler models lack the flexibility to model non-linear patterns, which in this context proved advantageous. Because they could not overfit to the fluctuations, LR and SVC maintained stable alarm-triggering behavior across varying fluctuation levels.



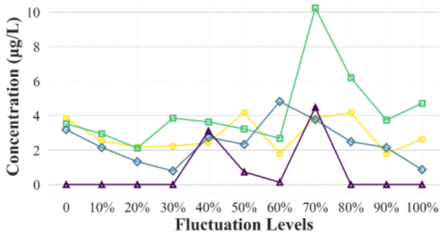
(a) Logistic Regression (LR)

(b) Support Vector Classifier (SVC)



(c) eXtreme Gradient Boosting (XGB)

(d) Random Forest Classifier (RF)



(e) Multi-Layer Perceptron (MLP)

**Figure 4.6**. BEX concentrations (μg/L) at which the five ML models triggered the contamination alarm under varying fluctuation levels added to the training data at all virtual warning wells.

In contrast, complex models such as XGB, RF, and MLP were more sensitive to these fluctuations. As the amplitude of the fluctuations increased, XGB and RF became increasingly conservative, delaying alarms until higher BEX concentrations were reached. RF only triggered alarms at concentrations exceeding 60 μg/L under extreme (100%) fluctuation conditions. The MLP model exhibited the most erratic behavior, likely due to the MLP’s complex internal representations and susceptibility to overfitting, particularly when the hyperparameters were tuned on clean data (Rynkiewicz, 2019).

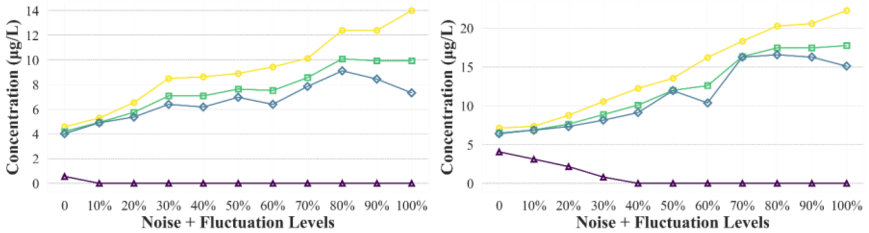
These findings suggest that fluctuations not representative of the true underlying signal can degrade the performance of complex models. Such models are prone to fitting high-frequency patterns in the training data, mistaking noise for meaningful trends (Hakkal and Lahcen, 2024). In contrast, simpler linear models such as LR and SVC cannot capture complex relationships such as these fluctuations. However, if the fluctuations were part of the actual signal, the inability of linear models to learn non-linear patterns would likely degrade their performance.

To account for possible seasonal trends in iWQPs, additional experiments were also conducted by shifting the sinusoidal peaks to represent different months. These seasonal adjustments did not significantly alter model behavior, and the general conclusions described above held across all scenarios.

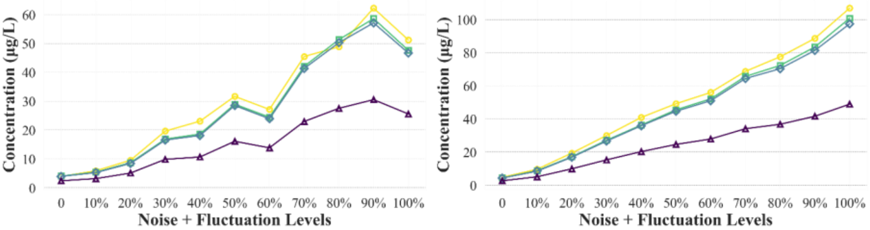
### Influence of Combined Noise and Seasonal Fluctuation

When both sensor noise and sinusoidal annual fluctuations were introduced into the training data, model performance degraded more noticeably than when only one type of variability was present (**Figure 4.7**). For instance, at 100% combined noise and fluctuation levels, LR delayed its alarm at virtual warning well X2Z1. The alarm was triggered only until BEX concentrations reached 14 µg/L which is well above the intended 5 µg/L threshold. In comparison, LR triggered the alarm at around 8 µg/L with only sensor noise, and correctly at 5 µg/L with only sinusoidal fluctuations. Similarly, XGB triggered the alarm at X2Z1 at 50 µg/L under combined noise and fluctuations, compared to 20 µg/L with only sensor noise and 13 µg/L with only sinusoidal fluctuations.

Interestingly, model performance did not always degrade linearly with increasing noise levels. In some cases, such as XGB at the four virtual warning wells, the model performed slightly better at 100% combined noise and fluctuation than at 90%. At 90%, the alarm was triggered at approximately 60 µg/L, compared to around 50 µg/L at 100%. This counterintuitive result may be due to instances where the sensor noise and the sinusoidal fluctuations partially canceled each other out.



(a) Logistic Regression (LR) (b) Support Vector Classifier (SVC)



(c) eXtreme Gradient Boosting (XGB) (d) Random Forest Classifier (RF)

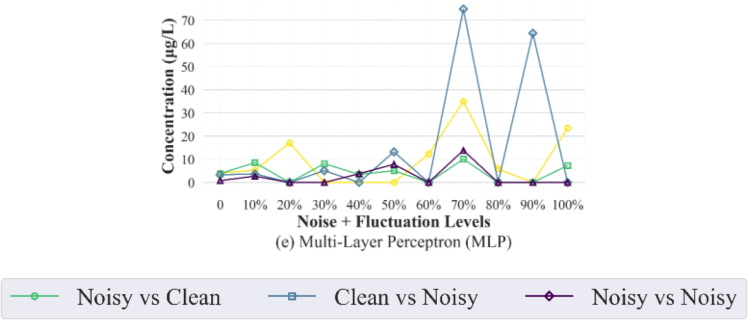
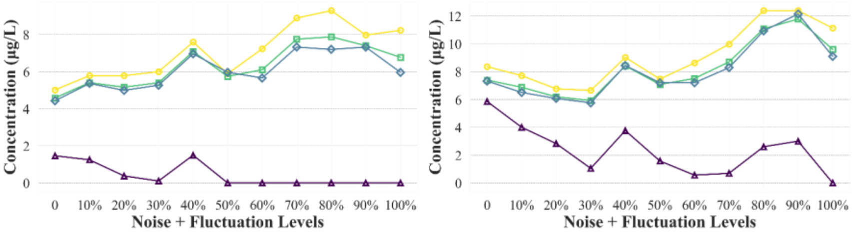


**Figure 4.7**. BEX concentrations (µg/L) at which the five ML models triggered the contamination alarm with combined sensor noise and sinusoidal fluctuation levels added to the training data at all virtual warning wells.

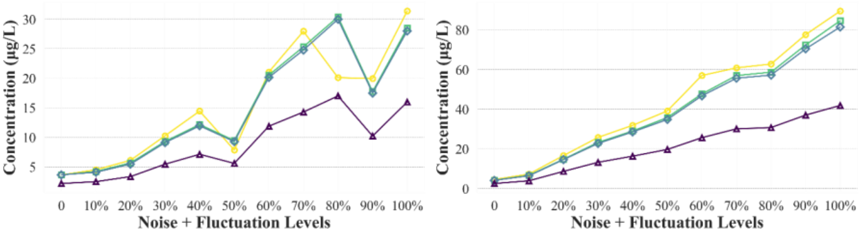
### 4.3.4. Improving Model Performance

To mitigate the effects of sensor noise and seasonal fluctuations, 3-day, 5-day, and 10-day moving average smoothing were applied to the training data after introducing Gaussian noise and sinusoidal fluctuations. Among these, the 5-day moving average yielded the best model performance (**Figure 4.8**). The results for 3-day and 10-day smoothing are provided in **Figure B10a** and **Figure B10b**.



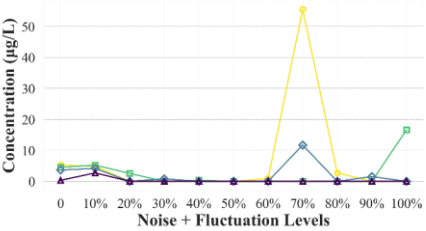
(a) Logistic Regression (LR)

(b) Support Vector Classifier (SVC)



(c) eXtreme Gradient Boosting (XGB)

(d) Random Forest Classifier (RF)



(e) Multi-Layer Perceptron (MLP)

**Figure 4.8**. BEX concentrations (µg/L) at which the five machine learning models triggered contamination alarms. A 5-day moving average was applied for smoothing after adding different levels of combined sensor noise and sinusoidal fluctuations in the training data at all virtual warning wells.

LR initially triggered the contamination alarm at ~10 µg/L at 100% noise and fluctuation levels in virtual warning well X2Z2 (**Figure 4.7**). However, LR detected BEX contamination at ~7 µg/L when 5-day smoothing was applied (**Figure 4.8**), which is closer to the 5 µg/L regulatory threshold. Similarly, SVC showed improved sensitivity, with alarms triggered at ~10 µg/L (smoothed) compared to ~17 µg/L (unsmoothed).

While smoothing can generally help reduce the impact of sensor noise in the ML model performance (Xiao et al., 2022), the choice of window size is critical: Too few days (e.g., 3-day) makes smoothing overly sensitive to noise, as short-term fluctuations disproportionately influence the average. On the other hand, too many days (e.g., 10-day) leads to over-smoothing and can delay the detection by masking short-term contamination spikes. The 5-day window in this case served as an optimal balance, effectively dampening random noise while still preserving meaningful sensor signals. However, seasonal fluctuations require a different approach such as Fourier decomposition or wavelet transforms (Bi et al., 2023) to separate periodic trends from contamination signals.

## 4.4 Conclusion

In this study, a proof-of-concept ML framework that integrates data from affordable in-situ sensors was presented to detect BEX in groundwater. The used virtual sensors in the presented case include pH, DO, EC, and ORP. The ML framework was trained to detect contamination at warning wells and provide timely alerts of contaminant migration before it reaches sensitive receptor areas.

Ensemble models such as RF and XGB consistently demonstrated reliable performance across diverse aquifer scenarios. This robustness suggests their potential suitability for future field deployment after further field validation. However, successful real-world implementation still requires careful consideration of site-specific hydrogeochemical properties. Scenario-specific deviations consistently emerged under three conditions: when altered transport times changed the degradation signatures, when background chemistry obscured the contamination signals, or when terminal electron acceptors were absent. These factors affect the relationships between iWQPs and BEX concentrations.

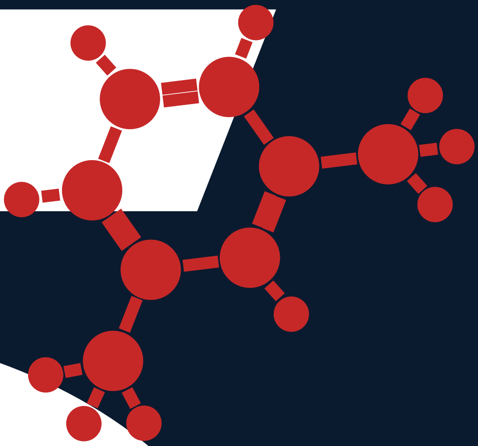
The integration of sensor data into ML systems raises challenges related to data quality and signal stability. Even modest levels of noise, such as 10–20% Gaussian variation, considerably impacted the model performance, especially for complex models such as the MLP. This highlights the need for both hardware and software-based solutions to stabilize data inputs. For example, hardware modifications such as flow-stabilization chambers in monitoring wells can help reduce turbulence and provide more consistent sensor readings. On the software side, preprocessing techniques to remove seasonal patterns and smooth noise can also enhance data quality. Fixed-window smoothing can improve model reliability; however, static and adaptive smoothing methods may mask short-term concentration spikes or fail to adjust to seasonal variability.

To maintain model accuracy over time, all deployed systems will require continuous learning mechanisms capable of incorporating new sensor data and adapting to evolving aquifer conditions. Without such updates, model predictions may degrade as site conditions shift. Moreover, since class imbalance can affect model performance, resampling or weighting techniques can be explored for future work to improve contaminant detection sensitivity. While this study demonstrates the feasibility of this approach in a controlled virtual environment, further field-based research is needed to validate its practical applicability.

Importantly, this framework is intended to complement, not replace, existing manual sampling strategies. Nevertheless, the integration of low-cost sensor networks, adaptive ML models, and robust validation strategies offers a promising path toward real-time, scalable, and continuous groundwater quality monitoring.

# Real-time Monitoring of Petroleum Hydrocarbons in Groundwater using Hybrid Machine Learning Architectures

## 5



## Abstract

Monitoring benzene, toluene, ethylbenzene, and xylenes (BTEX) in groundwater is essential for managing petroleum hydrocarbon (PHC) plumes but is often hindered by high costs. A machine learning (ML) framework was evaluated in terms of estimating concentrations of benzene, ethylbenzene, and xylenes (BEX), using affordable, in-situ water quality parameters (iWQPs) as inputs: pH, dissolved oxygen, electrical conductivity, and oxidation-reduction potential. Due to a scarcity of field data, the models were trained and tested on high-resolution virtual data generated by a reactive transport model. A Long Short-Term Memory (LSTM) network was compared against classical algorithms (multiple linear regression, random forest, support vector regression, XGBoost) and an LSTM-XGBoost hybrid. A key finding was that model performance critically depended on the underlying geochemical relationship between iWQPs and BEX. Accurate predictions ($R^2 \geq 0.80$, MAPE < 2.3%) were achieved when iWQPs were strongly correlated with BEX degradation (e.g., as a primary electron donor). Performance declined sharply ($R^2 < 0$) during periods where iWQPs were correlated with non-volatile dissolved organic carbon, which is another component of dissolved PHC. Incorporating hydraulic head data improved accuracy by informing the model of groundwater flow dynamics. While the LSTM model struggled to extrapolate beyond its training data (e.g., during extreme flow events), it reliably detected the direction of concentration trends, providing a valuable trigger for adaptive monitoring. It was also demonstrated how a hybrid Kalman filter could successfully capture concentration trends after source removal through recursive updating. The proposed ML strategy provides a feasible initial framework for estimating BEX concentrations. Based solely on iWQPs, the model yielded predictions within a 5% error margin for 70% of the test cases, and performance was substantially improved by incorporating hydraulic head data. This makes ML models a viable, cost-effective tool for improving groundwater monitoring, though their success is dependent on site-specific and time-varying conditions. Future work should validate the proposed strategy with high-resolution field data.

This chapter has been submitted for publication as:

Wu, C.L.R., Wagterveld, R.M., Rietveld, L.C., van Breukelen, B.M. Real-time Monitoring of Petroleum Hydrocarbons in Groundwater using Hybrid Machine Learning Architectures.

The repository associated with this chapter can be accessed at:

https://doi.org/10.4121/0a23147e-ba85-4ba2-a058-ba199c65d711

## 5.1 Introduction

Groundwater contamination by petroleum hydrocarbons (PHCs) remains an environmental challenge, particularly in areas affected by historical spills or leaks. Compounds such as benzene, toluene, ethylbenzene, and xylenes (BTEX) pose serious risks to ecosystems and human health due to their toxicity and persistence (Zanello et al., 2021). Monitored natural attenuation (MNA) is a widely adopted remediation strategy that relies on natural processes such as biodegradation, dilution, and sorption to reduce contaminant concentrations over time (Chiu et al., 2013). However, the effectiveness of MNA depends on monitoring contaminant levels accurately and continuously to ensure that attenuation is proceeding as expected (Beck and Mann, 2010). Traditional monitoring methods, which rely on periodic manual sampling and laboratory analysis, are time-consuming, costly, and spatially limited, often resulting in data gaps that hinder timely decision-making.

Existing technologies for continuous BTEX monitoring face limitations. In-situ optical sensors provide near-real-time data but are susceptible to interference from other dissolved organics and require frequent calibration (Buerck et al., 2001; Larsson and Dasgupta, 2003, Wong et al., 2024). Membrane interface probes offer high-resolution vertical contaminant profiling but are invasive and expensive, making them impractical for long-term, large-scale monitoring (Industrial Economics, Incorporated, 2023). Portable gas chromatography systems, while accurate, require specialized operators and regular maintenance (Ji et al., 2006). Even with these technologies, a gap remains in developing scalable, low-maintenance solutions for real-time monitoring of BTEX contamination in groundwater.

Machine learning (ML) has emerged as a promising alternative by leveraging easily measurable in-situ water quality parameters (iWQPs) to infer contaminant concentrations. As shown in **Chapter 3**, pH, dissolved oxygen (DO), electrical conductivity (EC), and oxidation-reduction potential (ORP), are correlated with dissolved PHCs. The RTM was used as the data source due to the lack of an available dataset on groundwater-contaminated sites for a robust statistical analysis. Building on this, a binary classification model was developed in **Chapter 4** to flag benzene, ethylbenzene, and xylenes (BEX) contamination for early warning systems.

Qiao et al. (2025) also proposed a two-stage predictive framework for non-aqueous phase liquid (NAPL) plumes, using the RTM in **Chapter 3** to generate 30 years of simulated daily data. They used year 30 as the prediction target and the preceding years as training data. The training and test datasets were generated for each of the 7,500 grid cells (150 × 50), resulting in thousands of synthetic data observations. In the first stage, a sliding-window random forest (RF) algorithm predicted iWQPs for the target year. In the second stage, these forecasted parameters were then combined with sparse NAPL measurements. ML models, including long short-term memory (LSTM) networks and extreme gradient boosting (XGBoost), estimated the future NAPL plume extent and concentration distributions in the simulated 2D cross-section.

Despite these advances, current ML applications for hydrocarbon monitoring have limitations. While existing frameworks can detect PHC presence and predict LNAPL plume extent based on historical virtual data, they lack real-time capability for estimating PHC concentrations in groundwater. Moreover, these models are often poorly suited to dynamic field conditions, such as active remediation or sudden shifts in hydraulic gradients. This stems from their feed-forward design, which processes data linearly without recursive updates from new measurements. These gaps illustrate the need for more responsive systems capable of continuous operation under variable site conditions.

Therefore, several ML regression models that address these limitations were evaluated by estimating BEX concentrations in real time through sensor data fusion (SDF). SDF involves integrating multiple sensor data to indirectly estimate unknown parameters that are otherwise difficult or costly to measure (Mitchell, 2007). The research objective was to quantify the prediction accuracy of ML models in estimating BEX concentrations using iWQPs from cheap sensors (i.e., pH, DO, EC, and ORP) under controlled conditions. Due to a lack of high-resolution temporal and spatial field data from contaminated sites, virtual data generated from a previous RTM (**Chapter 3**) were used.

To capture temporal dependencies, LSTM networks with variable sequence lengths (30-360 days) were implemented. The sequence length determines how many successive daily data points are considered when making each prediction. This architecture was selected for its demonstrated effectiveness in modeling environmental time-series data (Arsenault et al., 2023). The main advantage of LSTM networks is their ability to learn long-range dependencies while overcoming the vanishing gradient problem inherent in simpler recurrent neural networks. A hybrid architecture that combines LSTM-based sequence modeling with XGBoost regression was also developed. This configuration takes advantage of the LSTM's temporal pattern recognition capabilities alongside XGBoost's strengths in modeling nonlinear relationships and feature importance weighting (Karimi et al., 2025).

For comparison, traditional regression models were also implemented for their interpretability and established use in water quality monitoring (Singh et al., 2021): multiple linear regression (MLR) to provide a benchmark for linear relationships (Sani Gaya et al., 2020); support vector regression (SVR) to capture nonlinear patterns (Banadkooki et al., 2020); and RF and XGBoost to handle complex feature interactions (Szomolányi and Clement, 2023). While these models can represent temporal dynamics when supplied with engineered features (e.g., lagged variables), in this study they were applied using daily iWQP measurements as independent observations. This choice allows for a focused evaluation of the extent to which explicitly modeled temporal dependencies, such as LSTM-based approaches, improve predictive performance relative to models without embedded sequence learning. However, the primary goal is to assess the feasibility of ML approaches for BEX monitoring, rather than strict model comparison.

To enhance the reliability of BEX concentration estimates, the use of Kalman filtering was also explored. The Kalman filter (KF) is a recursive algorithm that integrates noisy sensor measurements over time to produce optimal state estimates (Simon, 2006). It is particularly well-suited for dynamic systems where measurements are uncertain and temporally correlated, making it a valuable addition to the proposed SDF framework. By incorporating Kalman filtering, BEX concentration estimates could be continuously refined as new iWQP data and BEX measurements became available.

The robustness of these models was evaluated through stress-testing scenarios, including source removal and increased hydraulic gradient. To further improve model accuracy, hydraulic head was incorporated as an additional input parameter. This parameter contains information on the advective transport and residence times, both of which are critical for capturing the influence of flow dynamics on solute distribution and reactive processes.

## 5.2 Materials and Methods

### 5.2.1. Data Generation

The RTM used in this study is based on the model developed in **Chapter 3**, implemented in Python using FloPy within a Jupyter Notebook environment. Groundwater flow was simulated using MODFLOW 2005, and contaminant transport was modeled with MT3DMS, incorporating advection and dispersion processes. Geochemical reactions were coupled through PHT3D and PHREEQC-2 to account for biodegradation, mineral-phase reactions, cation exchange, and outgassing. The model also incorporated transient water table fluctuations and spatial heterogeneity in hydraulic conductivity. The synthetic aquifer domain represents a saturated porous medium with a length of 300 m and a thickness of 10 m, inspired by an existing study on the Bemidji crude oil spill site (Ng et al., 2015). It was discretized into 0.2 m (vertical) × 2 m (horizontal) grid cells (150 × 50 grids).

To evaluate the robustness of the ML models under realistic stress-test conditions, three scenarios were simulated: (1) a base case scenario representing general conditions at the Bemidji site, USA, but with more generic parameterization; (2) a realistic remediation case in which the light non-aqueous phase liquid (LNAPL) source zone is fully excavated; and (3) a controlled hydraulic stress to represent a potential extreme anthropogenic influence, such as the use of an injection well for artificial aquifer recharge upstream or an excessive downstream pumping to simulate accelerated groundwater flow.

The second scenario involved completely suppressing PHC dissolution from the oil source zone in simulation year 40. This setup is particularly relevant for evaluating post-remediation monitoring strategies and plume tailing behavior, and to test whether the ML model can still accurately estimate BEX concentrations. The third scenario involved an injection well that continuously injected water for one year, starting in simulation year 40. The injected water, simulated in a larger groundwater flow model, increased the hydraulic gradient by approximately 10 m above the initial conditions. This resulting gradient was used as the boundary condition for the RTM simulation. The third scenario was designed to evaluate the ML model’s performance under extreme groundwater flow velocity. Although this scenario was simulated via upstream artificial recharge, extreme flow conditions could also result from excessive downstream pumping or extreme natural groundwater table fluctuations; these conditions can accelerate contaminant advection and dispersion (Ahmadi et al., 2021).

The third scenario was designed to evaluate the ML model's performance under conditions of extreme groundwater flow velocity, induced through upstream artificial recharge. Such extreme hydraulic gradients, which can also arise from excessive downstream pumping or natural water table fluctuations, are critical to study as they promote accelerated contaminant advection and dispersion (Ahmadi et al., 2021).

Time-series data from a 100-year simulation with daily timesteps were analyzed, capturing the long-term persistence of PHCs in groundwater systems. Eight virtual observation wells were placed at depths ranging from 1.5 to 5.2 m and distances of 0 to 165 m from the source zone (**Table B1**) to monitor plume dynamics (**Figure 5.1**). Each well was modeled with a 1-meter screened interval. Groundwater chemistry was calculated as a flow-rate-weighted average concentration; the weights were proportional to the hydraulic conductivity of each RTM grid cell (Höyng et al., 2015; Thouement & Van Breukelen, 2020) (**eqn 3.7** in **Chapter 3**).

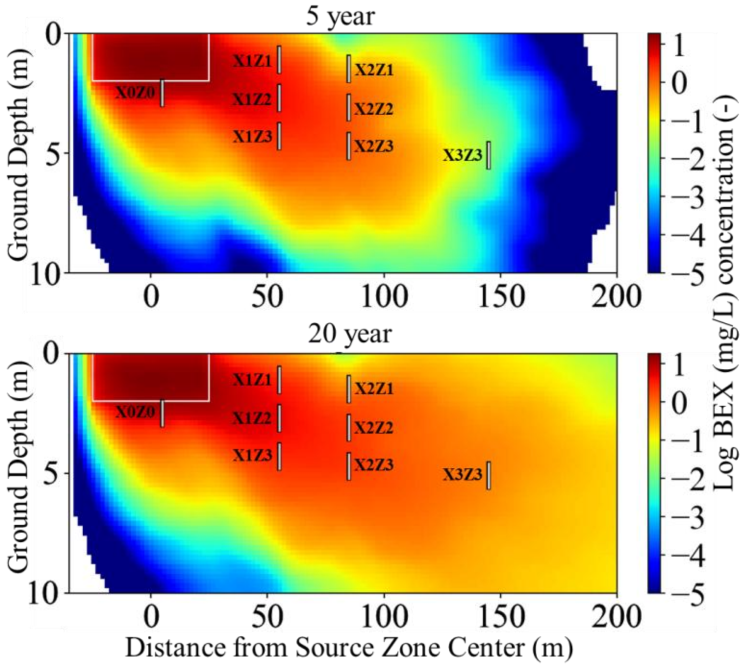


**Figure 5.1**. BEX plume distribution at 5 and 20 years of simulation. BEX concentration is converted to base 10 log form with 1 mg/L as reference. The white square indicates the oil source zone, and the eight vertical white lines represent virtual observation wells.

### 5.2.2. Data Preprocessing, Handling, and Modeling Approach

The goal in this research is to predict daily BEX concentrations using iWQPs at eight virtual observation wells. Hereafter, all observation wells and time-series data correspond to virtual datasets obtained from the RTM. For brevity, these were referred to as observation wells and time-series data throughout, without explicitly repeating the qualifier “virtual”. All analyses used daily time-series data, with input features standardized using Scikit-learn's StandardScaler (Ahsan et al., 2021) to ensure consistent scaling, while BEX concentrations served as the target variable. Separate models were developed for each observation well to account for spatial variability across the aquifer and allow localized calibration and performance assessment.

In a subsequent modeling phase, it was evaluated whether incorporating hydraulic head measurements would improve the predictive accuracy of the ML model. The average hydraulic head in an observation well was analyzed together with the four iWQPs to assess its contribution to model performance.

To account for seasonal variability while ensuring model adaptability, a rolling five-year training window approach was employed. Models were trained on daily data from each five-year window and subsequently tested on the following year's data (**Figure 5.2**). The window was then advanced by one year, with the model retrained on the new five-year segment and tested on the subsequent year. This one-year interval was chosen to capture a full hydrological cycle for a robust performance assessment. While more frequent retraining is possible in practice, the annual update provides a computationally efficient and conservative benchmark for evaluating the framework. If the ML model performs well with annual updates, it establishes a baseline from which it can be inferred that more frequent updates would perform at least as well as, if not better than, yearly updates. This approach enables performance evaluation across different time periods.

**Figure 5.2**. Illustration of the rolling window approach for model training and testing. The model is trained on a fixed-size window of past observations, which progressively moves forward through the dataset over time. The model is tested on the immediately succeeding data points to evaluate its performance (Modified from Amat Rodrigo and Escobar Ortiz, 2024).

The five-year training window duration was selected to balance model stability and responsiveness. While longer windows provide more data for learning complex relationships and reduce overfitting to short-term fluctuations, they may incorporate outdated patterns. By advancing the window yearly, the model adapts to recent data while maintaining sufficient historical context for robust generalization.

### 5.2.3. Regression Model Architectures and Hyperparameter Tuning

An LSTM network (**Text C1**) was implemented to predict BEX concentrations from timeseries iWQP data (i.e., pH, DO, EC, ORP). This architecture was selected for its ability to learn temporal dependencies directly from sequential input data, without requiring explicit feature engineering such as lagged variables and summary statistics (Zhang et al., 2023). Unlike models that operate on fixed input representations, LSTMs maintain an internal memory state through their gated architecture (input, forget, and output gates) (Hochreiter and Schmidhuber, 1997), as shown in **Appendix C** (**Figure C1**). The gates allow the LSTM to process data sequentially and selectively retain relevant information over time.

An important aspect of the LSTM is its sequence-based input. The hyperparameter governing this is the sequence length, which defines the number of consecutive daily observations used to make a single prediction. Hyperparameters are configuration settings for an ML algorithm that are established prior to the training process and govern how the model learns from data. Unlike internal model parameters that are learned during training, hyperparameters must be predefined and significantly impact predictive performance (Wu et al., 2019).

The LSTM model was trained on overlapping sequences to predict the next value in the sequence. For example, if the sequence length is 30, the model uses data from 30 consecutive days to predict the value for the next day. As shown in **Figure 5.2**, for each five-year training window, all possible overlapping sequences of the chosen length were generated. For instance, with a 30-day sequence length, the first training sample consists of days 1–30 of iWQP values, and the target output is the BEX concentration on day 31. The next sample uses days 2–31 to predict day 32, and so on, until the entire five-year period is covered. Four sequence lengths (i.e., 30, 90, 180, and 360 days) were tested to capture different temporal contexts. Shorter sequences (30 days) emphasize recent changes, while longer sequences (360 days) incorporate seasonal and long-term trends. These values provide a balance between capturing relevant patterns and maintaining computational feasibility. Furthermore, prediction uncertainty for the LSTM using Monte Carlo dropout to generate 95% confidence intervals for the estimates were quantified.

For comparative assessment, four classical ML models trained on the same 5-year rolling windows were implemented. These models were trained on single-day input, thus, the iWQPs from a single day were used to predict that day's BEX value. MLR served as a simple, interpretable baseline that captures linear relationships without tunable hyperparameters. SVR, RF, and XGBoost were also implemented to handle non-linear relationships. To optimize the performance of these non-linear models, the ML models' hyperparameters (**Text C2**) were fine-tuned using Tree-structured Parzen Estimator (TPE) from the Hyperopt library, a Bayesian optimization method (Bergstra et al., 2013). TPE efficiently navigates the hyperparameter space by building a probabilistic model from previous trials to intelligently select the most promising configurations for subsequent evaluation, thus finding high-performing settings with fewer evaluations than exhaustive methods.

An LSTM–XGBoost hybrid model was also tested that combined the strengths of temporal feature learning with robust feature-based decision making. In this architecture, the LSTM component was first trained on sequential iWQP data to learn latent temporal features. A feature extractor was then used to obtain fixed-length representations from the LSTM for each input sequence. These representations were subsequently used as inputs to an XGBoost regression model, which was trained separately.

To prevent data leakage, the dataset was partitioned prior to model training, and the LSTM was trained exclusively on the training data. Feature extraction for both training and test sets was performed independently using their respective inputs, and the XGBoost model was trained only on features derived from the training set. For clarity, the standalone LSTM has a different architecture than the LSTM used for the hybrid model.

All modeling was conducted in Python. Scikit-learn was employed for the regression models (Pedregosa et al., 2011), TensorFlow and Keras for the LSTM (Abadi et al., 2016; Chollet et al., 2015), Hyperopt for tuning (Bergstra et al., 2013), Pandas for data manipulation (McKinney, 2010), NumPy for numerical computing (Harris et al., 2020), and Matplotlib for visualization (Caswell et al., 2020).

#### 5.2.4. Hybrid Kalman Filtering for Source Removal Scenario

To improve model responsiveness during abrupt system changes, a hybrid KF (**Text C3**, **Figure C2**) was implemented specifically for the source removal scenario. This scenario simulates the complete excavation of the LNAPL source zone at year 40, resulting in a sudden drop in PHC dissolution. The KF recursively updated BEX concentration estimates by integrating incoming iWQP measurements with direct BEX measurements (from RTM output), enabling real-time adaptation to these transient conditions.

KF operates through two primary steps: prediction and update. In the prediction step, the state vector is projected forward in time using a state transition model. This model is defined by the state transition matrix F and contains the system's variables of interest. Thus, it predicts the BEX concentration at the next time step based on its current value and the system's dynamics. This step also incorporates the inherent uncertainty of the model itself, characterized as process noise with covariance matrix Q (Simon, 2006).

For the subsequent update step, this prediction is corrected using new observations typically obtained from sensors or laboratory measurements. An observation model maps the state vector to the measurement domain and is defined by the observation matrix. The Kalman gain is then computed to optimally balance the uncertainty between the model's prediction and the new measurement by weighting their respective covariances. The measurement noise, representing sensor inaccuracy, is defined by its covariance matrix. Both the process and measurement noise are assumed to be Gaussian.

In the KF implementation, BEX concentration and effective attenuation rate were considered as the state vectors in the prediction step. A first-order kinetic model was assumed, which is governed by the differential equation,

$$dC_{bex}/dt = -\lambda C_{bex},$$

where $C_{bex}$ is the BEX concentration, and λ is the effective attenuation rate constant. It was assumed that λ approximates the combined effects of advective transport and biodegradation after the LNAPL source was removed.

For the update step, an ensemble approach combining SVR with radial basis and polynomial kernels, and RF regression were employed. These models were trained to estimate BEX concentrations from iWQPs when direct BEX measurements were unavailable. The selection of these ML models aimed to balance nonlinear modeling capability with methodological simplicity, as the KF component was intended as a proof-of-concept rather than a comprehensive model comparison.

The hybrid KF model performance was evaluated using a rolling window validation strategy, with five-year training periods followed by 1-year testing windows. During testing, it was assumed that BEX measurements were only available yearly. Thus, the BEX concentration data was masked to simulate yearly sampling frequency, while iWQPs remained available daily.

### 5.2.5. Performance Metrics

The predictive performance of machine learning models was evaluated by comparing the estimated BEX concentrations against values simulated by the RTM, considered as the ground truth. Standard metrics were used for evaluation: mean absolute error (MAE), representing the arithmetic mean of absolute differences between predicted and actual values; coefficient of determination ($R^2$), indicating the proportion of variance in the observed data explained by the model; and mean absolute percentage error (MAPE), providing normalized error measurement independent of concentration magnitude.

## 5.3 Results and Discussion

In this research, the aim is to estimate daily BEX concentrations at observation wells using iWQPs by training separate LSTM models for each well and each test window.

The test window was defined as a one-year period following the 5-year training data, which was used to evaluate model performance on unseen data (i.e., the first test window corresponds to year 6 of the RTM simulation after an oil spill).

### 5.3.1. Sequence Length Selection

A 30-day sequence length for the LSTM model consistently produced the highest number of test windows with accurate estimates of BEX concentration from iWQPs ($R^2 \geq 0.80$), across all observation wells when compared to longer sequence lengths. For instance, at well X0Z0, 34 out of 94 test windows had $R^2 \geq 0.80$ using the 30-day sequence (green circles in **Figure 5.3**), compared to only 11, 32, and 21 test windows with 90-, 180-, and 360-day sequence lengths, respectively. While the 30-day model did not estimate BEX concentration accurately at all test windows, its better relative performance indicates that a month of historical iWQPs data (in this case, the virtual training data) provided the most effective temporal pattern for BEX estimation at the current timestep.

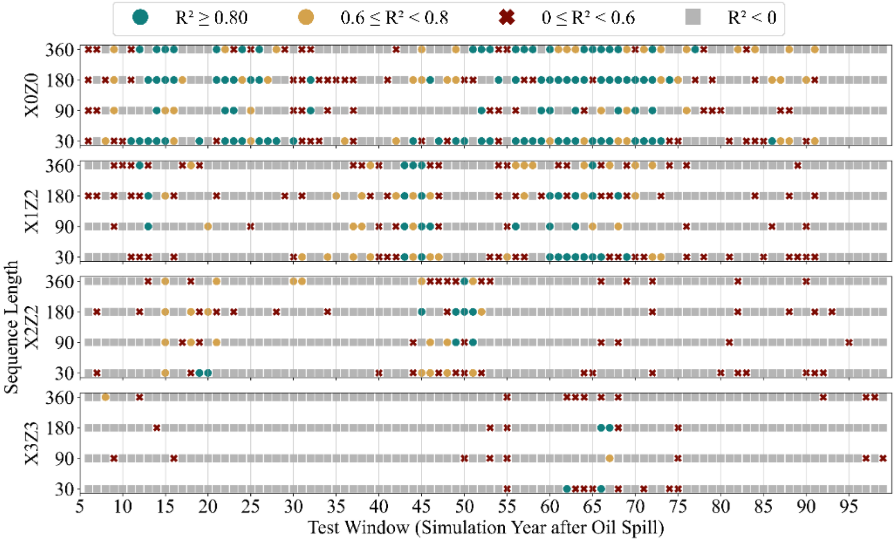


**Figure 5.3**. LSTM model predictions for observation wells X0Z0, X1Z2, X2Z2, and X3Z3 across time windows and sequence lengths of 30, 90, 180, and 360 days. Predictions with $R^2 \geq 0.80$ are shown as green circles, those with $0.60 \leq R^2 < 0.80$ as golden circles, and those with $0 \leq R^2 < 0.60$ as red crosses. Negative $R^2$ values were set to 0 and are shown as grey squares.

While the high $R^2$ value of 0.80 was not met across all test windows, the model demonstrated a potential for practical application in general contaminant monitoring; more than 70% of test windows at each observation well (68 out of 94) had a MAPE value below 5%, while over 87% (82 out of 94 test windows) had less than 10%

MAPE value at each observation well (**Figure 5.4**). A 5% error is widely considered as an acceptable margin of error for general monitoring of groundwater quality (Hounslow, 1995). However, this margin of error is based on field applications and should be interpreted with caution in the context of the idealized RTM-based dataset used in this study.

Consequently, $R^2$ values within the 0.6 to 0.8 range may still indicate useful predictive performance for monitoring purposes. This is especially true when the objective is to capture general contamination trend rather than exact concentration. For field applications, the model performance could be further improved through strategies such as more frequent retraining cycles and by limiting prediction horizons to shorter timeframes (e.g., one month instead of one year).

Furthermore, **Figure 5.3** shows suboptimal results as indicated by the predominance of red crosses and grey squares ($R^2 < 0.60$). At well X3Z3, for example, 98% of test windows had $R^2 < 0.80$. However, this scenario served as a baseline; its main purpose was to illustrate the model's limitations before introducing additional input parameters that could improve performance, as presented in the subsequent sections.

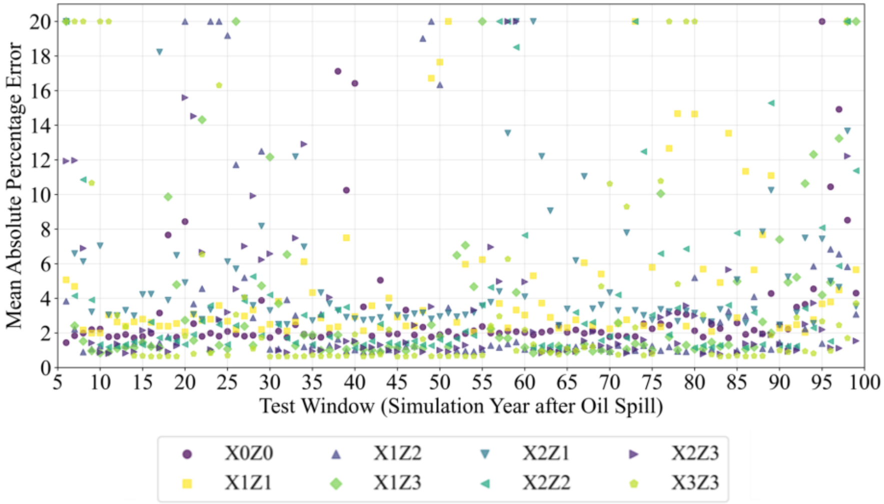


**Figure 5.4**. MAPE values for LSTM-predicted BEX concentrations across observation wells. Values above 20% were capped at 20% for visualization.

Results for other wells at different distances and depths relative to the source zone are included in **Figure C3** and **Figure C4**. The exception was observation well X2Z2, where four test windows had $R^2 \geq 0.80$ for the 180-day sequence and only two test windows for the 30-day sequence. Although the improved performance with a longer sequence could be due to its ability to capture hydrogeochemical processes, such as plume migration or degradation rates that are specific to this well's location and depth, the inconsistency across only a few test windows is more likely attributable to stochastic variations in model weight initialization during the LSTM's training phase than to a fundamentally different hydrogeochemical condition.

### 5.3.2. Baseline Model Performance in Well X0Z0

**Figure 5.5** illustrates the different performances of the LSTM model at well X0Z0 by comparing estimated versus actual (RTM-output) BEX concentrations across three representative test windows. These examples capture the range of observed outcomes: poor performance ($R^2 \approx -28.6$, MAE ≈ 567 µg/L, MAPE ≈ 5.85%), good performance ($R^2 \approx 0.80$, MAE ≈ 71 µg/L, MAPE ≈ 1.84%), and excellent performance ($R^2 \approx 0.95$, MAE ≈ 32 µg/L, MAPE ≈ 1.96%). In this study, focus was given on $R^2$ when comparing model performance due to its interpretability and more truthful nature (Chicco et al., 2021), while acknowledging that MAE and MAPE also provide complementary insights into the model's accuracy.

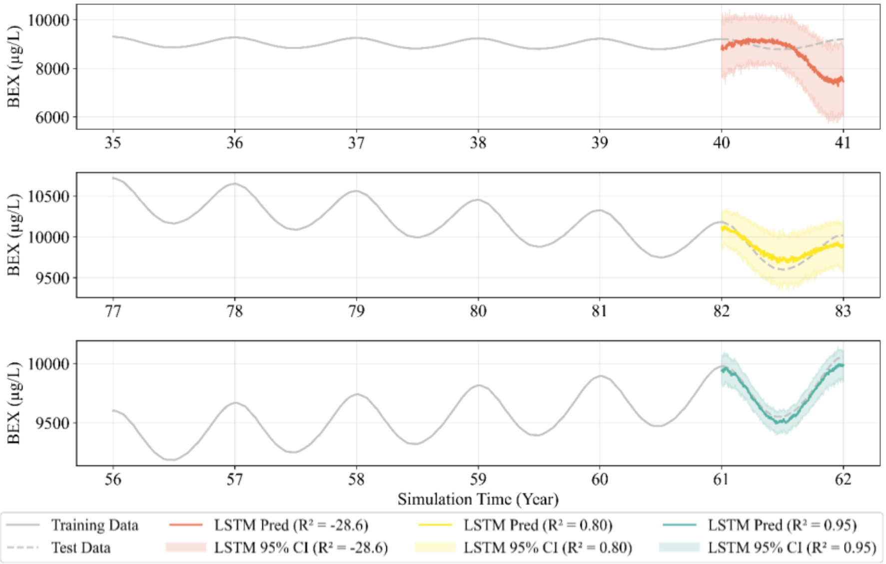


**Figure 5.5**. Sample training and testing windows with LSTM predictions using a 30-day input sequence at observation well X0Z0. Three predictions with different $R^2$ values are illustrated: approximately -28.6 (red), 0.80 (yellow), and 0.95 (green), corresponding to mean absolute error values of about 567 µg/L, 71 µg/L, and 32 µg/L, respectively. Shaded areas represent 95% confidence intervals estimated using Monte Carlo dropout.

To quantify predictive uncertainty of the model, Monte Carlo dropout was applied at inference time to generate probabilistic forecasts. In this approach, dropout remains active during prediction, and the LSTM model is evaluated multiple times with different dropout masks, producing an ensemble of outputs for each input sequence. The mean prediction and the corresponding 95% confidence intervals were then approximated and derived from these outputs, as shown by the shaded regions in **Figure 5.5**.

A negative $R^2$ indicates that the model performs worse than a simple mean predictor by failing to capture the underlying trend in the data (Chicco et al., 2021). Although its magnitude reflects the degree of deviation from the test data, any negative $R^2$ were treated as an indication of poor performance and truncated these values to zero.

This variability in predictive accuracy, as indicated by metric values such as $R^2$, MAE, and MAPE, is linked to the evolving hydrogeochemical conditions within the aquifer. The observed fluctuations in the LSTM model's performance align with the five identified hydrogeochemical periods defined in **Chapter 3**, where the relationship between iWQPs and BEX concentration changed in time.

The window with the poor fit ($R^2 < 0$) coincided with period 3 (around year 20 to year 60), where the BEX compounds reached a state of quasi-equilibrium. In this period, the rolling Spearman's correlation between iWQPs and BEX dropped to approximately zero from around year 25 to year 40. Instead, iWQPs became highly correlated with non-volatile dissolved organic carbon (NVDOC), a fraction of the dissolving light non-aqueous phase liquid (LNAPL). NVDOC became the primary electron donor, and thus the dominant driver of geochemical change. This results in the poor performance of the LSTM model for BEX concentration estimation.

Conversely, the window with a high $R^2$ of 0.95 coincided with the transition between period 3 and period 4 (around year 55 to year 65). During this time, iWQPs were strongly correlated with both BEX and NVDOC concentrations. The model successfully learned this combined, strong signal of the relationship between BEX and iWQPs, resulting in highly accurate predictions.

The window with good performance ($R^2 \approx 0.80$) corresponded to period 4 (around year 60 to year 90). Here, the dissolved NVDOC concentration peaked and began to decline. This reduction in NVDOC's influence allowed the correlation between iWQPs and the persistent BEX plume to re-strengthen to a moderate level. Even though the relationship was not as strong as in the transitional period, it provided a sufficient signal for the model to achieve reliable estimation accuracy. Also, while the specific $R^2$ values varied, the general performance trend tied to previously identified hydrogeochemical periods (**Chapter 3**) was consistently observed at other test windows.

A key assumption in this analysis was that the models were not retrained during the one-year test period. In actual model deployment, models can and should be retrained more frequently, especially under highly variable field conditions. For test windows where the LSTM model performance declines, implementing a more frequent retraining schedule (e.g., weekly, monthly, or quarterly depending on data availability) would likely mitigate performance decay and maintain model accuracy.

### 5.3.3. Comparison Across Observation Wells

**Figure 5.3** also shows that more test windows had accurate predictions at observation well X0Z0 (34 windows with $R^2 \geq 0.80$) than at other wells (fewer than 18 windows) for a 30-day sequence length. This spatial difference in model performance aligns with previous findings that the relationship between BEX and iWQPs is not only temporally variable but also spatially heterogeneous. The better performance at X0Z0 is likely due to two factors, as established in the previous chapter (**Chapter 3**).

First, the identified periods 1 to 4 began and ended earlier at well X0Z0 (closer to the source zone), than at distant wells such as X3Z3 (see **Figure 3.3** and **Figure 3.5**). For instance, the shift from BEX to NVDOC as the primary electron donor occurred after approximately 20 years at X0Z0 but was delayed until around 31 years at X3Z3 . Consequently, there were more test windows at X0Z0 where iWQPs were strongly correlated with BEX (e.g., the transition between Periods 3 and 4). At more distant wells, these predictive phases were less distinct, and their signal was weaker within any given five-year window, resulting in fewer windows with good accuracy ($R^2 \geq 0.80$).

Second, the signal (i.e., the change in iWQPs due to PHC degradation) was attenuated with distance from the source zone due to aquifer heterogeneity. From **Chapter 3**, it can be concluded that the amplitude of water table fluctuations and their consequent impact on iWQP signals (e.g., pH and EC) were dampened at farther wells like X3Z3. This attenuation weakened the strength of the BEX-iWQP relationship, which the LSTM model relies upon. The stronger, less attenuated signal recorded at X0Z0 provides a clearer input pattern for the model to learn, thereby producing more consistently accurate predictions across multiple test windows. Thus, the location of the observation well was a key factor in determining whether BEX concentration can be accurately estimated from iWQPs, depending on travel times and signal strength.

### 5.3.4. Hydraulic Head as an Additional Input

**Figure 5.6** presents heatmaps of $R^2$ values for LSTM predictions (30-day sequence length) across all observation wells, comparing model performance (a) without and (b) with hydraulic head included as an input feature alongside the four iWQPs. To avoid visual clutter, the figure illustrates a subset of the results, showing only every fifth test window starting from year 6 after the oil spill simulation. The addition of hydraulic head, which exhibited a high correlation with BEX concentration, markedly improved overall model performance.

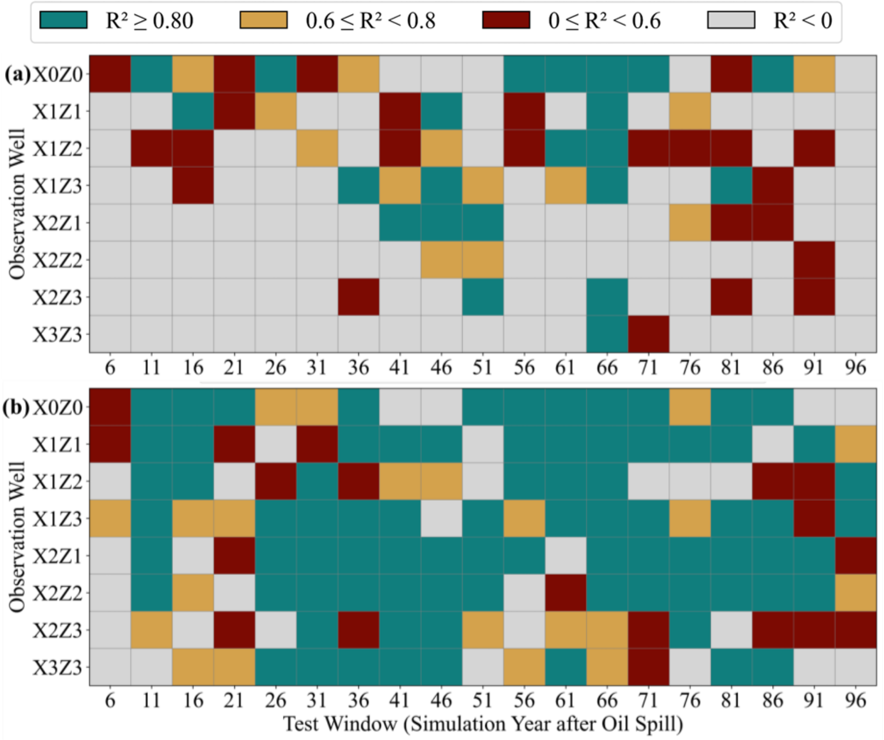


**Figure 5.6**. Comparison of LSTM model performance with and without hydraulic head as an input feature. Heatmaps show the R² values for estimating BEX concentrations across all observation wells and test windows for the 30-day sequence length LSTM model using (a) only iWQPs and (b) iWQPs and hydraulic head. Results are presented at every fifth test window for visual clarity.

This improvement was particularly evident at specific wells. At X0Z0, the number of test windows achieving $R^2 \geq 0.80$ increased from 34 to 43 out of 94 test windows. The highest improvement occurred at the farthest well, X3Z3, where the number of windows with $R^2 \geq 0.80$ increased from only 2 (~2%) to 35 (~37%) out of 94 test windows (**Figure C5**). This demonstrates that hydraulic head was a critical feature for estimating BEX concentration. Its importance is mechanistic: hydraulic head contains information on groundwater flow direction and advective transport of contaminants. In the RTM, simulated sinusoidal water table fluctuations were a key driver of concentration changes. By providing this signal explicitly, the LSTM could learn to recognize the fluctuation pattern as a coherent driver of BEX variability, rather than treating it as noise to be overcome using iWQPs alone.

On the other hand, a model trained only on hydraulic head performed poorly (yielding negative $R^2$ values; results not shown). While this model learned the sinusoidal temporal pattern, it failed to predict the correct magnitude of BEX concentration, showing that the geochemical information from iWQPs was essential for quantifying concentration levels.

However, performance gains were not the only observed results. **Figure 5.7** illustrates the per-window change in $R^2$ across all eight wells after adding hydraulic head. While performance generally increased, especially at X3Z3, at well X0Z0 it decreased in more than 20 test windows. It must be noted that for **Figure 5.7**, the test windows with a negative $R^2$ both before and after adding hydraulic head were categorized as showing "no change" in LSTM model performance (marked by yellow boxes). This classification was applied regardless of any variations in their mean absolute error (MAE).

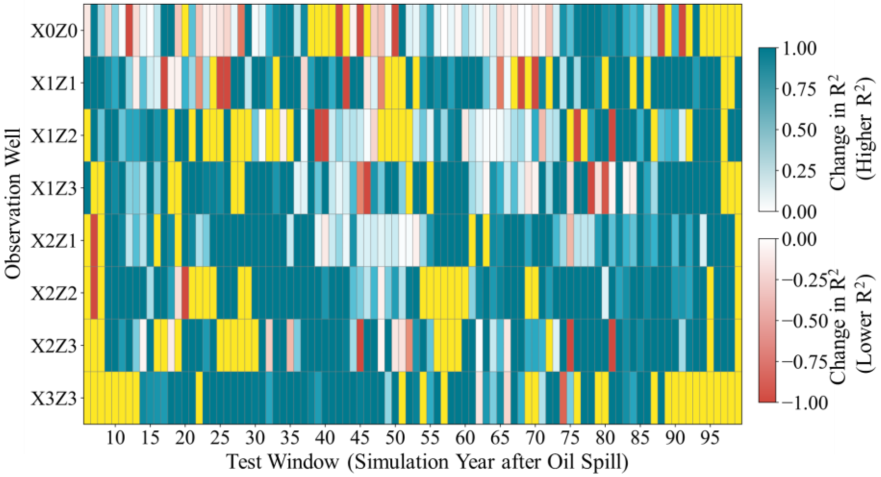


**Figure 5.7**. Changes in $R^2$ across test windows at eight observation wells after adding hydraulic head as an input feature alongside iWQPs in LSTM models (30-day sequence length). Yellow boxes denote windows with negative $R^2$ in both models (with and without hydraulic head). $R^2$ differences were clipped to the range [-1, 1] for visualization.

For example, during the test window from simulation year 87 to 88, a gradual decrease in BEX concentration induced changes in the iWQPs, whereas hydraulic head fluctuations remained stable. The model trained on the four iWQPs correctly captured this decreasing trend (**Figure 5.8a**). However, when hydraulic head was added as an input parameter, the model's performance decreased (**Figure 5.8b**). Although the hydraulic head was a useful predictive feature during the training period, its behavior in this specific test window created a conflicting signal: the average hydraulic head remained stable while the average BEX concentration decreased and other iWQPs changed (**Figure C6**).

The ML model incorrectly attributed a portion of the BEX behavior to the stable hydraulic head signal it had learned during training. This reduced the influence of the more predictive iWQPs, leading to less accurate estimates and a lower $R^2$. This suggests that while the hydraulic head can be beneficial, it can potentially introduce noise and reduce model accuracy in periods where its behavior deviates from the target contaminant's behavior. Dynamically adjusting the importance or weight of features could potentially solve this. For example, if iWQPs show strong variation but the average hydraulic head is flat, the model could learn to reduce the weight of hydraulic head (Yao and Ge, 2021).

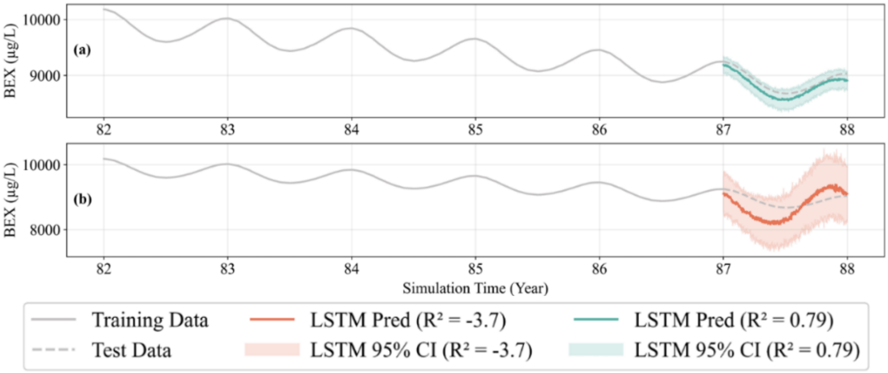


**Figure 5.8**. Comparison of LSTM model performance in estimating BEX concentration (µg/L) at observation well X0Z0 before (a) and after (b) including hydraulic head as an additional input.

Beyond these hydrogeochemical explanations, the stochastic nature of model training is thus, again, a contributing factor. Similar to other neural networks, the random initialization of model weights at the start of training influences which patterns the LSTM prioritizes and its convergence path (Narkhede et al., 2022). This randomness means that the exact number of test windows with an $R^2 \geq 0.8$ can vary between independent training runs. For example, in a separate LSTM model run (training and testing) under identical conditions, high-accuracy windows ($R^2 \geq 0.80$) at well X0Z0 for the 30-day sequence length were 27. This was lower than the 34 test windows observed in the initial run. Therefore, while the reported counts indicate a strong positive trend, they must be interpreted within the context of this inherent variability.

### 5.3.5. Comparison between Regression Models

The hypothesis that the LSTM model would outperform classical regression models such as MLR, SVR, RF, and XGB, was supported in several, but not all, test windows. The LSTM's ability to learn temporal patterns from sequences of iWQP data provided an advantage for estimation, particularly in scenarios where the correlation between iWQPs and BEX evolved over time. This better performance was especially noticeable at observation well X0Z0, where classical models had fewer than 20 test windows with $R^2 \geq 0.8$.

A representative example of the LSTM's better performance is shown in the test window at simulation years 33-34 at well X0Z0 (**Figure 5.9a**). Here, the LSTM achieved a high accuracy ($R^2 \approx 0.84$), while all classical models produced negative $R^2$ values with high error (MAE > 150 µg/L). Even an LSTM-XGB hybrid model performed poorly ($R^2 \approx -0.86$) in this instance. A feasible explanation for the LSTM's success is that, while the instantaneous correlation (Spearman’s rolling correlation) between iWQPs and BEX was low during the test year, the model leveraged its memory of stronger correlations in the preceding months within its 30-day input sequence. This historical context allowed it to make accurate predictions even as real-time correlations weakened, a capability the static models inherently lack. The subsequent rise in MAE after approximately 60 days suggests a limit to this predictive persistence as the system dynamics eventually deviated too far from the learned historical patterns.

However, the LSTM's reliance on temporal context could also be a disadvantage. In test window 13 at well X0Z0 (**Figure 5.9b**), the LSTM performed very poorly ($R^2 \approx -16.6$), while MLR and the LSTM-XGB hybrid excelled ($R^2 \approx 0.88$ and 0.93, respectively). In this scenario, the relationship between iWQPs and BEX was likely strong and consistent within the specific test window but may have been dissimilar to patterns in the immediate historical data used by the LSTM's sequence. The classical models, which treat each day independently, could directly map this strong contemporaneous correlation without the "confusion" of irrelevant historical data. The superior performance of the LSTM-XGB model here suggests its architecture successfully fused the LSTM's feature extraction with the gradient boosting's ability to leverage strong, static feature-target relationships, making it the most robust and adaptable model overall.

These observations confirm that model efficacy is not universal but is instead determined by the nature of the BEX-iWQP relationship at a given time. Optimal model depends on whether the predictive signal is embedded in temporal patterns (favoring LSTM) or in strong instantaneous correlations (favoring models like MLR or the LSTM-XGB hybrid). Temporal patterns were crucial when the geochemical processes were more complex, for example, a change in the dominant electron acceptor. However, the existence of windows where static correlations dominate, such as when the geochemical reactions reach a quasi-equilibrium state, justified the use of a hybrid modeling approach or a model selection framework.

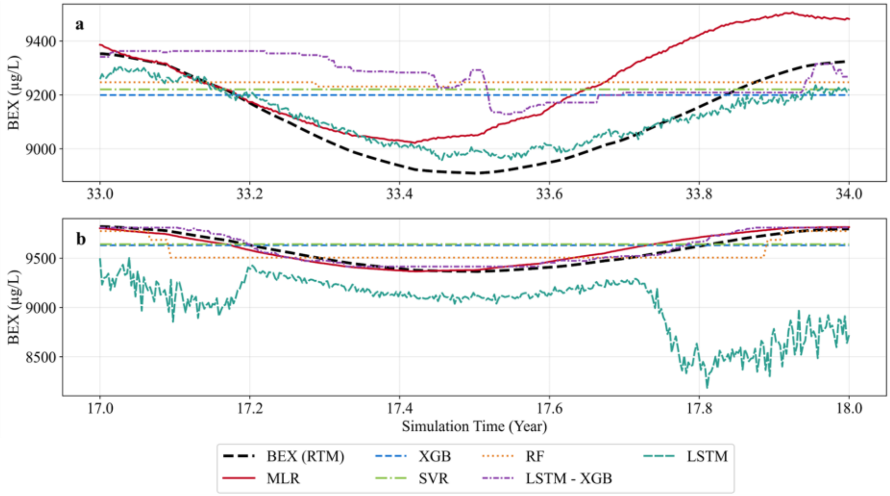


**Figure 5.9**. Comparison of BEX concentration estimates from six regression models at two test windows. The LSTM model (a) outperforms and (b) underperforms the other five models.

### 5.3.6. Evaluation under Increased Hydraulic Gradient and Source Removal

#### Increased Hydraulic Gradient

A 10x increase in the hydraulic gradient degraded the LSTM performance across the observation wells. Probably, the sudden increase in flow velocity enhanced the dilution and “flushing” of the BEX plume, causing an immediate drop in BEX concentration at well X0Z0 (**Figure C7**). In other wells farther from the source zone, the faster groundwater flow initially pushed a concentrated BEX slug forward (causing a spike) before subsequently diluting it (causing the decline). The trained LSTM model failed to estimate BEX concentration and struggled to generalize under this hydrological condition (**Figure C8**). This behavior confirms a known limitation of deep learning models. LSTMs and other ML models often fail to extrapolate accurately when faced with conditions outside their training range (Baste et al., 2025). Improving the training strategy could address these limitations: adding more training data with a wider range or adjusting the model configurations.

Despite the poor estimation of BEX concentration, the LSTM correctly predicted the direction of concentration trends (i.e., increasing or decreasing) in six of the eight wells (X0Z0, X1Z1, X1Z2, X1Z3, X2Z1, and X2Z3); the model failed to capture the increasing BEX concentration at wells X2Z2 and X3Z3. This ability to detect a sudden change in trend, particularly in the six wells where directional shifts were correctly predicted even when the magnitude was inaccurate, serves as an effective indicator of an anomaly in groundwater processes. Thus, governing bodies could use these predictions to trigger a manual sampling campaign, changing a model's limitation into a practical tool for adaptive monitoring.

### LNAPL Source Removal

The LSTM model's performance after source removal was notably poor, failing to capture the rapid decline in BEX concentration. The model predicted that concentrations would remain at an equilibrium state, despite the actual decrease observed in the RTM output (**Figure C9**). This failure probably occurred because the relationship between BEX concentration and the iWQPs was altered by the source removal; the existing BEX plume continued to undergo biodegradation without additional source, as observed from the RTM results. This sustained the geochemical reactions that affect the iWQPs. However, the sudden drop in BEX concentration due to the removed source was much faster than the decline from biodegradation. This created a discrepancy between the actual BEX concentration and the concentration effectively governing the iWQP signal. The LSTM model, trained on a system with an active source, lacked the necessary information to detect this change. It therefore interpreted the stable geochemical conditions as evidence of a persistent, steady-state plume, leading to the model's inaccurate predictions.

An exception was observed at well X2Z1, where the LSTM model correctly predicted a decreasing trend in BEX concentration following source removal (**Figure C9**). This response, however, may be attributed to a pre-existing decline in concentration at that location rather than an accurate prediction of the source removal's impact.

The LSTM's inability to predict the BEX concentration decline after source removal emphasizes an important limitation of the model that could not extrapolate beyond the conditions seen in its training data. This demonstrates the necessity for a recursive updating method like the Kalman Filter (KF), which can continuously adjust its estimates as new information becomes available (Simon, 2006).

In the subsequent KF experiment (**Figure C10**), the BEX concentration measurements were assumed to be available only annually, while iWQP measurements were available daily. The KF implementation used a first-order attenuation process model for BEX, with an effective attenuation rate (λ) updated every five years. This parameter intended to capture the combined effects of advection and biodegradation. However, the KF did not immediately capture the initial decrease in BEX concentration (**Figure C10)** either. Its estimation only aligned with the RTM output after incorporating the first annual BEX measurement following source removal. While the filter did not estimate the exact concentration values well, it successfully captured the general decreasing trend following the update.

The KF framework thus served as a foundation for applying data assimilation to dynamic systems like source removal. Future improvements could involve developing a more complex process model that better represents subsurface physics, potentially even integrating the RTM itself. Additionally, the observation model could be enhanced by creating a more robust statistical translator between iWQP sensor data and BEX concentration. Finally, more frequent BEX measurements, such as quarterly or monthly, would significantly improve the accuracy and timeliness of the KF's estimations.

### 5.3.7. Final Discussion and Recommendations

Incorporating hydraulic head data significantly enhanced BEX concentration estimates. This improvement was most pronounced in observation wells located farther from the source zone. In heterogeneous aquifers, additional information on groundwater flow and contaminant transport pathways becomes critical in these locations. The inclusion of hydraulic head offers a practical and cost-effective strategy to boost model performance, particularly at sites with strong hydrological seasonality. It must be considered that the hydraulic head adds a cost for the divers, along with the four in-situ water quality sensors (**Table C1**). Moreover, the model reliability could be further improved by implementing more frequent retraining cycles and reducing prediction horizons (e.g., monthly instead of yearly).

Selecting the most suitable regression model for field application remains challenging due to deviations between real-world conditions and RTM simulations. Nevertheless, the proposed monitoring strategies in this research were effective for estimating concentration trends and identifying hotspot areas. This enables the development of adaptive monitoring systems that can issue alerts when regulatory thresholds are exceeded, and trigger increased measurement frequency to identify the causes and characterize sudden concentration changes. It should be emphasized that for compliance monitoring affecting critical areas such as drinking water wells, more intensive and costly direct measurements remain necessary.

There is also potential to enhance the KF framework. The process model could be refined by incorporating more complex mechanisms such as dissolution, advection, and biodegradation, bringing it closer to a full RTM. Similarly, the measurement model could account for chemical reactions or be updated with more frequent PHC measurements to increase accuracy.

Furthermore, using hydraulic head data from both upstream and downstream wells relative to the target point was recommend. This configuration provides a direct measurement of the hydraulic gradient, allowing the model to better infer groundwater flow velocity and the timing of contaminant arrival; better information on hydraulic gradient could improve the model's predictive performance.

A primary limitation of this study is the lack of available extensive field data. There is a scarcity of BEX concentration and iWQP measurements for real-world model validation (**Figure C11**). Thus, future research should prioritize collecting high-resolution temporal data to capture contaminant plume dynamics more accurately and allow for model evaluation under field conditions.

## 5.4 Conclusion

The prediction accuracy of LSTM and classical regression models in estimating BEX concentrations using iWQPs were evaluated using virtual data from an RTM under controlled conditions. It was found that:

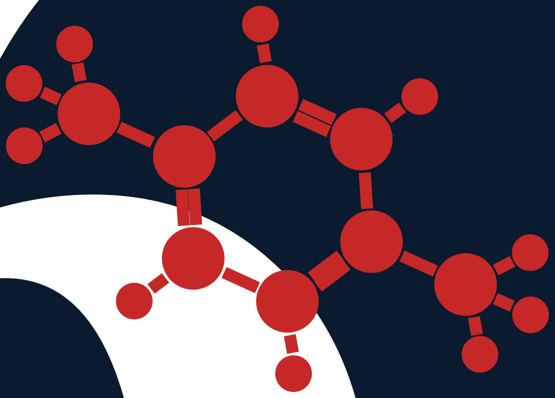
- While the LSTM model showed limited performance based on $R^2$ ($R^2 < 0.80$ in 98% of test windows at well X3Z3), over 70% of test windows fall within a 5% error margin (MAPE). This indicates potential under controlled conditions but should be interpreted cautiously for real-world application given the observed sensitivity to noise.
- Incorporating hydraulic head considerably improved LSTM accuracy, increasing the number of test windows with $R^2 \geq 0.80$ at well X3Z3 from 2% to 37%. However, while hydraulic head contained information on contaminant transport, it can introduce noise and reduce accuracy when contaminant behavior differs from groundwater flow (e.g., when biodegradation dominates over advection).
- Model performance for predicting BEX concentrations depended on the BEX-iWQP relationship. LSTMs showed potential for capturing temporal patterns from sequential inputs, while MLR and the LSTM-XGB hybrid performed well for stronger instantaneous relationships.
- During an extreme hydraulic gradient increase, the LSTM's predictive accuracy declined, confirming its inability to extrapolate beyond training data. Nonetheless, it correctly predicted the direction of concentration trends in 75% of the observation wells. However, the model remained valuable as an anomaly detection system to trigger manual sampling.
- Following source removal, the LSTM model failed to capture the rapid concentration decline, incorrectly predicting a steady-state equilibrium. In contrast, a recursive method like the KF successfully captured the

decreasing trend after source removal by continuously updating its estimates with new measurements.

- Future work should prioritize collecting high-resolution temporal field data from PHC-contaminated aquifers to evaluate and refine these models under real-world conditions.

# Synthesis

# 6



*"A century ago, petroleum – what we call oil – was just an obscure commodity; today it is almost as vital to human existence as water. Oil transports us, powers our machines, warms us and lights us. It clothes us, wraps our food and encases our computers. It gives us medicines, cosmetics, CDs and car tyres. Even those things that are not made from oil are often made with oil, with the energy it gives. Life without oil, in fact, would be so different that it is frightening to contemplate."*

*– J. Buchan, 2006*

# 6. Synthesis

## 6.1 Overall Conclusions on the Scientific Findings

Microbial communities in aquifers can degrade and thereby attenuate PHCs, making monitored natural attenuation a preferred long-term management strategy for PHC contaminated sites. Conventional monitored natural attenuation strategies rely on manual sampling and laboratory analysis, which are costly, labor-intensive, and often fail to capture critical temporal variations in contaminant behavior. This limits the ability to generate high-resolution datasets needed to validate attenuation processes and ensure regulatory compliance in case of unforeseen events, such as unaccounted oil leakage source or extreme events that affect plume migration. Highly toxic compounds such as BTEX, especially benzene with a drinking water limit of just 5 µg/L in the United States (US EPA, 2024), must be monitored closely.

The overall objective of the thesis was to determine and assess the potential of in-situ water quality parameters (iWQPs) for detecting and estimating petroleum hydrocarbon (PHC) contamination in groundwater. The thesis presents an integrated framework for using iWQPs in groundwater monitoring, effectively determining and assessing their potential for detecting and estimating PHC contamination. The findings demonstrate that reactive transport model (RTM)-driven insights and data-driven models can enable a cost-effective approach to groundwater monitoring. RTM explains why and how iWQPs change in response to PHC contamination, while machine learning (ML) models use these changes as training data to provide real-time early warning and continuous estimation of PHC concentrations in groundwater.

First, using an RTM which is a physics-based model, the dynamic correlation between iWQPs and PHCs affected by the aquifer's spatiotemporal and geochemical conditions was established (**Chapter 3**). Building on this process-based understanding, it was demonstrated that these correlations can be used for SDF: ensemble ML models such as RF and XGBoost can reliably convert iWQP parameters into early contamination warnings, even under diverse and noisy field conditions (**Chapter 4**). Finally, the research progressed from detection to quantification where LSTM models were shown to estimate PHC concentrations with practical accuracy (**Chapter 5**). The performance also improved significantly when hydraulic head data were included, providing essential context on plume movement.

By demonstrating that ML models can estimate BEX concentrations using low-cost iWQPs (pH, DO, EC, ORP), a strategy was shown on how traditional monitoring could be enhanced in both spatial and temporal resolution. The proposed framework could enable continuous data collection and real-time analysis, enhancing the detection of concentration trends and anomalies. When deployed at downstream warning wells, these models can serve as early warning systems, triggering alerts and manual sampling when predicted concentrations exceed safety thresholds. As a result, sensitive receptors such as drinking water wells can be protected. Although this research primarily addressed high frequency temporal monitoring, the proposed ML framework is equally suited for spatial scalability. ML models can be deployed at multiple monitoring wells to capture a broader area of the plume, thereby improving spatial coverage for monitoring. Moreover, the ML models can have good predictive performance at varying downgradient distances when information on these distances were incorporated during training.

However, interpreting ML model performance requires context. Common evaluation metrics like $R^2$ and MAPE provide statistical benchmarks, but their practical significance depends on the monitoring objectives. For example, a 5% error may be negligible at sites with high contamination levels (~10,000 µg/L), but critical when assessing compliance near regulatory limits. Additionally, the resolution and accuracy of commercially available sensors must be considered, as small but meaningful changes in iWQPs may be difficult to detect. Despite these limitations, the results showed that the combined response of multiple sensor parameters can still reliably support early warning systems even when individual deviations are small, as described in **Chapter 4**.

## 6.2. Conclusions on Research Questions

Three main research questions were defined in **Chapter 1**:

> **Research question 1:** How do iWQPs correlate with dissolved PHC concentration in groundwater, and how are these correlations influenced by spatiotemporal, hydrogeological, and geochemical characteristics of the aquifer? (Research Gap 3)
>
> **Research question 2:** Can ML models trained on iWQPs reliably provide early warning of PHC contamination in groundwater under varying aquifer conditions? (Research Gap 1 and 2)
>
> **Research question 3:** Can ML models accurately estimate toxic PHC concentration in groundwater using iWQPs, and maintain accuracy during abrupt changes (i.e., source removal, hydraulic gradient shifts)? (Research Gap 1 and 2)

To answer these research questions, a modeling approach was adopted. Focus was first given on understanding the fate, including biogeochemical and physical processes, and migration of dissolved phase PHC plumes in groundwater. Then, an RTM was developed to simulate groundwater flow and reactions under hypothetical but realistic PHC-contaminated aquifer conditions. This is to provide temporal and spatial datasets for further analysis. The model domain and parameters were based on existing studies that examine PHC contamination, specifically the crude oil spill site in Bemidji, USA (Ng et al., 2015), which is one of the most studied oil contaminated sites in the world. In the following, the research questions are addressed.

***Research question 1:*** *How do iWQPs correlate with dissolved PHC concentration in groundwater, and how are these correlations influenced by spatiotemporal, hydrogeological, and geochemical characteristics of the aquifer?*

The released hydrocarbons, such as from accidental leakage of underground storage tanks or buried oil pipes, tend to migrate towards the water table, causing a portion of the light non-aqueous phase liquid (LNAPL) to be entrapped in the saturated zone. LNAPLs are usually responsible for the continuous release of PHCs such as benzene, toluene, ethybenzene, and xylenes (BTEX) in the subsurface, which could then generate plumes of dissolved phase PHC in groundwater. The transport of LNAPL in the subsurface and its partitioning into different phases are governed by complex processes affected by several factors. Furthermore, various equilibrium and kinetic reactions can be triggered which could include, but are not limited to: the terminal electron-accepting processes facilitated by microbial activities, the contaminant partitioning (dissolution and volatilization) and degassing, the mass transport of BTEX compounds with respect to space and time; the dissolution and precipitation of minerals; the homogeneous and heterogeneous reactions with acids and bases; and sorption and surface complexation. Current understanding of these processes was discussed in **Chapter 2**. The migration time of PHC in the vadose zone is affected by the following: 1) the oil leakage rate and pore water saturation; 2) the geophysical characteristics of the media such as soil characteristics and capillary pressures; and 3) the rate of water infiltration into the soil. Furthermore, the solubility of the different LNAPL components dictates the distribution of these compounds in groundwater. The dissolution of LNAPL can be influenced by groundwater flow regimes, with higher flow velocity resulting in higher dissolution rate, water table fluctuations, and other geophysical and biochemical properties.

Due to lack of available field data, combined with the high cost and safety concerns associated with collecting high-resolution data from contaminated sites, a computer simulation modeling approach was followed. Therefore, in **Chapter 3**, an RTM that incorporates essential processes governing groundwater flow, PHC contaminant transport, and hydrogeochemical reactions is presented. The correlations between iWQPs and dissolved PHC concentrations in groundwater were then examined. The focus was on the iWQPs, pH, EC, DO, and ORP, and BEX and non-volatile dissolved organic carbon (NVDOC). Toluene was excluded from the BTEX group due to its rapid degradation, and it was assumed to degrade directly from the source zone. Strong correlations (Spearman's correlation > |0.8|) between iWQPs and PHCs at varying periods were determined, which were highly dependent on the location of observation wells, aquifer properties, and the available electron acceptors.

Five distinct periods were identified over a 100-year simulation, each characterized by different dominant electron donors and redox conditions. Initially, aerobic degradation of BEX led to rapid DO depletion and a sharp drop in ORP, accompanied by changes in pH and EC. As conditions transitioned to anaerobic, BEX continued to degrade, further acidifying the groundwater and increasing ion concentrations. Later, NVDOC became the dominant electron donor, producing more $H^+$ per molecule and resulting in lower pH and higher EC values. These changes were delayed relative to peak NVDOC concentrations, with lag times varying by well location due to differences in travel time and reaction kinetics. In the final stages, PHC concentrations declined, and iWQPs began to stabilize, though they did not return to baseline levels, indicating long-term aquifer impacts.

Spatial variabilities across the eight virtual observation wells located at increasing distances from the source zone were also observed. Wells farther from the source exhibited delayed PHC arrival, longer lag times between PHC peaks and iWQP responses, and attenuated signals due to aquifer heterogeneity. Geochemical processes such as calcite dissolution, iron reduction, and siderite precipitation also affected the iWQPs. PHC degradation produced $H^+$ and $HCO_3^-$, lowering pH and increasing EC. Calcite dissolution buffered acidity and contributed to ion release, while iron reduction and subsequent siderite formation influenced redox conditions. These processes further influenced the groundwater chemistry.

Overall, iWQPs were found to be highly correlated with PHC in groundwater when spatiotemporal variability was considered. DO and ORP were especially highly correlated with PHC during the initial years of contamination, while pH and EC were variably affected throughout the contamination period. Furthermore, the aquifer's hydrogeology impacts hydrocarbon-water quality correlations by influencing buffering capacity and flow velocity, among others.

***Research question 2:*** *Can ML models trained on iWQPs reliably provide early warning of PHC contamination in groundwater under varying aquifer conditions?*

Building on the findings in **Chapter 3**, it was studies whether ML models, trained in the four iWQPs, can reliably provide early warnings of PHC contamination in groundwater, as presented in **Chapter 4**. Using a binary classification framework, five ML classifiers (LR, RF, XGBoost, MLP, and SVC) were trained to detect when BEX concentrations reached or exceeded the regulatory threshold. To prioritize health protection, the U.S. Environmental Protection Agency's maximum benzene limit of 5 µg/L was adopted as a conservative threshold for the entire BEX group. It was based on benzene since it is recognized as the most toxic constituent within the BEX group. In comparison, ethylbenzene and xylenes have higher allowable concentrations in drinking water, which are 700 µg/L and 10,000 µg/L, respectively. At the threshold of 5 µg/L total BEX concentration, benzene concentration is below its 5 µg/L limit since the value represents the sum of all constituent concentrations. This approach prioritizes the detection of any potential benzene presence, ensuring a safety margin that flags samples for further investigation even if the immediate cumulative risk is low. The models were trained on virtual wells near the contamination source and tested on downstream virtual warning wells to simulate real-world early detection scenarios.

Most models triggered contamination alarms before BEX concentrations reached the threshold, with lead times ranging from 4 (acceptable) to 189 days (too premature/false alarm) depending on the well location. SVC, however, was more conservative, often triggering alarms only after the threshold was exceeded. At the farthest well, LR and MLP triggered alarms at very low BEX concentrations (0.6 and 1.2 µg/L respectively), while RF and XGB showed early detection. These discrepancies were attributed to differences in flow and reaction dynamics between training and test locations, which altered the relationships between BEX and iWQPs. Geochemical reactions had more time to affect iWQPs at distant wells, resulting in different data distributions between training and testing sets and causing premature or false alarms. In the scenario simulations, it was also found that aquifer properties affect model performance. Reduced hydraulic conductivity increased residence time and enhanced redox-driven degradation. As a consequence, iWQP–BEX correlations were reduced while iWQP–NVDOC correlations were increased. Under these conditions, LR became overly conservative, triggering alarms at 17.4 µg/L, while SVC falsely detected contamination in wells with non-detectable BEX levels. Elevated background salinity masked subtle changes in EC caused by BEX degradation, reducing the signal-to-noise ratio and leading to false positives. Changes in electron acceptor availability, such as elevated nitrate or absence of oxygen/$SO_4^{2-}$, disrupted the expected degradation pathways, which in turn delayed alarm triggers. For example, RF and XGB only flagged contamination when BEX concentrations exceeded 30 µg/L under high nitrate conditions.

The impact of synthetic sensor noise and seasonal fluctuations in iWQPs on ML model performance was also assessed. Gaussian noise and sinusoidal fluctuation amplitude as low as 10–20% of maximum field observable values caused premature, false, and delayed alarms. To mitigate the effect of the synthetically added noise, moving average smoothing was applied to the training data, with the 5-day window yielding the best model performance by balancing noise reduction and signal preservation.

Based on the findings using RTM-generated data, ensemble ML models such as RF and XGBoost demonstrated the ability to provide early warning of PHC contamination across varying aquifer conditions when trained on iWQPs. These results suggest a degree of robustness under the simulated groundwater scenarios considered. However, performance may be more variable under real-world conditions, particularly with noise. Coupling these models with low-cost sensors therefore offers a potentially promising and scalable approach for real-time groundwater monitoring, although further validation under field conditions is required. In this context, the ML frameworks may enhance early detection and reduce long-term monitoring costs by complementing, rather than replacing, manual sampling.

***Research question 3:*** *Can ML models accurately estimate toxic PHC concentration in groundwater using iWQPs, and maintain accuracy during abrupt changes (i.e., source removal, hydraulic gradient shifts)?*

In **Chapter 5**, the research moved one step further: estimating concentrations of BEX in groundwater using the four iWQPs. An ML-based framework, specifically an LSTM network, was proposed and its predictive accuracy was evaluated. To provide additional insight and explore alternative approaches, classical regression models and hybrid architectures were also implemented as baselines rather than for strict comparison. The framework's robustness was further assessed under dynamic scenarios, including source removal and hydraulic gradient shifts. To ensure adaptability while accounting for seasonal variability, rolling five-year training window approach was used. Models were trained on daily data from each five-year segment (trained on first four years, and validated on fifth year) and tested on the following year (referred to as test windows). This annual update offers a conservative yet computationally efficient benchmark, allowing for the performance assessment across different time periods. If the ML model performs well with annual updates, it establishes a baseline from which it can be inferred that more frequent updates would perform at least as well as, if not better than, yearly updates.

Initially, the LSTM model showed limited performance based on $R^2$ values (e.g., <0.80 in 32 out of 34 test windows at the farthest virtual observation well X3Z3). However, over 70% of test windows have predictions below or within a 5% error margin (MAPE), suggesting some potential for capturing general contamination trends under the conditions of this study. This level of performance should be interpreted with caution, as analyses demonstrate sensitivity to noise and variable fluctuations. The model's accuracy was highly dependent on the temporal and geochemical relationship between iWQPs and BEX concentrations. When iWQPs were strongly correlated with BEX degradation, the LSTM performed well; however, its accuracy declined when iWQPs were more correlated with NVDOC due to the shift in dominant electron donor.

Incorporating hydraulic head as an input feature significantly improved model performance, particularly in wells located farther from the source zone. At well X3Z3, the proportion of test windows with $R^2 \geq 0.80$ increased from only 2 (~2%) to 35 (~37%) out of 94 test windows, when hydraulic head was included. This improvement demonstrates that hydraulic head was a critical feature for estimating BEX concentration since it contains information on groundwater flow direction and advective transport of contaminants. However, the added feature also introduced noise in scenarios where biodegradation dominated over advection. Moreover, the hybrid LSTM-XGBoost model and classical regressors like MLR performed well in cases with stronger instantaneous correlations, suggesting that model choice should be adapted to the nature of the BEX-iWQP relationship.

Stress testing revealed limitations in the LSTM's ability to extrapolate beyond its training data. During extreme hydraulic gradient shifts, predictive accuracy declined, though the model still correctly identified the direction of concentration trends in 6 out of 8 virtual observation wells. This makes the model possibly useful for anomaly detection which could trigger increased measurement frequency campaigns to identify the causes and characterize sudden concentration changes. Following source removal, LSTM failed to capture the rapid concentration decline, instead predicting a false steady state. In contrast, KF successfully tracked the decreasing trend by continuously updating its estimates with new measurements. These findings support the use of recursive methods for dynamic monitoring and also show the potential of ML models for groundwater monitoring systems.

## 6.3 Research Perspectives

### Sorption and Desorption in the Reactive Transport Model

Sorption processes play an important role in the fate and transport of BTEX compounds in groundwater. These processes include physical, chemical, and electrostatic interactions, which can reduce the aqueous-phase concentration and mobility of BTEX compounds (Delle Site, 2001). The resulting retardation alters the arrival time and concentration profiles at an observation well, altering the relationship between iWQPs and PHC concentrations. This relationship is further complicated since water quality parameters (e.g., pH, dissolved oxygen) and soil characteristics also influence sorption rates (Fiore and Zanetti, 2009; Li et al., 2024).

In this study, sorption was excluded due to the complexity of representing the individual behaviors of different PHC fractions; volatile and non-volatile dissolved organic carbon fractions exhibit distinct sorption properties. Based on batch sorption isotherms, the BTEX compounds alone have different sorption tendencies to various media (Zytner, 1994; Seagren and Becker, 2002). Nevertheless, incorporating sorption can improve the predictive accuracy of ML models. Numerical studies confirm that sorption, coupled with dissolution and biodegradation, governs the spatial distribution and mobility of dissolved BTEX (Valsala and Govindarajan, 2018b)

To address this gap, future work should focus on numerically and experimentally understanding how sorption parameters affect the correlation of PHCs with iWPQs. This requires numerical simulations and controlled batch and column experiments, supported by field-scale validation, to evaluate the effect of different sorption isotherms (e.g., Linear, Freundlich, Langmuir). Comparative analysis of these models will help investigate how non-linear and variable sorption behaviors at different aquifer conditions affect the PHC-iWQPs correlations. Furthermore, RTM should integrate spatial heterogeneity in organic carbon content of aquifers and account for long-term contaminant release via desorption from residual source zones (Cvetkovic et al., 1998; Eckert and Appelo, 2002). Such investigations can enable more realistic RTM simulations, provide more representative training data, and improve the robustness of data-driven sensor data fusion frameworks for BTEX estimation and early warning systems.

### Applicability of the SDF Approach to other Contaminants

Beyond PHC compounds, the SDF approach proposed in this study may also be applicable to other contaminants that undergo natural attenuation in aquifers. These include organic micropollutants such as pharmaceuticals, agricultural pesticides, and nutrient-related compounds from fertilizers (e.g., $NO_3^-$ and $PO_4^{3-}$). At high $NO_3^-$ levels greater than 600 mg/L in groundwater such as those in agricultural areas of India (Pant et al., 2019; Abascal et al., 2022), denitrification with highly reactive organic carbon as electron donor can be expected to alter redox conditions detectable using sensors (Richard-Cerda et al., 2024). The spatial distribution of geogenic phosphorus has been successfully predicted in aquifers using field parameters including well depth, pH, ORP, EC, $Fe^{2+}$ and $NH_4^+$-N (Y. Li et al., 2022). Thus, it could be hypothesized that $PO_4^{3-}$ from fertilizers can be detected using iWQPs. Furthermore, ammonium from fertilizers or wastewater undergoes biooxidation in the environment which consumes oxygen and reduces pH (Prinčič et al., 1998).

Although organic micropollutants occur at trace levels (ng/L–µg/L), their biodegradation often involves the consumption of electron acceptors (e.g., oxygen and $NO_3^-$) (Fenner et al., 2013; Helbling, 2015). These reactions could potentially cause measurable shifts in redox potential and dissolved oxygen, particularly in low-buffered aquifers. Furthermore, cumulative effects of multiple co-occurring contaminants and their intermediates may amplify these changes.

Intermediate degradation products of PHCs, such as phenol and benzoate, also represent important targets for future work. Phenol, formed via hydroxylation of benzene, and benzoate, an intermediate in anaerobic benzene degradation (Caldwell and Suflita, 2000; Zhang et al., 2013), can both influence groundwater chemistry and could serve as indicators of ongoing biodegradation. Investigating their unique signatures in iWQPs may improve the interpretability of sensor-based models and help distinguish between parent compounds and intermediates.

### Development of In-situ BTEX Sensors

Spectrophotometric sensors such as UV-based sensors hold significant potential for advancing BTEX monitoring because they offer rapid, sensitive, and non-destructive analysis. Future work could explore the integration of UV-LED spectrophotometry with fluorescence techniques as a cost-effective and selective approach. UV spectrophotometry has been applied for BTEX detection in gas-phase soil remediation (Dupuit et al., 2000) and indoor air quality monitoring (Khan et al., 2020).

Recent developments include low-cost UV-induced fluorescence sensors for monitoring aromatic hydrocarbons in coastal marine water (Lo Savio et al., 2024) and integrated platforms combining UV/Vis and fluorescence for analyzing unsaturated and aromatic compounds in aquatic ecosystems (Goblirsch et al., 2023). The principle is that BTEX compounds exhibit specific excitation-emission wavelengths; thus, multi-wavelength excitation and fluorescence measurements could improve specificity and enable real-time detection with minimal sample preparation for continuous water quality monitoring. Despite these advantages, deep UV spectrophotometry for BTEX detection still faces critical challenges. Important areas for investigation include reducing the cost of materials, mitigating interference from natural organic matter, enhancing sensitivity at sub-ppm concentrations, and addressing optical limitations inherent to liquid-phase analysis. Research focused on novel calibration strategies, advanced signal processing, and hybrid sensing technologies could help overcome these constraints and unlock the full potential of spectrophotometric methods for in-situ applications. These in-situ sensors can then be used to complement the ML-based frameworks for a more informed contaminant monitoring.

### Field Validation of the SDF Approach

A primary limitation of this study is the lack of extensive field data for validating the proposed ML-based framework in practice. Currently, there is a scarcity of high-resolution temporal datasets that include both BEX concentrations and iWQPs, which are essential for assessing model performance under real-world conditions. Although ORP sensors have been deployed in the field for groundwater monitoring of PHC in contaminated locations (Sale et al., 2021), future research should prioritize the collection of iWQPs including pH, DO, EC and ORP, along with PHC concentration, for field validation.

It should be noted that model performance observed in this study, based on simulated RTM data, represents an optimistic scenario compared to real-world conditions. In practice, additional sources of uncertainty are expected to reduce predictive performance. These sources may include sensor noise, environmental variability, and even incomplete system representation.

For field validation and deployment, sensors for measuring iWQPs should be installed at monitoring wells to collect continuous water quality data. High-resolution spatiotemporal data are essential for training ML models and establishing site-specific baselines. During the initial contamination phase, monthly BTEX measurements via laboratory analysis are recommended, with more frequent sampling (e.g., weekly or daily) improving model robustness. Once trained, models can be deployed at downstream warning wells, with sensor data streamed to a central system for ML-based analysis. Alerts would trigger manual sampling to confirm contamination. Initial costs include sensor installation and calibration, while ongoing costs involve maintenance and periodic model retraining. Compared to manual grab sampling (~€130 per BTEX analysis), in-situ water quality sensors (~€400 total) offer continuous monitoring and faster detection, potentially reducing long-term costs. This can also be combined with the installation and testing of in-situ BTEX sensors using UV-LED spectrophotometry.

A possible topic for future research is the use of simulated data for model pretraining, followed by fine-tuning with site-specific field data. Such an approach may help overcome data scarcity, particularly in the early stages of monitoring. However, its effectiveness is expected to depend on how well the simulated data represent the hydrogeological and geochemical conditions of the target site. Further investigation is therefore required to evaluate the transferability of models trained on synthetic data to real-world applications.

### Incorporating Environmental and Anthropogenic Factors

Pollutant selectivity remains a major challenge. In aquifers with multiple co-occurring contaminants, changes in iWQPs may reflect combined effects, complicating attribution to a single compound. At PHC-contaminated sites, PHCs typically dominate electron donor dynamics, whereas for other pollutants, their individual and combined impacts on water chemistry must be carefully characterized.

Groundwater sites contaminated with total PHC were identified in China using parameters such as wind speed, population density, evaporation rate, total potassium in the soil, precipitation, and leakage accident as inputs to ML models (Huang et al., 2023). Thus, beyond traditional in-situ sensors, satellite-based remote sensing can offer an important, spatially extensive data source for groundwater contamination assessment. Geospatial imagery can provide context in terms of land-use patterns and potential pollutant sources. For instance, high-resolution imagery and land-cover classification can delineate agricultural zones (associated with $NO_3^-$ and pesticides), industrial sites, and hospitals (linked to organic micropollutants). Integrating this geospatial data into SDF frameworks can potentially enhance the predictive modeling performance. It can also enable the early detection of land-use changes that may introduce new contamination threats in aquifers with different pollutants. Other factors could include management practices, geographic and hydrological context (e.g., proximity to surface water bodies affecting recharge and flow paths), and meteorological conditions.

## 6.4 Final Remarks

This thesis shed light on the relationship between PHCs and iWQPs in groundwater. It was found that this correlation is not static but varies depending on factors such as the elapsed time since contamination and the monitoring location relative to the contaminant source. Distinct periods of high and low correlation were identified, primarily governed by the dominant terminal electron-accepting processes and prevailing groundwater flow conditions.

To translate these insights into actionable solutions, the study introduces proof-of-concept groundwater monitoring systems designed to detect and estimate PHC concentrations using iWQPs with operational accuracy. ML-based frameworks were developed and evaluated to show how PHC-iWQPs correlations can be used for practical applications. Specifically, these frameworks enable early detection when contamination reaches previously uncontaminated wells and provide concentration estimates for contaminated using only iWQP data. The frameworks developed in this thesis offer cost-effective solutions for continuous and real-time groundwater contaminant monitoring. Rather than replacing conventional monitoring practices, these frameworks serve as a complimentary tool that can provide high-resolution data to improve decision-making and safeguard water resources.

# Appendices

## Appendix A

**Text A1.** Larger model used for determining the general head boundary condition at the right boundary of the base case model.

A 15 m by 5,200 m 2D groundwater flow model was set up in FloPy. The grid cells were 0.5 m (in y-direction) by 1 m (in x-direction). The hydraulic conductivity was set at 6.1 m/d, the aquifer storage coefficient at 0.1, and the aquifer storage yield at 0.2, similar with the base case model. For the transient groundwater flow simulation, the general head boundary condition was used at the left border, while the constant head boundary condition at the right border of the model. No recharge was assumed, and a no-flow boundary condition was implemented at the bottom of the model.

**Table A1**. Set of parameters used in the flow and transport model.

| Parameters | Value, unit | Source |
|---|---|---|
| Model dimensions $l_x$ x $l_z$ | 300 x 10 m | Ng et al., 2015 |
| Discretization $d_x$ x $d_z$ | 2 m x 0.2 m | This research |
| Hydraulic conductivity (heterogeneous) | 6.2 m/d | This research |
| Porosity | 0.38 | Dillard et al., 1997 |
| Transverse dispersivity | 0.04 m | Ng et al., 2015 |
| Longitudinal dispersivity | 1 m | Ng et al., 2015 |
| Water table gradient | 0.0035 m/m | Ng et al., 2015 |
| Groundwater recharge rate (average) | 4.88 x $10^{-4}$ m/day | Ng et al., 2015 |
| Aquifer storage coefficient | 0.01 | This research |
| Aquifer storage yield | 0.20 | This research |

**Table A2**. Initial groundwater composition of the base case model.

| Parameter | Unit | Value |
|---|---|---|
| Temperature | ° Celsius | 10 |
| pH | - | 7.7 |
| pe | - | 14.2 |
| Cl | mmol/L | 0.31 |
| DIC | mmol/L | 3.60 |
| $O_2$ | mmol/L | 0.25 |
| $NO_3$ | mmol/L | 0.36 |
| $SO_4$ | mmol/L | 0.52 |
| Ca | mmol/L | 1.31 |
| Mg | mmol/L | 0.58 |
| Na | mmol/L | 1.36 |
| K | mmol/L | 0.021 |
| Fe(II) | mmol/L | 0 |

**Table A3.** Equilibrium mineral phases considered in the base case model.

| Minerals | Available Amount | SI | Log K | Remarks |
|---|---|---|---|---|
| $Fe(OH)_3$ | 0.00759 mol/kg | 0 | 0.239[a] | From Ng et al. (2015) |
| Siderite | 0 | 0 | -9.35[b] | Allowed to precipitate |
| Pyrite | 0 | 0 | -18.479 | Allowed to precipitate |
| Calcite | 0.1316 mol/kg | 0.05 | -8.48 | Estimated from Ng et al. (2015) |

[a] *Calibrated by Ng et al. (2015)*

[b] *Adjusted in PHREEQC-2*

**Table A4**. Scenarios considered in the scenario simulation with different aquifer properties and hydrogeologic conditions.

| **Scenarios (Symbol)** | **Units** | **Lower Value** | **Base Case** | **Higher Value** | |
|---|---|---|---|---|---|
| Water table fluctuation (amplitude) (WF) | - | - | 0.25 | 0.5 | 0.75 |
| Mean value of hydraulic conductivity (HK) | m/day | 3.1 | 6.2 | 9.3 | |
| Electrical conductivity of aquifer (EC) | µS/cm | - | 339 | 1078 | 2039 |
| Dissolved oxygen concentration (DO) | mg/L | 0 | 7.87 | - | |
| Nitrate concentration ($NO_3$) | mg/L | 0 | 5 | 50 | |
| Sulfate concentration ($SO_4$) | mg/L | 0 | 50 | 250 | |
| Calcite content of aquifer (Cal) | - | no calcite | with calcite | - | |
| Exchangeable cation capacity (CEC) | mol/kg | 2.19 | 10.94 | 54.70 | |
| Amorphous ferric hydroxide (FHO) | mmol/kg | 0.759 | 7.59 | | |

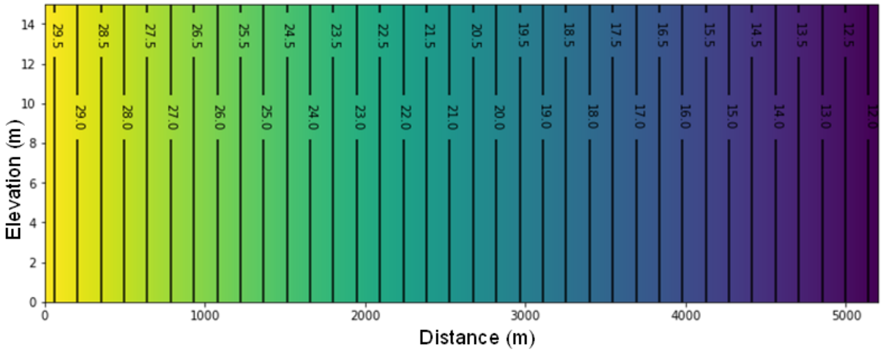


**Figure A1**. Initial steady-state groundwater head distribution of the larger model used to determine the water table fluctuation at the right boundary of the base case model.

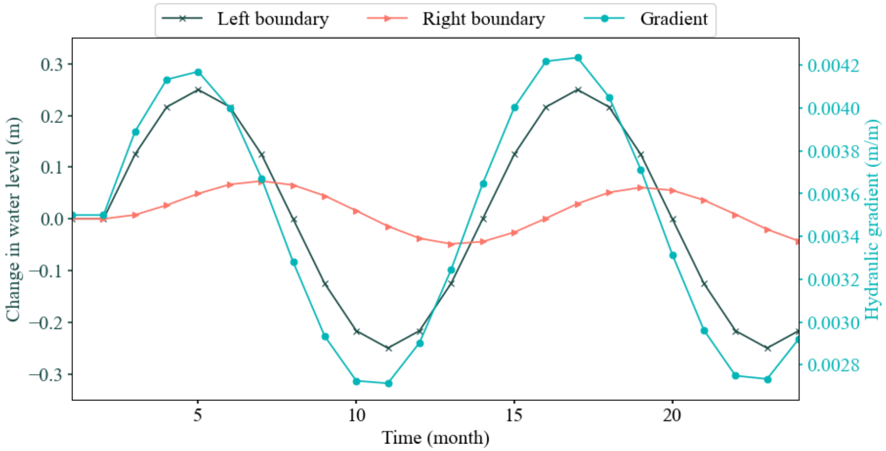


**Figure A2**. Monthly water level variations as general head boundary condition used in the base case model for 100 years. The sinusoidal function represents the water table fluctuation under transient groundwater flow conditions. The figure depicts a 24-month period for illustrative purposes.

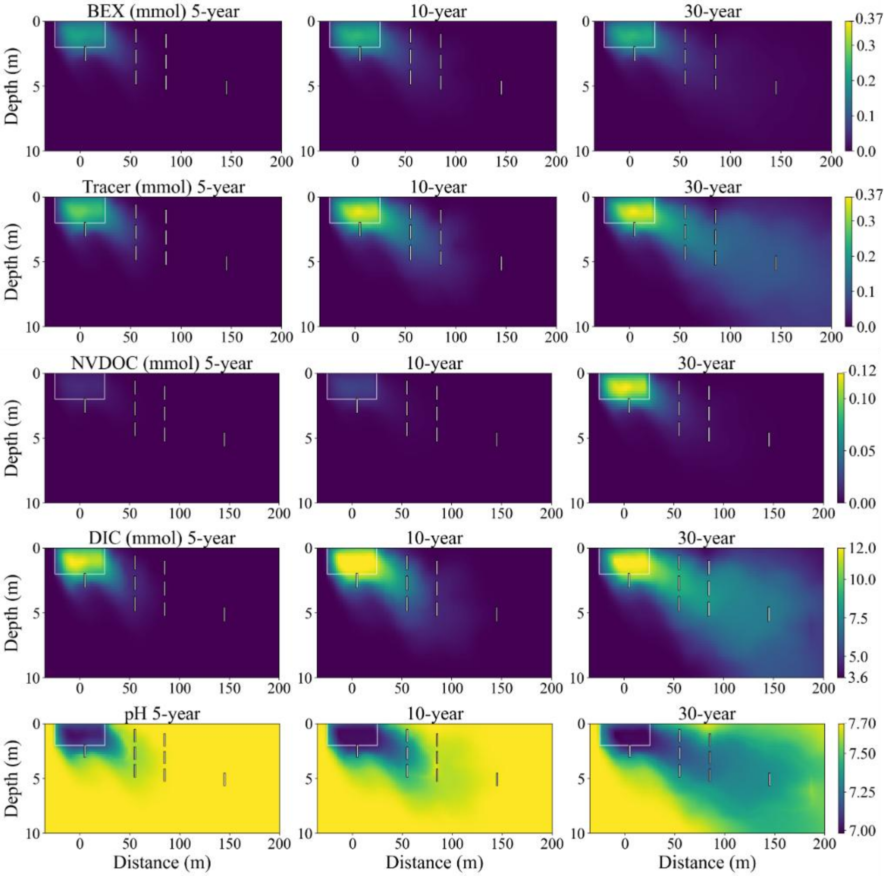


**Figure A3**. Model simulation results for major electron donors (BEX and NVDOC), the tracer representing BEX when it would not degrade, and pH, at three moments in time. DIC is produced by all TEAPs. The source zone is delineated by white rectangles.

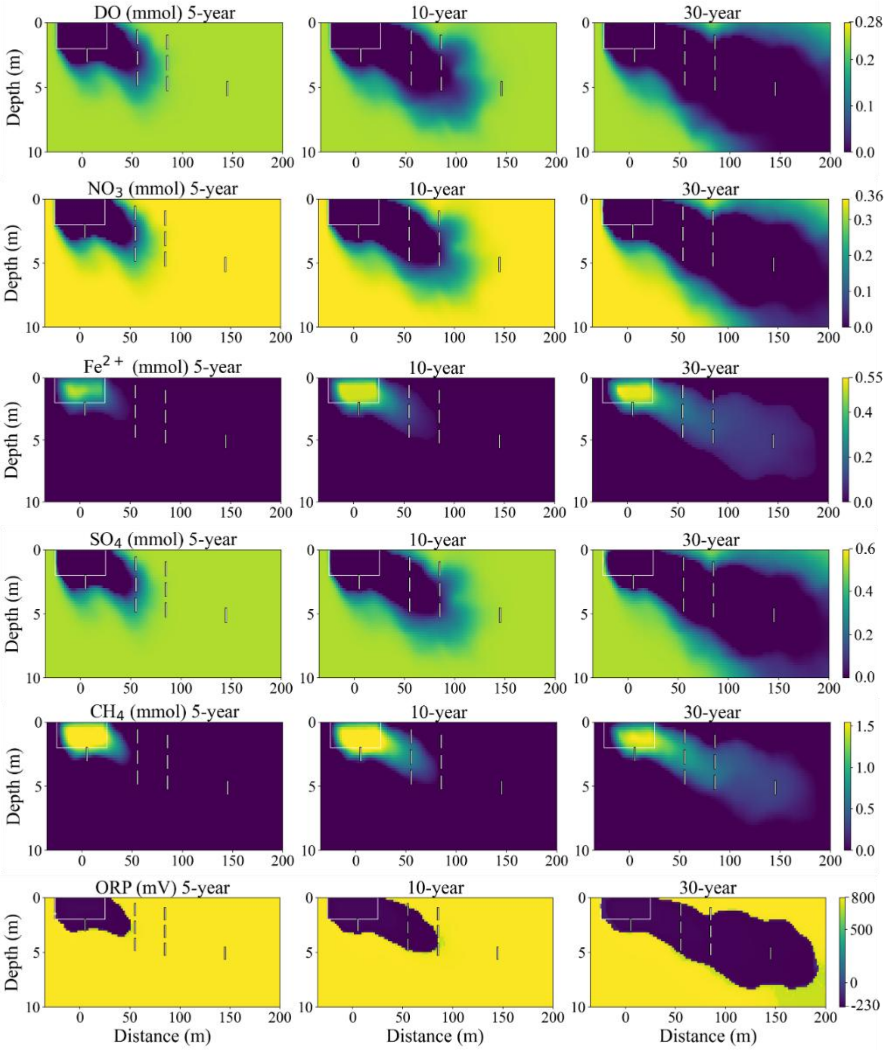


**Figure A4**. Simulated plumes of depleted electron acceptors with redox zones development (ORP plume) due to hydrocarbon degradation at three moments in time. The source zone is delineated by white rectangles.

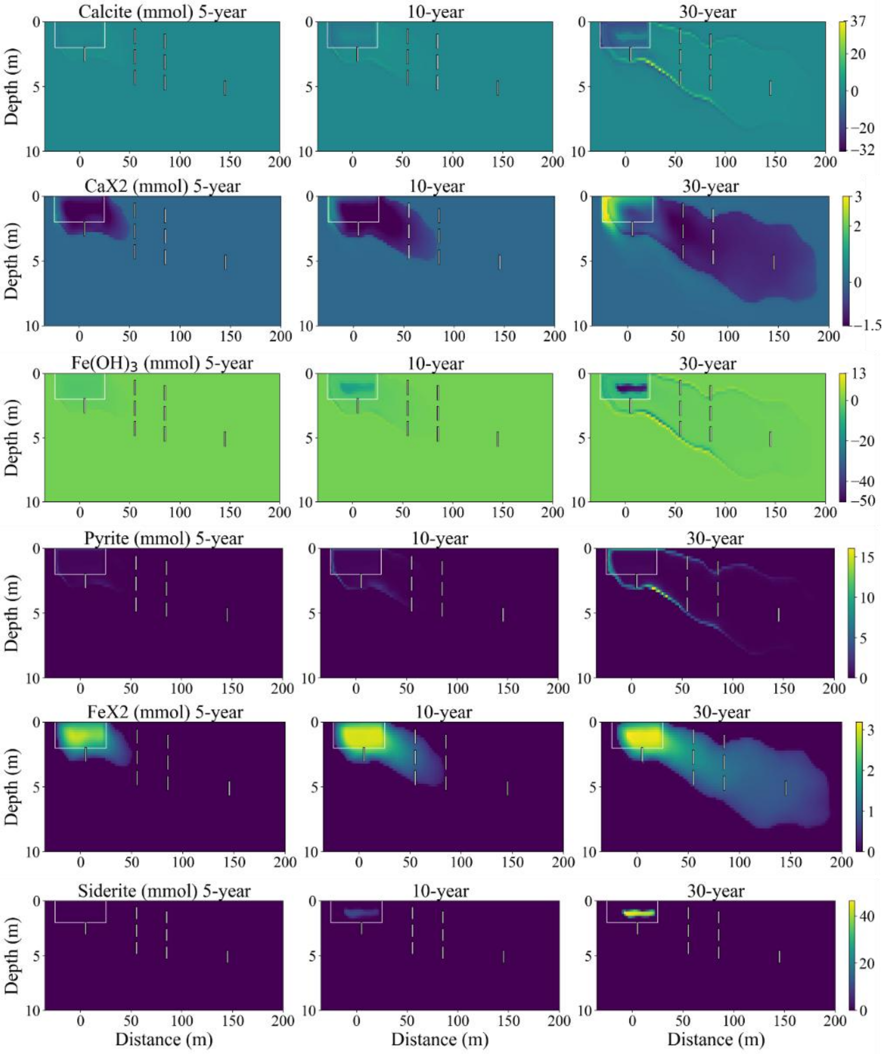


**Figure A5**. Simulation results at three distinct moments in time for solid phase minerals and cation exchangers (X represents a sorption site). $Fe(OH)_3$ represents the $Fe^{3+}$ electron acceptor, while FeX2 and the precipitated siderite ($FeCO_3$) and pyrite ($FeS_2$) represent the reduced solid phase $Fe^{2+}$. The plumes are depicted as the difference between the final and initial values. The source zone is delineated by white rectangles.

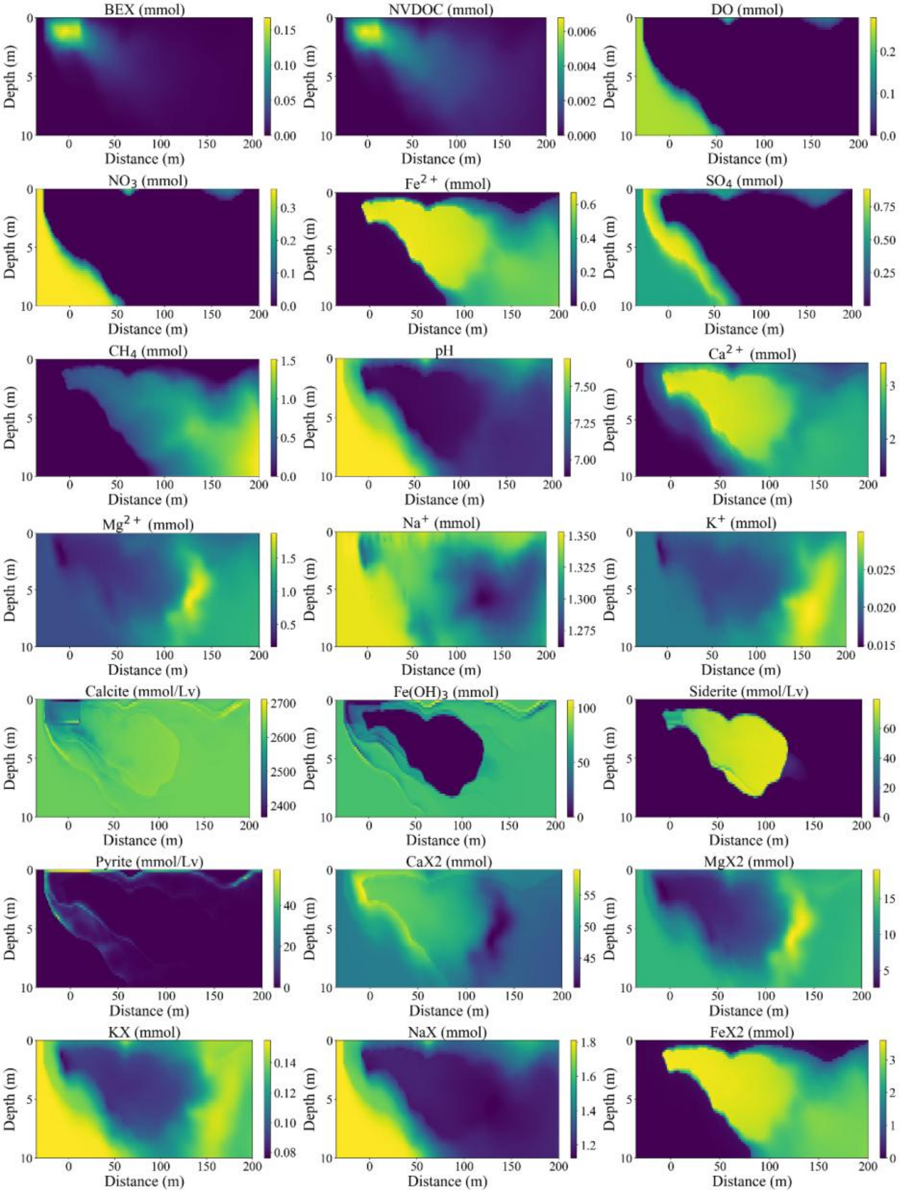


**Figure A6**. Full plume model simulation results for petroleum hydrocarbons BEX and NVDOC, electron acceptors, pH, major cations, reduced aqueous species from anaerobic terminal electron-accepting processes, and solid phase components after 100 years. X is the cation exchanger that represents sorption.

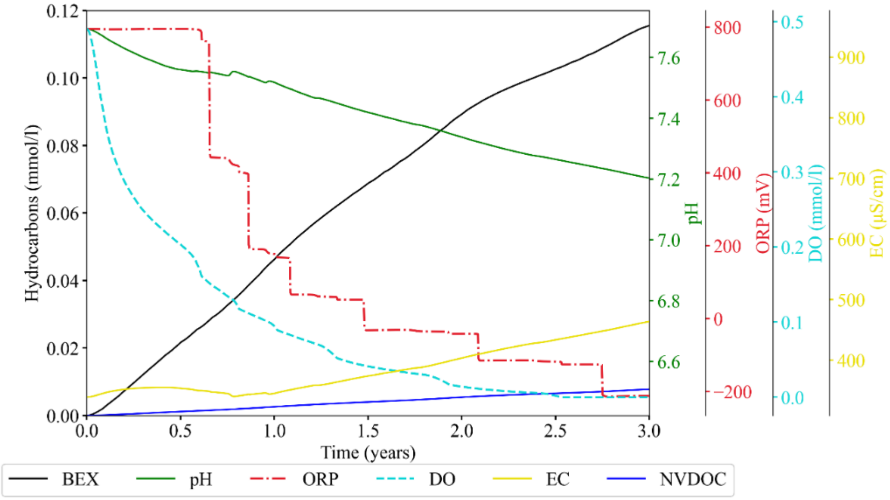


**Figure A7**. Water quality parameters (pH, ORP, DO, and EC) and petroleum hydrocarbon concentration (BEX and NVDOC) at observation well X0Z0 during the initial 3-year simulation. DO is consumed after about 2.5 years.

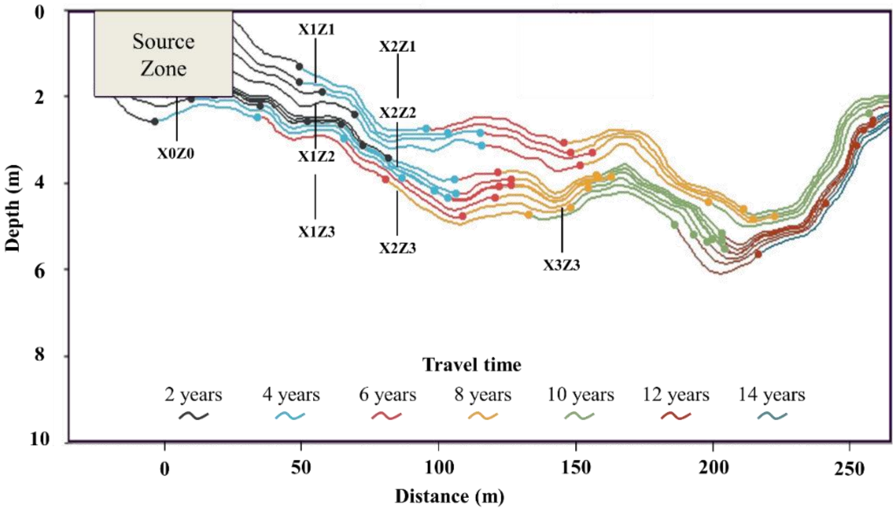


**Figure A8**. Steady state flow paths and average travel time from the edges of the source zone to the observation wells. Approximate travel times are based on steady-state particle tracking simulation results using MODPATH.

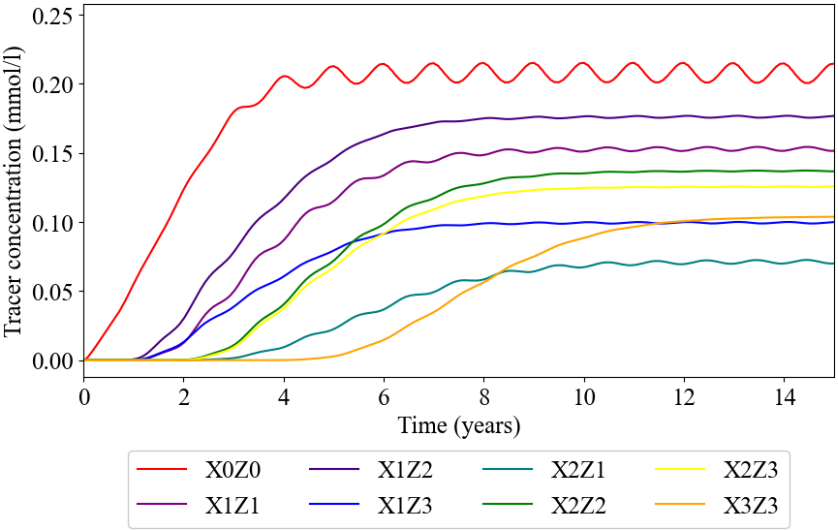


**Figure A9**. Concentration of the BEX tracer (mmol/l) at various observation wells highlighting the differences in arrival times.

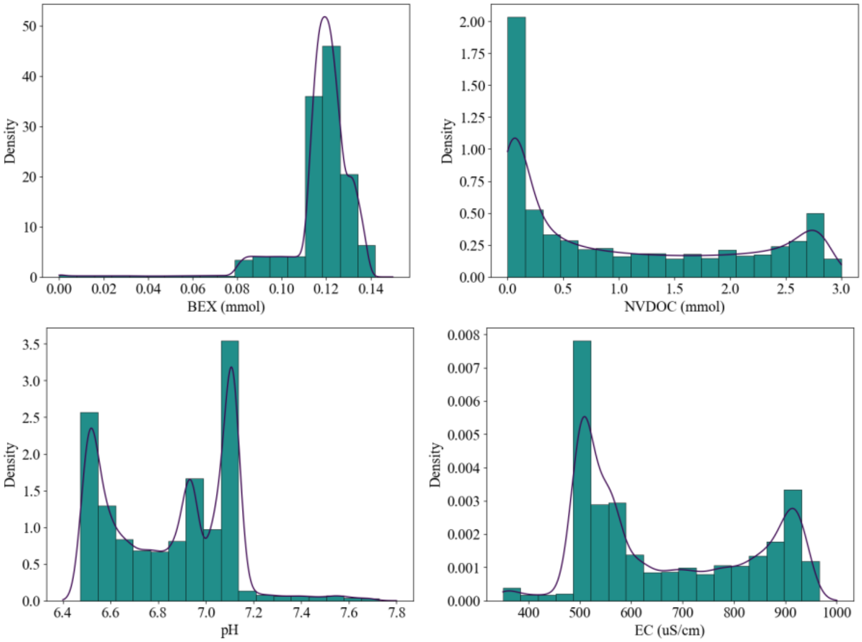


**Figure A10**. Histograms showing the KDE plot (line graph) and probability density function of petroleum hydrocarbon concentration (BEX and NVDOC) and water quality parameters (pH and EC) at observation well below the oil source zone (X0Z0).

# Appendix B

**Text B1.** Description of the five machine learning classification models that were trained and evaluated.

Using Jupyter notebook, a suite of Python libraries were employed including Scikit-learn for implementing and evaluating ML models (Pedregosa et al., 2011), Pandas for data manipulation and analysis (McKinney, 2010), NumPy for numerical computing and array operations (Harris et al., 2020), Matplotlib for data visualization (Caswell et al., 2020), and SciPy for scientific and technical computing tasks (Virtanen et al., 2020).

Logistic regression (LR) is a simple, interpretable, and computationally efficient classification algorithm. It models the probability that a given input belongs to a particular class (Starbuck 2023). This is achieved by fitting a logistic function to the input data, making it effective for linearly separable classes. However, LR struggles with complex, non-linear relationships and requires several observations to achieve stable parameter estimates. The model also assumes a linear relationship between the independent variables and the log odds of the outcome, which might not always be valid (Bisong, 2019).

In classification, support vector machines (SVMs) find the optimal hyperplane that separates data into distinct classes. The goal is to position the hyperplane so that all data points of one class lie on one side and all data points of the other class lie on the opposite side. This is achieved by maximizing the margin between the hyperplane and the closest data points from each class, known as support vectors. SVMs implicitly project data into a higher-dimensional feature space using a kernel function, enabling effective separation of classes (Vapnik, 1982; Gualtieri et al., 1999). SVMs are particularly useful for high-dimensional data, can handle noisy data, and are robust to overfitting.

Ensemble algorithms combine predictions from multiple classifiers to improve overall performance. These techniques enhance accuracy, reduce overfitting, and provide more robust models by leveraging the strengths of diverse algorithms (Briem et al., 2002).

Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and outputs the mode of their predictions for classification tasks (Breiman, 2001). It is robust to overfitting, can handle non-linear relationships, and maintain accuracy even with high-dimensional data (Parmar et al., 2019). However, it can be computationally intensive, especially with large datasets and many trees, and may become less interpretable as the number of trees increases, making it harder to understand the decision-making process.

XGBoost or eXtreme Gradient Boosting is a highly efficient and scalable ensemble algorithm based on gradient boosting. It builds trees sequentially, where each new tree corrects errors made by the previous ones. XGBoost offers high performance and speed, efficiently handles missing data, and includes built-in regularization to prevent overfitting (Chen and Guestrin, 2016). Despite its power, XGBoost can be complex to tune, requiring careful parameter optimization to achieve the best results. It also demands significant computational resources, which can be a drawback for very large datasets.

The most frequently used type of neural network is the Multi-layer Perceptron (MLP). This deep learning model is composed of multiple layers of neurons, capable of learning complex non-linear relationships (Popescu et al., 2009). They are adaptable to various types of data and can be fine-tuned for specific tasks and architectures, making them highly flexible. However, MLPs require significant computational power and large amounts of training data to perform well. They can be prone to overfitting without proper regularization and hyperparameter tuning. Additionally, the training process can be time-consuming, especially for deep networks.

**Text B2.** Details of the model training and hyperparameter tuning process

Hyperparameter tuning is essential for optimizing ML models, as it directly influences their performance and generalization. Hyperparameters are configuration settings that govern both the learning process and the model structure (Schratz et al., 2019). Unlike model parameters, which are learned from data, hyperparameters must be predefined and optimized through experimentation (Pedregosa et al., 2011). This process involves searching for the optimal combination of hyperparameters within a given space to maximize validation performance.

In this study, a two-layer tuning approach that combines the Tree-Structured Parzen Estimator (TPE) and Leave-One-Group-Out (LOGO) Cross-validation (Refaeilzadeh et al., 2009) was used. TPE is a Bayesian optimization algorithm that refines hyperparameter selection by modeling probability distributions of performance (Watanabe, 2023). TPE from Hyperopt (Bergstra et al., 2013) was implemented, since it is more computationally efficient compared to traditional Grid Search. Grid Search tests all combinations of hyperparameters, whereas TPE explores the hyperparameter space more efficiently, resulting in faster convergence.

TPE was combined with LOGO Cross-validation to ensure the robustness of the hyperparameter tuning process. LOGO was used to train the model on diverse subsets of the data by iteratively partitioning the four virtual training observation wells. In each iteration, one well was reserved as the validation set, while the remaining three wells were used for training. For each configuration, TPE optimized the hyperparameters using the three virtual training wells, and the resulting models were evaluated on the withheld virtual validation well using a custom scorer function. The average score across splits was then used as the “loss” for that hyperparameter set. This process was repeated until each well had once served as the validation set. The best hyperparameters, presented in **Tables B4a** to **B4e**, were those that yielded the lowest average loss score across all LOGO splits.

For this study, a custom scorer function was designed for each validation set to prioritize the timely detection of BEX, while still considering the prediction accuracy. The score was computed as

$$score = |delay^{0.7} * (1 - F1)^{0.3}| ,$$

where $delay$ is the predicted minus actual arrival time of BEX at the virtual validation well, and $F1$ is the harmonic mean of *precision* and *recall*. *Precision* is the proportion of true positive predictions out of all positive predictions, while *recall* is the proportion of true positive predictions out of all actual positive instances. This custom scorer penalizes large delays in the predicted arrival of BEX (weight of 0.7) and imbalanced or inaccurate predictions *(1 –* $F1$*)* (weight of 0.3). The timing of alarm triggers were included in the evaluation criteria, instead of solely focusing on accuracy. Detecting contamination prematurely or with minimal delay can prevent significant health risks. Although too early or premature alarms and false alarms lead to increased groundwater sampling costs and are therefore undesirable, delayed contamination detection is both costly and hazardous.

Although a lower weight was assigned to the overall model accuracy ($F1$ score), since $F1$ score does not account for the temporal aspect of predictions, slightly lower overall accuracy was considered acceptable in exchange for more precise prediction timing. During training, the model was heavily penalized when BEX contamination occurred, but the model did not trigger the alarm. This custom scorer “forces” the ML models to classify correctly and to predict contamination events with better timing, enhancing their practicality for decision-making. Additionally, the scorer required at least two consecutive contamination events to trigger an alarm, which helped reduce the impact of isolated false positives. After optimizing the hyperparameters for each model, the models were trained on the full training dataset and their performance was evaluated using the test datasets.

**Table B1.** Locations of eight observation wells in the reactive transport model.

| **Well** | **Distance from Source Zone Center** | **Depth from Ground Surface** |
|---|---|---|
| X0Z0 | 5 m | 2 m – 3 m |
| X1Z1 | 55 m | 0.6 m – 1.6 m |
| X1Z2 | 55 m | 2.2 m – 3.2 m |
| X1Z3 | 55 m | 3.8 m – 4.8 |
| X2Z1 | 85 m | 1 m – 2 m |
| X2Z2 | 85 m | 2.6 m – 3.6 m |
| X2Z3 | 85 m | 4.2 m – 5.2 m |
| X3Z3 | 145 m | 4.6 m – 5.6 m |

**Table B2**. Accuracy and resolution of sensors based on manufacturer's specifications (Atlas Scientific LLC, 2022; 2024a; 2024b; 2025).

| **Sensor** | **Accuracy** | **Resolution** |
|---|---|---|
| pH | 0.002 | 0.001 |
| Dissolved oxygen | 0.2 mg/l | 0.01 mg/l |
| Electrical conductivity | 2% of reading | 0.1 µS/cm |
| Redox potential | 1 mV | 0.1 mV |

**Table B3.** Mean and standard deviation for groundwater pH, dissolved oxygen (DO) (mg/L), and electrical conductivity (EC) (µS/cm) across various uncontaminated locations in the United States, including aquifer type and monitoring well depth (U.S. Geological Survey, n.d.). Due to the lack of continuous redox potential (ORP) (mV) data in the USGS dataset, redox potential measurements were obtained from the North Ostrobothnia region of northern Finland (Lyons et al., 2023).

| Location | Aquifer Type | Well Depth | Parameter | Mean | Standard Deviation |
|---|---|---|---|---|---|
| Morrow County, Oregon | Unconfined Single Aquifer | 80 ft (24.4m) | pH | 7.52 | 0.10 |
| | | | DO (mg/L) | 0.40 | 0.54 |
| | | | EC (µS/cm) | 187.13 | 4.01 |
| Sauk County, Wisconsin | Sand and Gravel | 50 ft (15.2 m) | pH | 7.6 | 0.08 |
| | | | DO (mg/L) | 2.36 | 0.90 |
| | | | EC (µS/cm) | 621.79 | 29.27 |
| Rockingham County, N. Hampshire | Ice-Contact Deposits, Pleistocene | 83 ft (25.3 m) | pH | 7.5 | 0.09 |
| | | | DO (mg/L) | 0.13 | 0.14 |
| | | | EC (µS/cm) | 673.51 | 14.1 |
| Portage County, Wisconsin | Sand and Gravel Unconfined Single Aquifer | 34.5 ft (10.5 m) | pH | 7.3 | 0.064 |
| | | | DO (mg/L) | 9.58 | 0.36 |
| | | | EC (µS/cm) | 615.47 | 45.61 |
| North Ostrobothnia, Northern Finland | Sand/Gravel | 6 m | ORP (mV) | 333.18 | 26.46 |

**Table B4a**. Optimal hyperparameters for Logistic Regression for different scenarios (BC: base case, WF: water table fluctuation, HK: hydraulic conductivity, EC: electrical conductivity, DO: dissolved oxygen, NO3: nitrate, SO4: sulfate, Cal: calcite, CEC: cation exchange capacity, FHO: amorphous ferric hydroxide). Parameters were changed relative to BC (+ for increased, - for decreased, 0 for removed).

| **Scenario** | **C** | **solver** |
|---|---|---|
| BC | 98.788966 | newton-cg |
| WF+ | 98.788966 | newton-cg |
| WF++ | 98.788966 | newton-cg |
| HK- | 1.2606513 | newton-cg |
| HK+ | 98.788966 | newton-cg |
| EC+ | 95.219172 | newton-cg |
| EC++ | 98.788966 | newton-cg |
| DO0 | 73.401661 | newton-cg |
| $NO_3$0 | 95.219172 | newton-cg |
| $NO_3$+ | 0.0149723 | saga |
| $SO_4$0 | 0.0161472 | sag |
| $SO_4$+ | 98.788966 | newton-cg |
| Cal0 | 95.219172 | newton-cg |
| CEC- | 98.788966 | newton-cg |
| CEC+ | 95.219172 | newton-cg |
| FHO- | 98.788966 | newton-cg |

**Table B4b**. Optimal hyperparameters for Support Vector Classifier for different scenarios.

| **Scenario** | **C** | **coef0** | **degree** | **gamma** | **kernel** |
|---|---|---|---|---|---|
| BC | 98.029131 | 0.3027024 | 6 | scale | linear |
| WF+ | 99.12663 | 0.9155417 | 6 | scale | linear |
| WF++ | 98.029131 | 0.3027024 | 6 | scale | linear |
| HK- | 99.826484 | 0.180156 | 6 | auto | rbf |
| HK+ | 93.320612 | 0.6066291 | 3 | auto | linear |
| EC+ | 94.611208 | 0.0076169 | 1 | scale | rbf |
| EC++ | 98.029131 | 0.3027024 | 6 | scale | linear |
| DO0 | 71.624874 | 0.2477173 | 6 | scale | linear |
| $NO_30$ | 96.690177 | 0.709633 | 6 | scale | linear |
| $NO_3+$ | 0.0314532 | 0.5942296 | 3 | auto | linear |
| $SO_40$ | 0.0366309 | 0.7619566 | 6 | scale | linear |
| $SO_4+$ | 98.567722 | 0.2161513 | 4 | scale | linear |
| Cal0 | 98.029131 | 0.3027024 | 6 | scale | linear |
| CEC- | 96.850778 | 0.5664778 | 6 | auto | linear |
| CEC+ | 83.305352 | 0.3402919 | 6 | scale | rbf |
| FHO- | 98.029131 | 0.3027024 | 6 | scale | linear |

**Table B4c**. Optimal hyperparameters for XGBoost for different scenarios.

| **Scenario** | **gamma** | **learning_rate** | **max_depth** | **n_estimators** | **reg_alpha** | **reg_lambda** |
|---|---|---|---|---|---|---|
| BC | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| WF+ | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| WF++ | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| HK- | 3.2638489 | 0.2141989 | 10 | 152 | 0.4805572 | 3.6697721 |
| HK+ | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| EC+ | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| EC++ | 0.0081877 | 0.028657 | 21 | 97 | 0.6267611 | 7.2380802 |
| DO0 | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| $NO_3$0 | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| $NO_3$+ | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| $SO_4$0 | 0.3695827 | 0.2142934 | 13 | 128 | 0.2122171 | 5.0284444 |
| $SO_4$+ | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| Cal0 | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| CEC- | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |
| CEC+ | 0.3695827 | 0.2142934 | 13 | 128 | 0.2122171 | 5.0284444 |
| FHO- | 3.839282 | 0.1990053 | 19 | 94 | 0.620119 | 25.909809 |

**Table B4d**. Optimal hyperparameters for Random Forest Classifier for different scenarios.

| Scenario | max_depth | max_features | max_leaf_nodes | min_samples_leaf | min_samples_split | n_estimators |
|---|---|---|---|---|---|---|
| BC | 2 | sqrt | 125 | 9 | 3 | 41 |
| WF+ | 5 | None | 55 | 10 | 8 | 41 |
| WF++ | 2 | sqrt | 155 | 4 | 15 | 7 |
| HK- | 10 | None | 62 | 7 | 10 | 31 |
| HK+ | 13 | sqrt | 15 | 4 | 4 | 24 |
| EC+ | 9 | sqrt | 33 | 5 | 16 | 147 |
| EC++ | 11 | sqrt | 151 | 8 | 17 | 29 |
| DO0 | 18 | sqrt | 125 | 8 | 17 | 6 |
| $NO_3$0 | 1 | sqrt | 182 | 1 | 6 | 108 |
| $NO_3$+ | 20 | sqrt | 58 | 5 | 2 | 159 |
| $SO_4$0 | 18 | sqrt | 125 | 8 | 17 | 6 |
| $SO_4$+ | 4 | sqrt | 55 | 8 | 19 | 29 |
| Cal0 | 10 | None | 102 | 2 | 16 | 29 |
| CEC- | 18 | None | 122 | 4 | 7 | 170 |
| CEC+ | 13 | sqrt | 15 | 4 | 4 | 24 |
| FHO- | 5 | sqrt | 198 | 5 | 18 | 40 |

**Table B4e**. Optimal hyperparameters for Multi-layer Perceptron for different scenarios.

| Scenario | activation | alpha | batch_size | hidden_layer_sizes | learning_rate | learning_rate_init | momentum | solver |
|---|---|---|---|---|---|---|---|---|
| BC | tanh | 0.0312019 | 64 | (103,) | constant | 0.0067669 | 0.6690047 | lbfgs |
| WF+ | tanh | 0.0377477 | 64 | (103,) | constant | 0.0079757 | 0.6709251 | lbfgs |
| WF++ | tanh | 0.0433795 | 16 | (151,) | invscaling | 0.0002137 | 0.5087271 | lbfgs |
| HK- | tanh | 6.32E-05 | 16 | (111, 109, 57) | constant | 0.0592971 | 0.5657249 | lbfgs |
| HK+ | tanh | 0.0229649 | 64 | (133,) | constant | 0.0212277 | 0.6149177 | lbfgs |
| EC+ | tanh | 0.0997503 | 64 | (133,) | constant | 0.0098147 | 0.6232301 | lbfgs |
| EC++ | tanh | 0.0794141 | 256 | (94,) | constant | 0.004667 | 0.5651141 | lbfgs |
| DO0 | tanh | 0.0214553 | 16 | (139,) | invscaling | 0.0003475 | 0.5418823 | lbfgs |
| $NO_3$0 | tanh | 0.0474201 | 64 | (199,) | invscaling | 0.0459553 | 0.6279344 | lbfgs |
| $NO_3$+ | relu | 0.0923423 | 16 | (63,) | invscaling | 0.007083 | 0.6513219 | sgd |
| $SO_4$0 | tanh | 2.2062771 | 32 | (115, 139, 185) | invscaling | 0.0168526 | 0.6054661 | lbfgs |
| $SO_4$+ | tanh | 0.0632695 | 64 | (155,) | constant | 0.0042625 | 0.6653128 | lbfgs |
| Cal0 | tanh | 0.0398131 | 64 | (156,) | constant | 0.0102258 | 0.6989061 | lbfgs |
| CEC- | tanh | 0.0485464 | 16 | (162,) | constant | 0.0117709 | 0.6201666 | lbfgs |
| CEC+ | tanh | 0.0264337 | 64 | (63,) | adaptive | 0.0039515 | 0.7186159 | lbfgs |
| FHO- | tanh | 0.0213282 | 64 | (91,) | constant | 0.0141023 | 0.598417 | lbfgs |

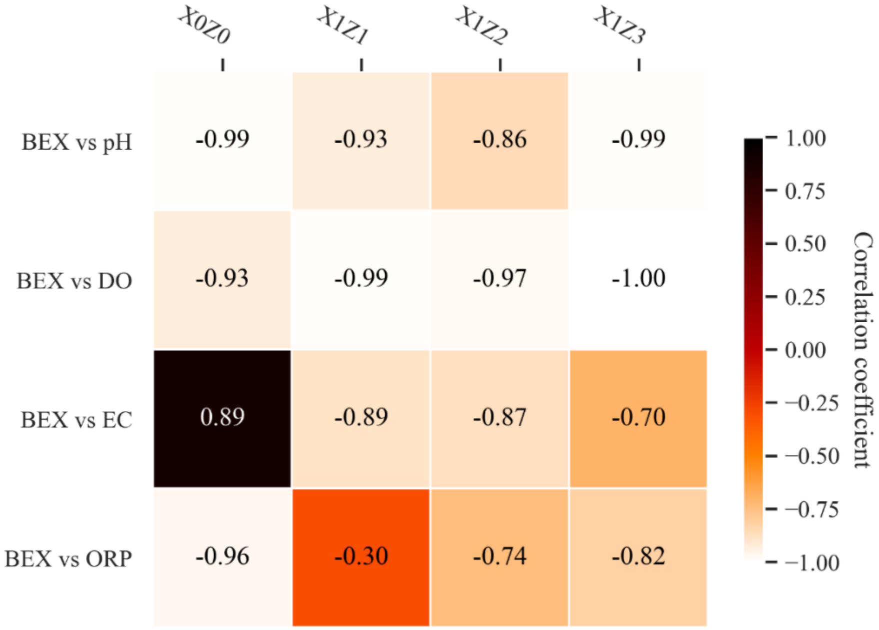


**Figure B1**. Correlation coefficients between in-situ water quality parameters and BEX at virtual observation wells near the source zone during the first three years of reactive transport model simulations (**Chapter 3**). All correlations are statistically significant (p-value < 0.05).

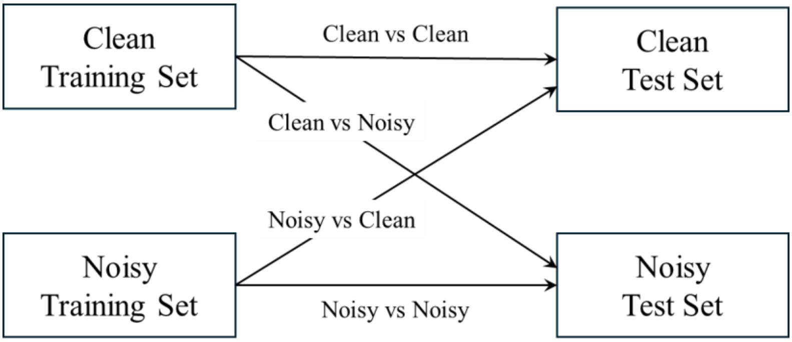


**Figure B2**. Method for evaluating the effects of noise in datasets on machine learning model performance. Note: Adapted from “Class Noise vs Attribute Noise: A Quantitative Study of Their Impacts” (Zhu and Wu, 2004).

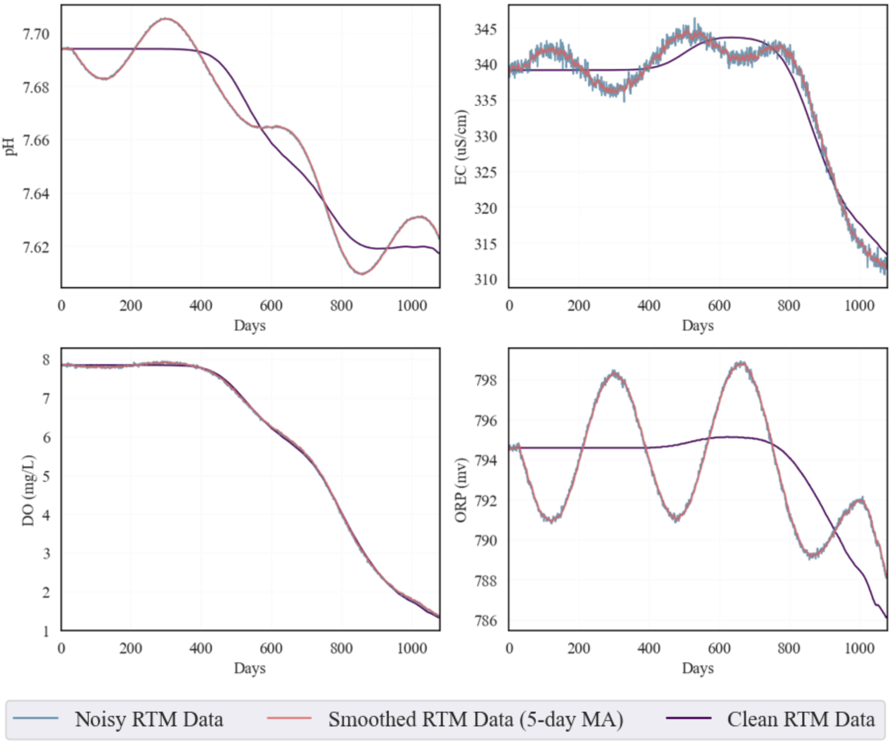


**Figure B3a**. In-situ water quality parameters from the reactive transport model (Clean RTM Data) with 10% Gaussian noise and sinusoidal fluctuations (Noisy RTM Data), then smoothed using a 5-day moving average (Smoothed RTM Data) at virtual well X0Z0.

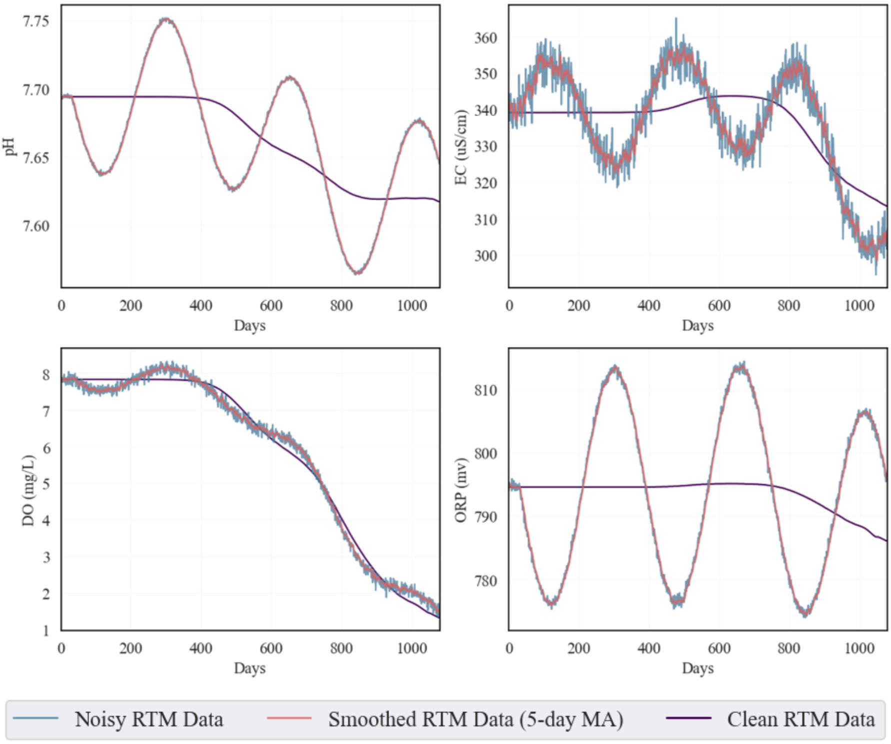


**Figure B3b**. In-situ water quality parameters from the reactive transport model (Clean RTM Data) with 50% Gaussian noise and sinusoidal fluctuations (Noisy RTM Data), then smoothed using a 5-day moving average (Smoothed RTM Data) at virtual well X0Z0.

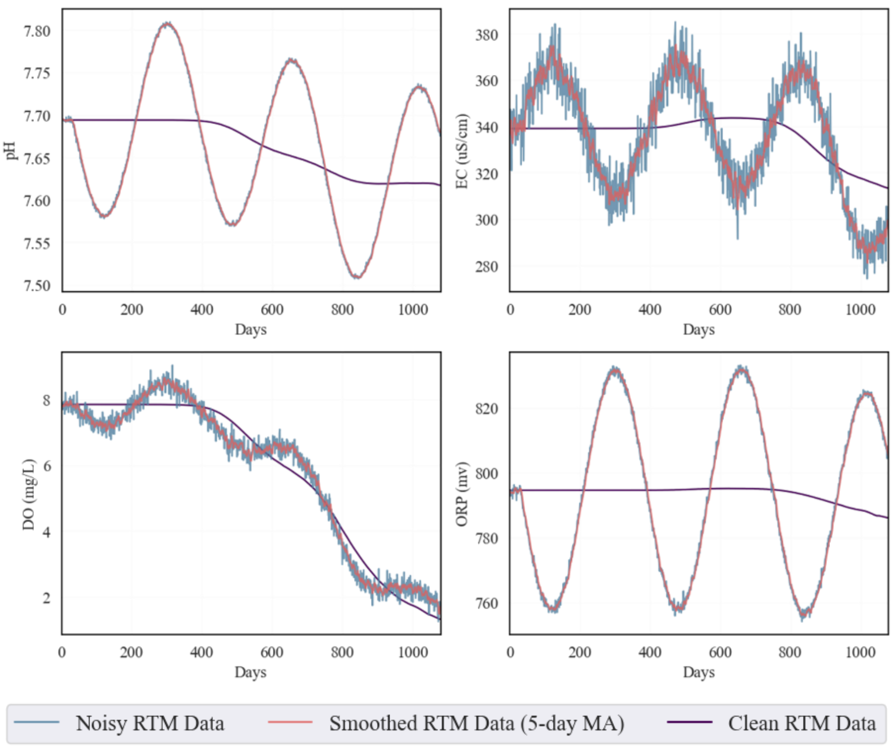


**Figure B3c**. In-situ water quality parameters from the reactive transport model (Clean RTM Data) with 100% Gaussian noise and sinusoidal fluctuations (Noisy RTM Data), then smoothed using a 5-day moving average (Smoothed RTM Data) at virtual well X0Z0.

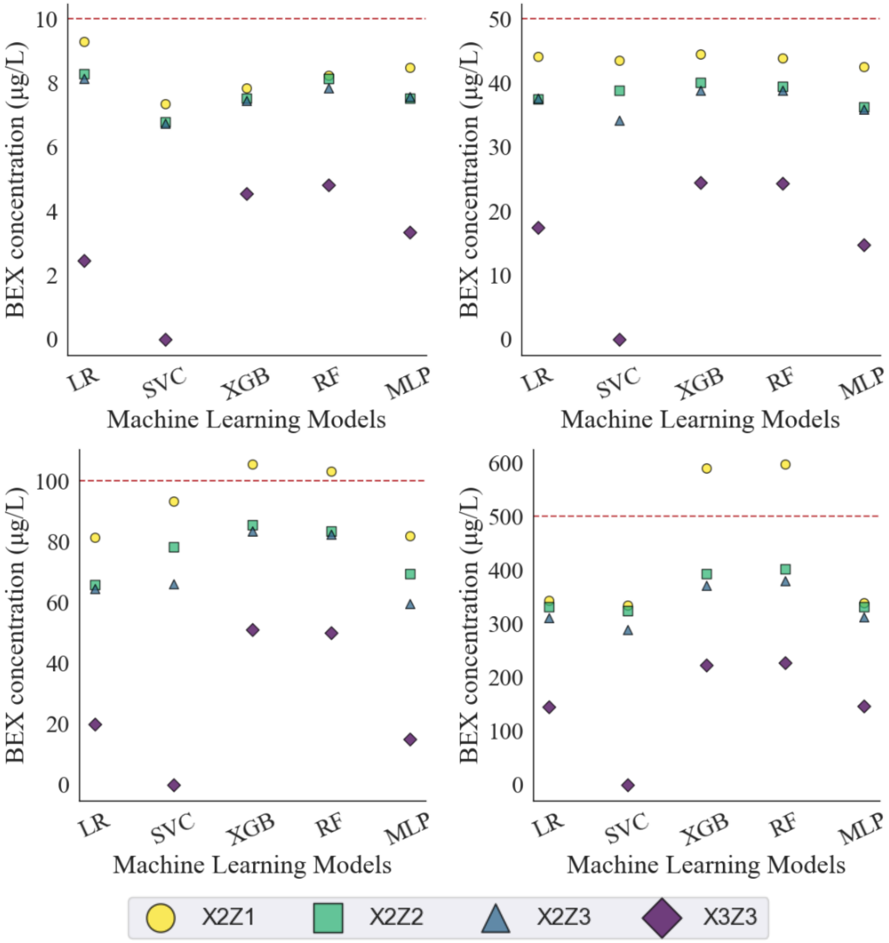


**Figure B4**. Concentrations of BEX at which different ML models triggered contamination alarms at the virtual warning wells, as contamination thresholds in the training and test data were increased from 5 µg/L to 10 µg/L, 50 µg/L, 100 µg/L, and 500 µg/L. Red dashed lines represent the contamination thresholds used for model training and evaluation.

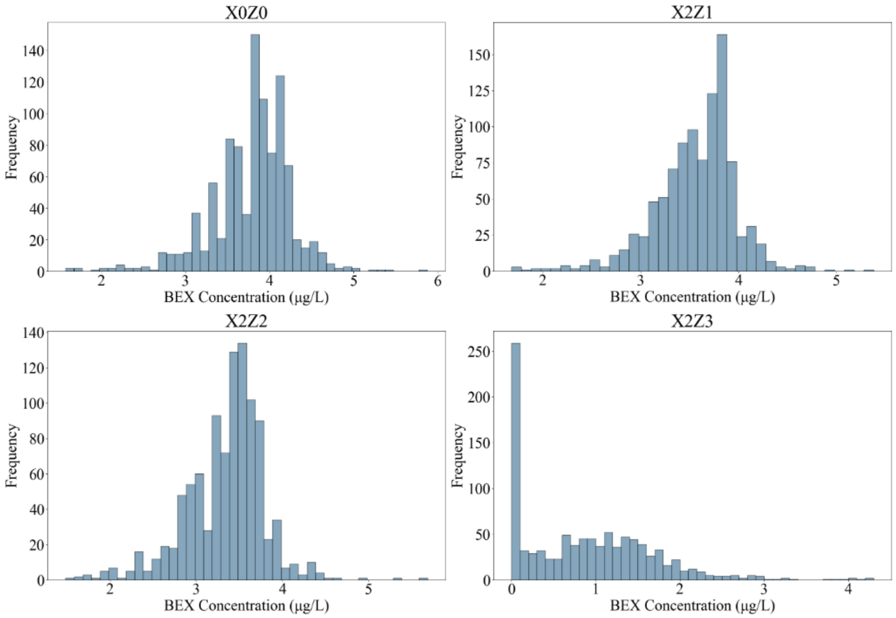


**Figure B5**. Histogram illustrating the distribution of BEX concentrations when multi-layer perceptron model triggered the contamination alarm at all virtual warning wells, with 1,000 prediction simulation runs.

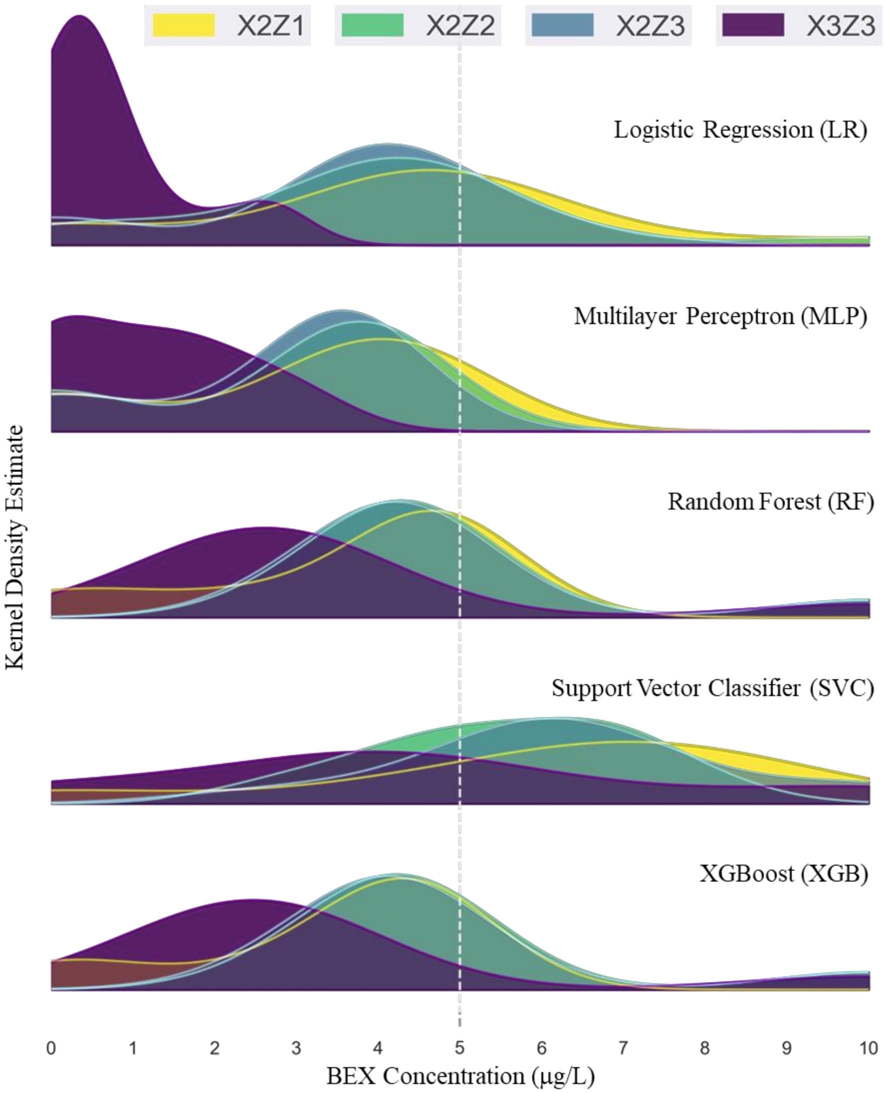


**Figure B6.** Kernel density estimate illustrating the BEX concentration at which the ML models triggered contamination alarms under the base case and across all 15 RTM scenario simulations.

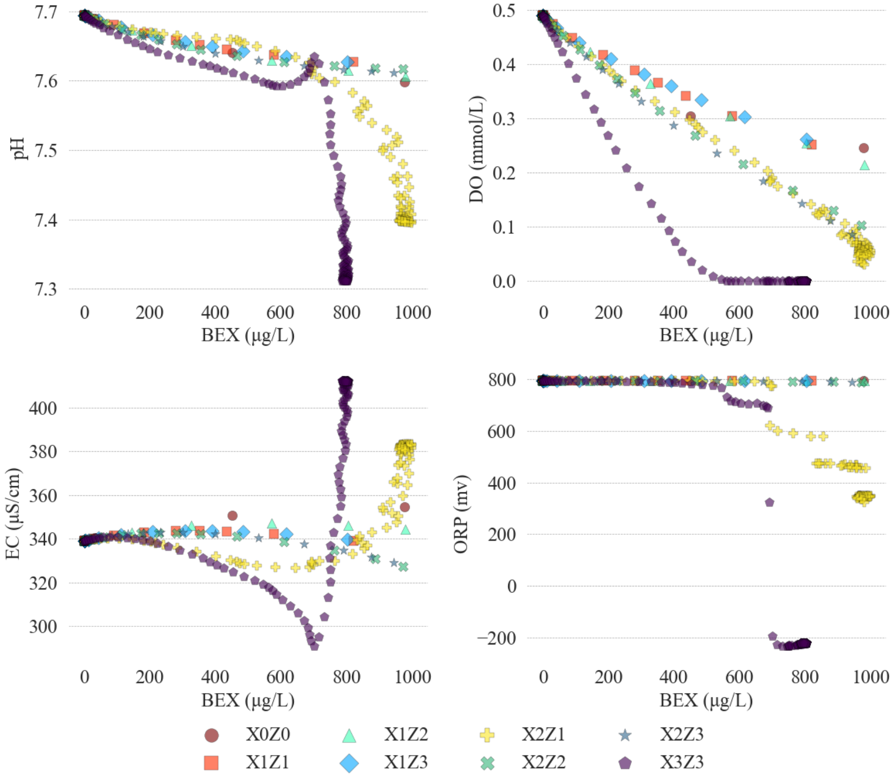


**Figure B7**. Scatter plots of four in-situ water quality parameters versus BEX concentration (μg/L) based on RTM results for BEX concentration until 1000 μg/L. Data points are plotted at 50-point intervals to reduce visual clutter while preserving trend resolution.

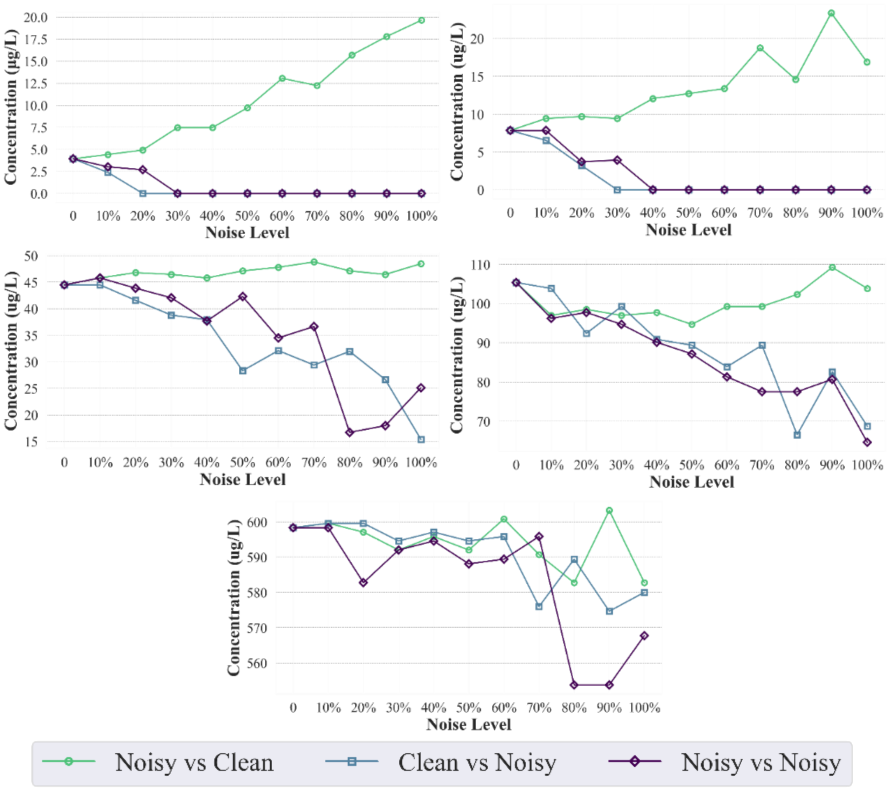
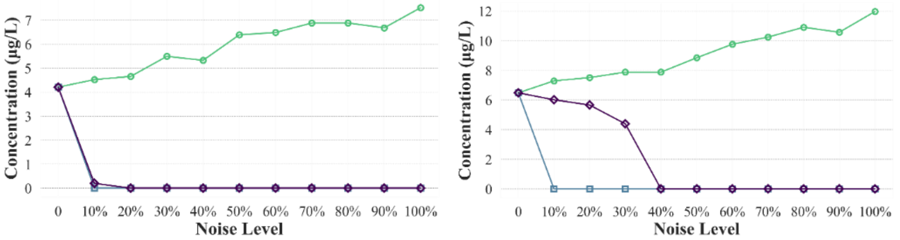
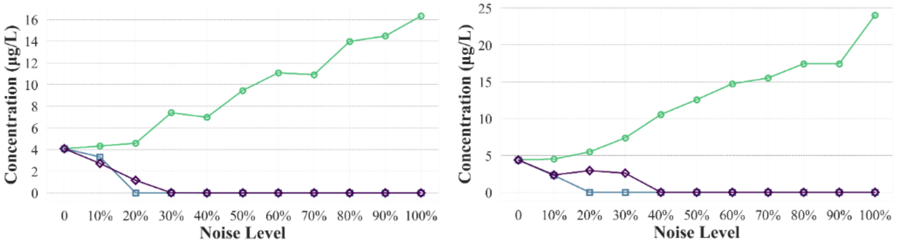


**Figure B8**. BEX concentrations (µg/L) triggering contamination alarms in the XGB model at the virtual warning well X2Z1, under varying sensor noise levels and BEX thresholds of a) 5 µg/L, b) 10 µg/L, c) 50 µg/L, d) 100 µg/L, and e) 500 µg/L: Noisy vs Clean (noisy training, clean test data), Clean vs Noisy (clean training, noisy test data), and Noisy vs Noisy (noisy training and test data).



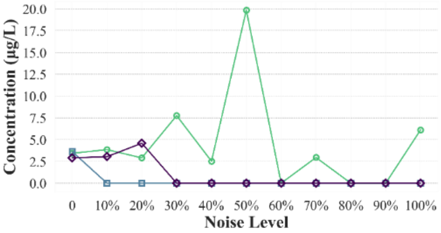
(a) Logistic Regression (LR)

(b) Support Vector Classifier (SVC)



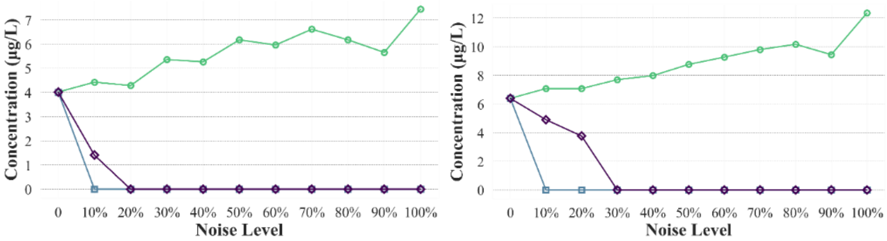
(c) eXtreme Gradient Boosting (XGB)

(d) Random Forest Classifier (RF)



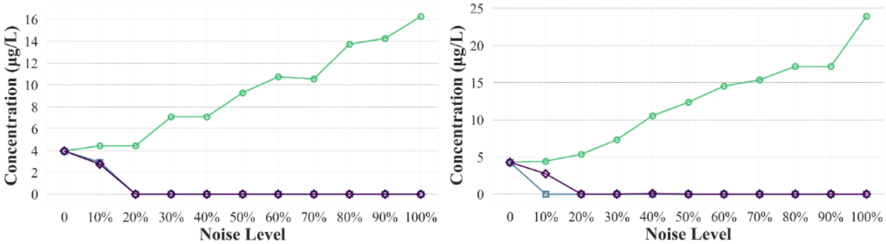
(e) Multi-Layer Perceptron (MLP)

**Figure B9a**. BEX concentrations (µg/L) at which the five ML models triggered the contamination alarm under varying sensor noise levels at the virtual warning well X2Z2: Noisy vs Clean (noisy training, clean test data), Clean vs Noisy (clean training, noisy test data), and Noisy vs Noisy (noisy training and test data).



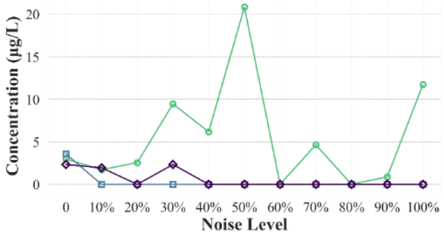
(a) Logistic Regression (LR)

(b) Support Vector Classifier (SVC)



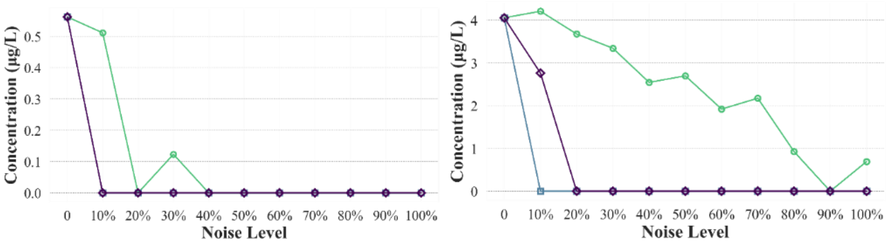
(c) eXtreme Gradient Boosting (XGB)

(d) Random Forest Classifier (RF)



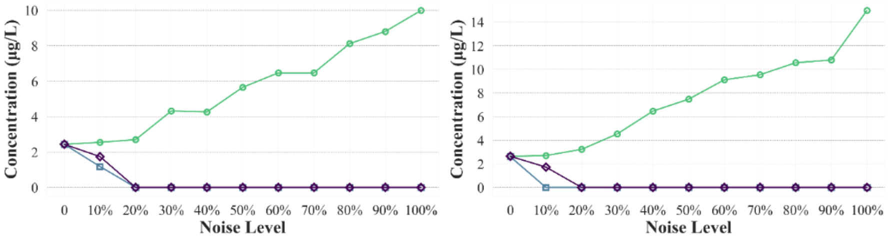
(e) Multi-Layer Perceptron (MLP)

**Figure B9b**. BEX concentrations (µg/L) at which the five ML models triggered the contamination alarm under varying sensor noise levels at the virtual warning well X2Z3: Noisy vs Clean (noisy training, clean test data), Clean vs Noisy (clean training, noisy test data), and Noisy vs Noisy (noisy training and test data).



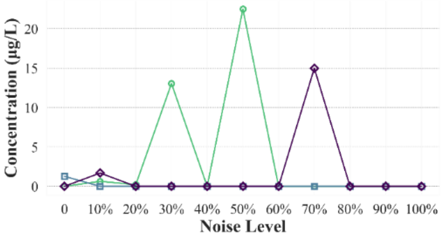
(a) Logistic Regression (LR)

(b) Support Vector Classifier (SVC)



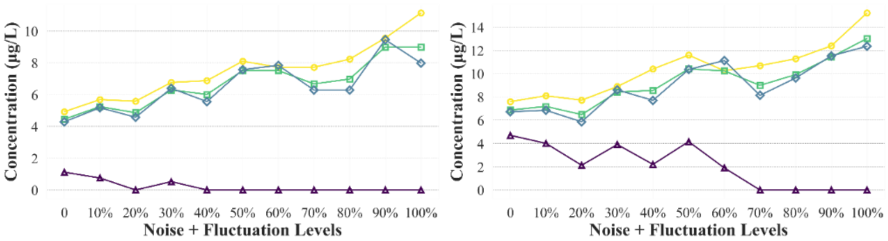
(c) eXtreme Gradient Boosting (XGB)

(d) Random Forest Classifier (RF)



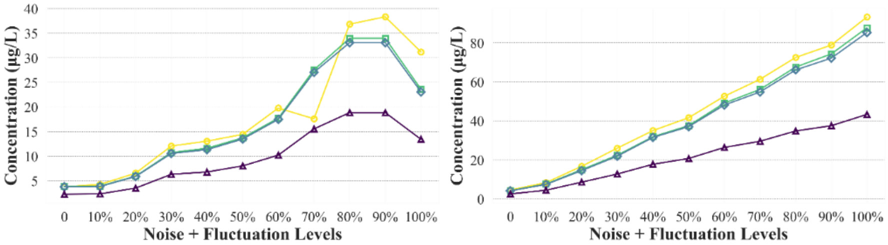
(e) Multi-Layer Perceptron (MLP)

**Figure B9c**. BEX concentrations (µg/L) at which the five ML models triggered the contamination alarm under varying sensor noise levels at the virtual warning well X3Z3: Noisy vs Clean (noisy training, clean test data), Clean vs Noisy (clean training, noisy test data), and Noisy vs Noisy (noisy training and test data).



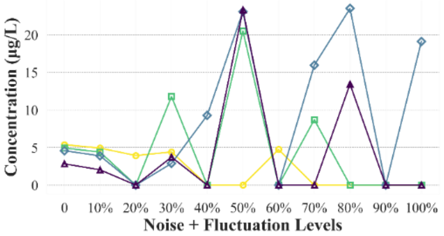
(a) Logistic Regression (LR)

(b) Support Vector Classifier (SVC)



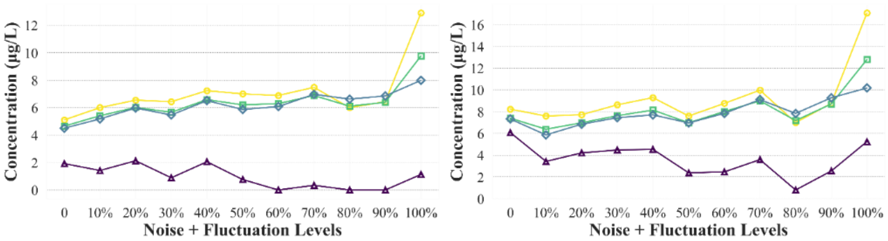
(c) eXtreme Gradient Boosting (XGB)

(d) Random Forest Classifier (RF)



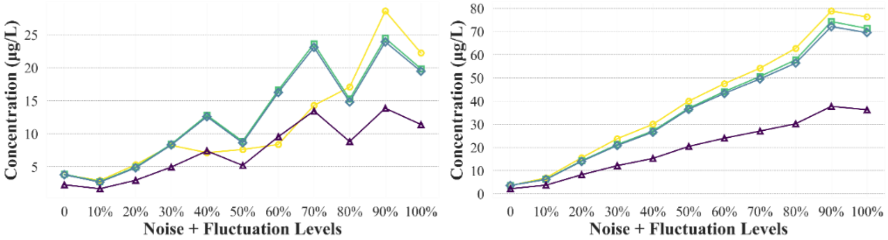
(e) Multi-Layer Perceptron (MLP)

**Figure B10a**. BEX concentrations (µg/L) at which the five machine learning models triggered contamination alarms. A 3-day moving average was applied for smoothing after adding different levels of combined sensor noise and sinusoidal fluctuations in the training data at all virtual warning wells.



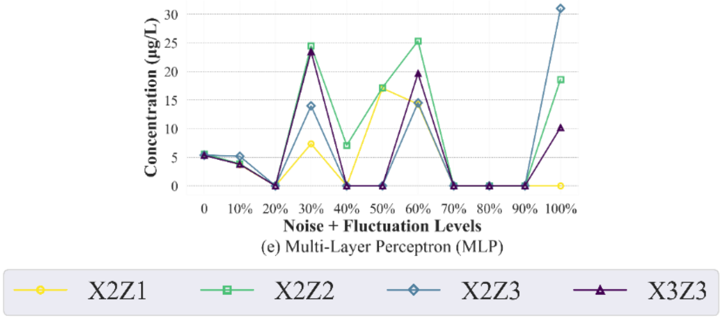
(a) Logistic Regression (LR)

(b) Support Vector Classifier (SVC)



(c) eXtreme Gradient Boosting (XGB)

(d) Random Forest Classifier (RF)



(e) Multi-Layer Perceptron (MLP)

**Figure B10b**. BEX concentrations (µg/L) at which the five machine learning models triggered contamination alarms. A 10-day moving average was applied for smoothing after adding different levels of combined sensor noise and sinusoidal fluctuations in the training data at all virtual warning wells.

# Appendix C

**Text C1**. LSTM Architecture and Hyperparameters

The LSTM model was constructed with two layers of 256 units each to capture complex temporal dependencies. A dropout rate of 0.1 was applied between layers, regularized by L2 regularization ($\lambda$=0.001), to mitigate overfitting. The learned sequences were passed to a dense feedforward layer with 32 units for final output processing. The model was trained with the Adam optimizer using a learning rate of 0.001.

**Text C2**. Hyperparameter Search Space of Classical Regression Models

The search spaces for the classical regression models were defined as follows: for support vector regression (SVR), tuning was done for the kernel function (i.e., radial basis function, polynomial and, sigmoid), the regularization parameter C from $e^1$ to $e^5$, epsilon $\varepsilon$ from $e^{-10}$ to $e^{10}$, and kernel coefficient $\gamma$ from $e^{-1}$ to $e^{10}$; for random forest (RF), optimization was done for the number of trees from 2 to 500, the maximum depth from 1 to 100, min_samples_split from 0.001 to 0.2, min_samples_leaf from 0.001 to 0.1, and the feature selection strategy (i.e., sqrt, log2, or none); for XGBoost, tuning was done for the number of estimators from 2 to 500, max depth from 1 to 100, the learning rate log-uniformly from $e^{-5}$ to $e^{-0.7}$, the subsampling ratio from 0.5 to 1.0, the column sampling from 0.5 to 1.0, and gamma from 0 to 7.

**Text C3**. Hybrid Kalman Filter Architecture

The hybrid predictive framework integrates Kalman filtering with machine learning to estimate BEX concentrations using both direct BEX and iWQPs values from the RTM (**Figure C2**). The system architecture consists of four key components: data preparation, effective attenuation rate estimation, ensemble observation modeling, and Kalman filtering with rolling window validation. Daily time-series from the RTM were compiled for BEX concentrations and four iWQPs at each observation well.

To model the attenuation of BEX following source removal, a first-order kinetic model was assumed, governed by the differential equation

$$dC_{bex}/dt = -\lambda C_{bex},$$

where $C_{bex}$ is the BEX concentration and $\lambda$ is the effective attenuation rate constant. Because of time scale decomposition, the effective attenuation rate was estimated per 5-year training window using constrained nonlinear least-squares fitting of the exponential model:

$$C_{bex,t} = (C_{bex,t-1}) \cdot e^{-\lambda_{t-1}\Delta t},$$

implemented via SciPy's curve fit function with bounds to prevent non-physical negative values. It was assumed that λ approximates the combined effects of advective transport and biodegradation.

Residuals from the fitting process were used to derive the process noise covariance matrix Q, which incorporated both concentration variance and λ uncertainty. An off-diagonal term was included to reflect the expected negative correlation between these variables.

For observation modeling, an ensemble approach combining two SVR models with an RF regressor was employed. One of the SVR had a radial basis function kernel (kernel='rbf', C=100, epsilon=0.01) while the other has a second-degree polynomial kernel (kernel='poly', degree=2, C=100). These ML models were selected to provide a balance between nonlinear modeling capability and methodological simplicity, as the KF component was intended primarily as a proof-of-concept rather than a comprehensive comparison of ML approaches. The RF regressor has n_estimators=100. These models were trained to estimate BEX concentrations from iWQPs when direct BEX measurements were unavailable. Prior to training, all input features were standardized using StandardScaler() from Scikit-learn. The prediction uncertainty of the ensemble, quantified as the standard deviation of the training residuals, was used to dynamically define the uncertainty for the BEX estimate in the observation covariance matrix R. For the physically measured iWQPs, the diagonal elements of R were set to the reported accuracies of commercial in-situ water quality sensors (**Table C1**).

The KF tracked both BEX concentration and effective attenuation rate in a two-dimensional state vector $[C_{bex}, \lambda]$. The state transition matrix encoded the first-order kinetic model using the analytical solution of the governing equation, while allowing sensitivity to variations in λ. Unlike classical KFs that rely on a fixed observation model $H$, this implementation uses an ensemble of ML models to estimate BEX from iWQPs when direct measurements are missing. In this hybrid configuration, the ensemble ML model predictions are treated as surrogate observations with dynamically computed uncertainty, which is incorporated into the measurement noise covariance matrix R. A sparse observation matrix $H_k$ is defined to link the predicted BEX to the state vector, enabling computation of the Kalman gain even in the absence of direct physical mapping. This approach allows the filter to adaptively balance model predictions and ensemble ML model estimates based on their respective uncertainties. Additional enhancements included adaptive scaling of process noise based on the time elapsed since the last measurement, Mahalanobis distance-based outlier detection (Chang, 2014), and numerical stabilization of the covariance matrix through eigen decomposition and regularization.

**Table C1**. Accuracy and price of commercial sensors based on manufacturer's specifications (Atlas Scientific LLC, 2022; 2024a; 2024b; 2025; van Essen Instruments 2025a; 2025b).

| Sensor | Accuracy | Estimated Price |
|---|---|---|
| pH | 0.002 | €72 |
| Dissolved oxygen | 0.2 mg/l | €170 |
| Electrical conductivity | 2% of reading | €94 |
| Redox potential | 1 mV | €72 |
| TD-Diver (water level) | ±0.5 cm $H_2O$ | €558 |
| Baro-Diver (atmospheric pressure) | ±0.5 cm $H_2O$ | €470 |

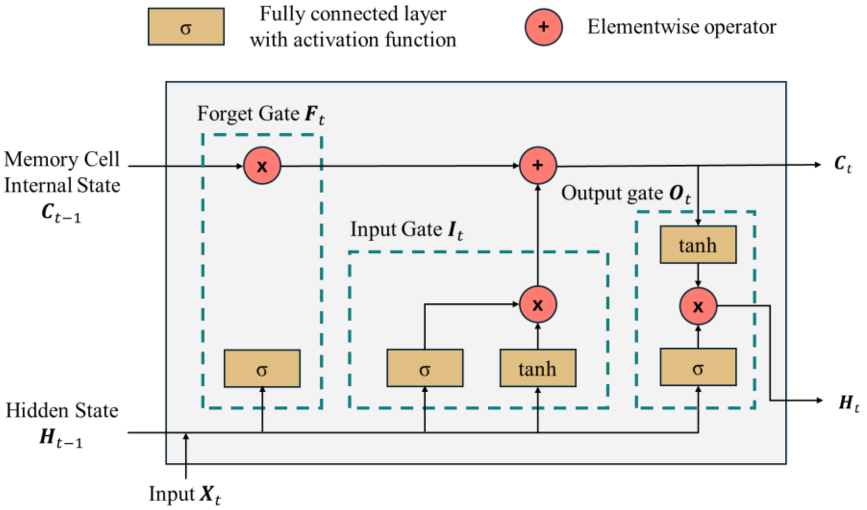


**Figure C1.** Framework of the LSTM model showing the fully connected layers with activation functions, gated mechanisms, and the computation of the hidden state. Modified from Zhang et al. (2023).

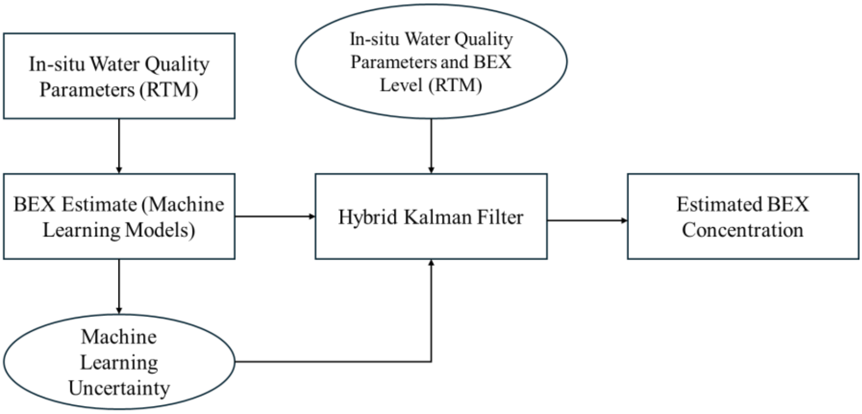


**Figure C2**. Schematic diagram of the hybrid Kalman filter framework. Monthly BEX levels were considered as input for the filter.

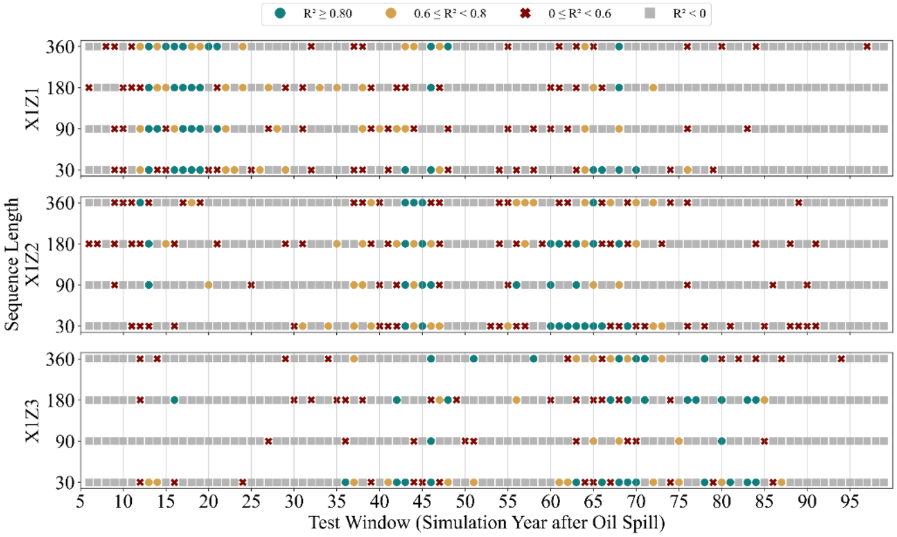


**Figure C3**. LSTM model predictions for observation wells X1Z1, X1Z2, and X1Z3 across time windows and sequence lengths of 30, 90, 180, and 360 days. Predictions with $R^2 \geq 0.80$ are shown as green circles, those with $0.60 \leq R^2 < 0.80$ as golden circles, and those with $0 \leq R^2 < 0.60$ as red crosses. Negative $R^2$ values were set to 0 and are shown as grey squares. **Figures C4** and **C5** reuse the legend from this figure for consistency across panels.

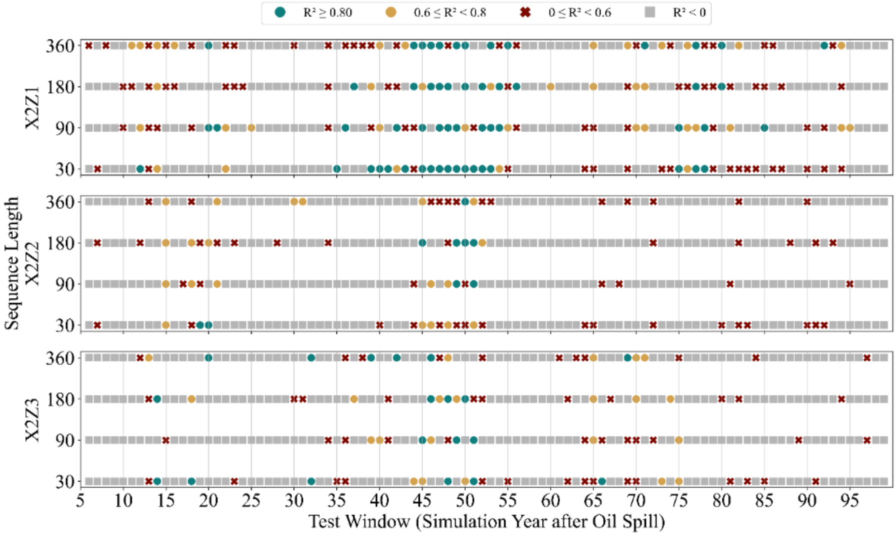


**Figure C4**. LSTM model predictions for observation wells X2Z1, X2Z2, and X2Z3 across time windows and sequence lengths of 30, 90, 180, and 360 days.

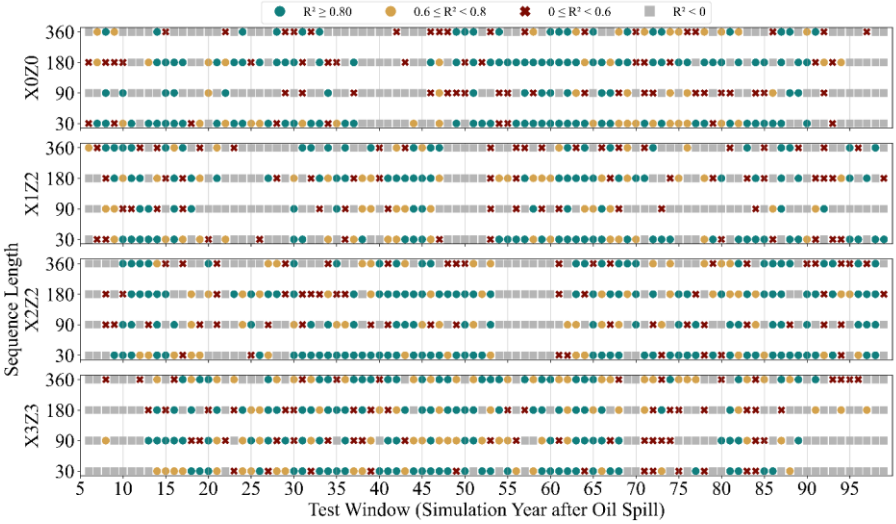


**Figure C5**. LSTM model predictions with hydraulic head as additional input parameter for observation wells X0Z0, X1Z2, X2Z2, and X3Z3 across time windows and sequence lengths of 30, 90, 180, and 360 days.

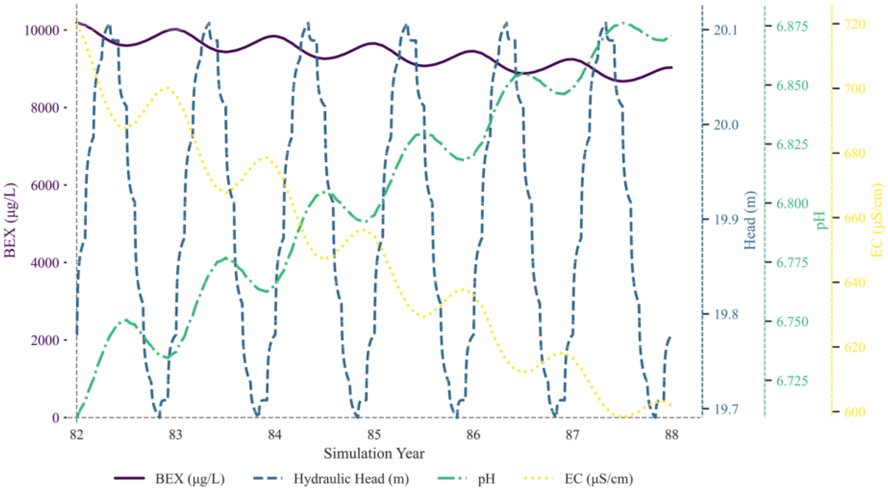


**Figure C6**. BEX concentrations with hydraulic head, pH, and EC at well X0Z0 during simulation years 82–88.

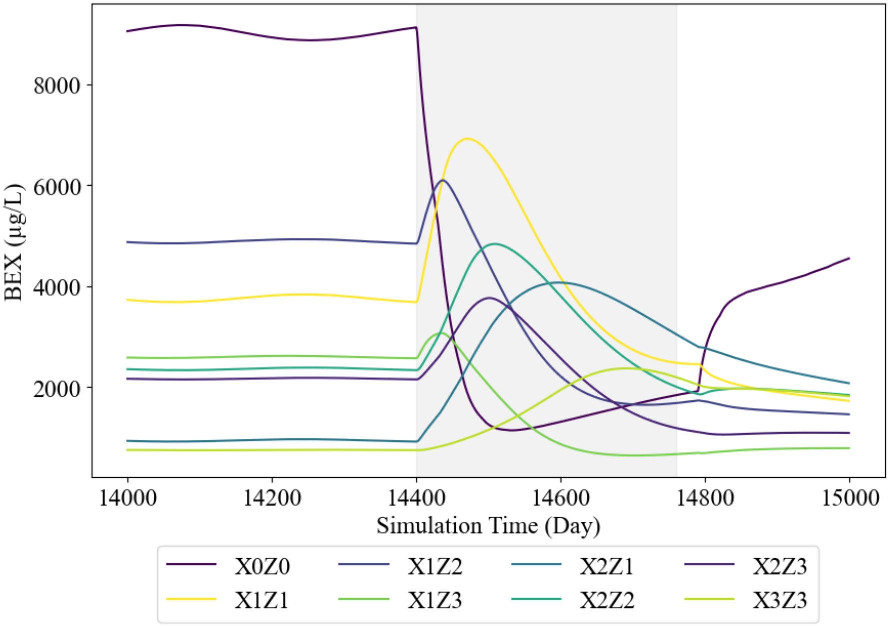


**Figure C7**. BEX concentration across eight observation wells with a year of increased hydraulic gradient starting from simulation year 40 (shaded region).

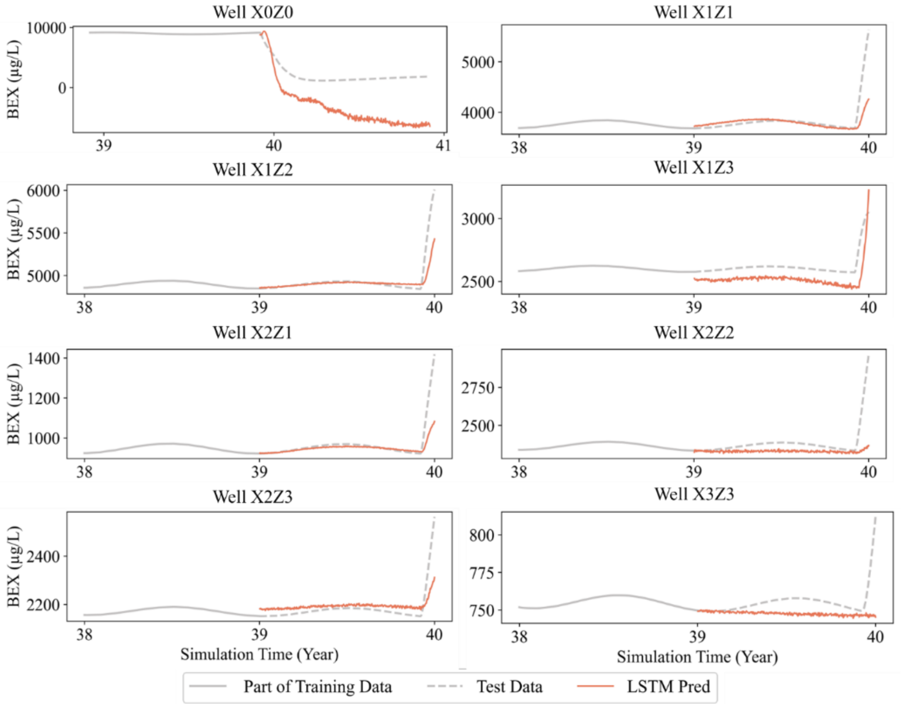


**Figure C8**. LSTM-predicted BEX concentrations under increased hydraulic gradient, showing one year of training data prior to the gradient increase and the subsequent one-year test period across eight observation wells.

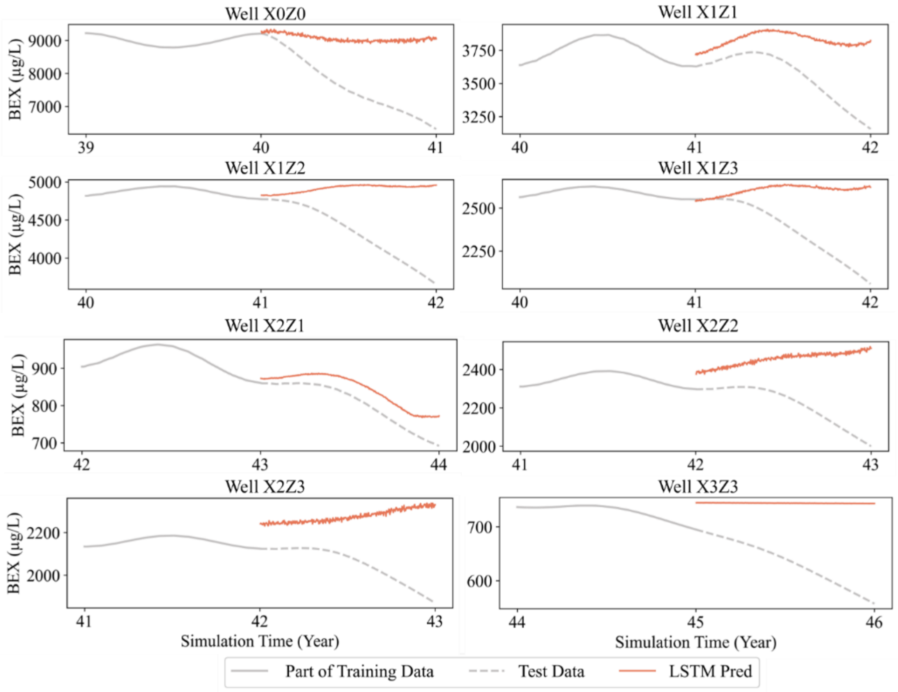


**Figure C9**. LSTM-predicted BEX concentrations after LNAPL source removal, showing one year of training data prior to the source removal and the subsequent one-year test period across eight observation wells.

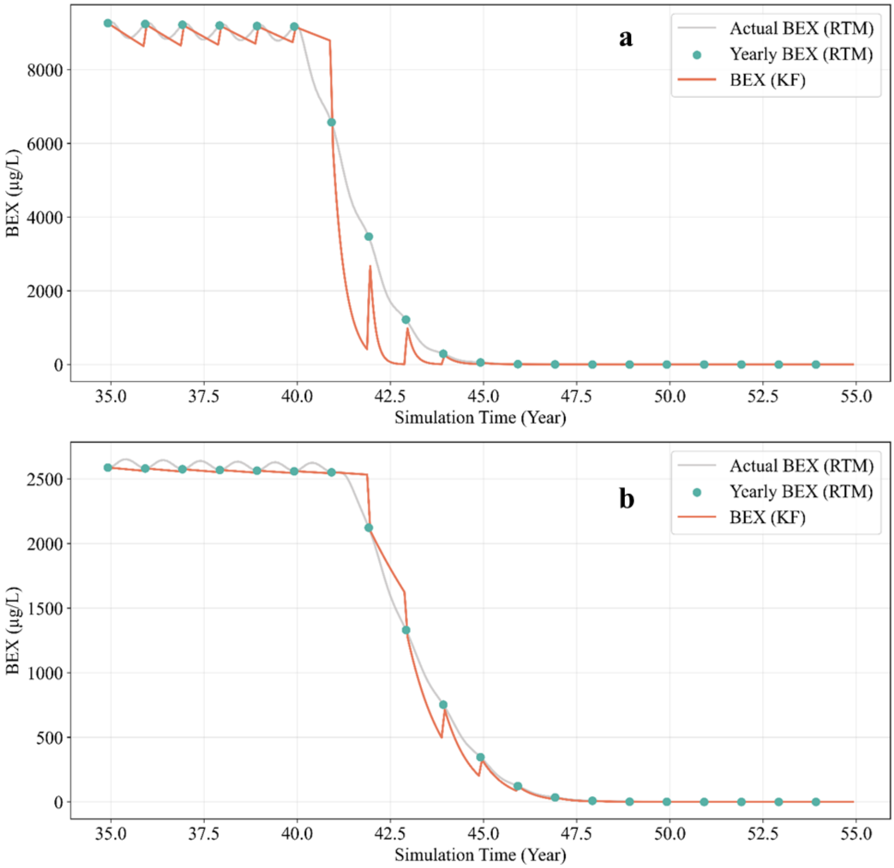


**Figure C10**. Sample Kalman filter results at observation wells X0Z0 (a) and X1Z3 (b) after LNAPL source removal.

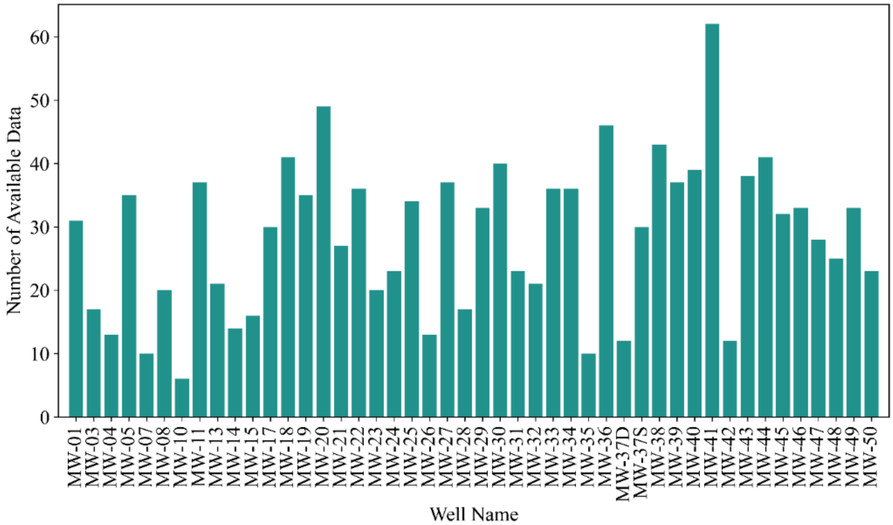


**Figure C11**. Number of available field data from an oil company containing both BTEX and in-situ water quality measurements at multiple monitoring wells from 2002 to 2020.

# References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Brain, G. (2016). TensorFlow: A System for Large-Scale Machine Learning. In Business Opp (Vol. 10, Issue July). https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

Abascal, E., Gómez-Coma, L., Ortiz, I., & Ortiz, A. (2022). Global diagnosis of nitrate pollution in groundwater and review of removal technologies. Science of The Total Environment, 810, 152233. https://doi.org/10.1016/J.SCITOTENV.2021.152233

ABB (n.d.). UviTec Benzene Sensor. Retrieved from https://new.abb.com/products/measurement-products/analytical/continuous-water-analysis/benzene-monitoring/uvitec-benzene-sensor on (18 February 2025)

Abbas, M., Jardani, A., Soueid Ahmed, A., Revil, A., Brigaud, L., Bégassat, P., & Dupont, J. P. (2017). Redox potential distribution of an organic-rich contaminated site obtained by the inversion of self-potential data. Journal of Hydrology, 554, 111–127. https://doi.org/10.1016/J.JHYDROL.2017.08.053

Agency for Toxic Substances and Disease Registry (2007). Toxicological profile for Benzene. U.S. Department of Health and Human Services. Public Health Service. Division of Toxicology and Environmental Medicine/Applied Toxicology Branch

Ahmadi, H., Kilanehei, F., Nazari-Sharabian, M. (2021). Impact of Pumping Rate on Contaminant Transport in Groundwater—A Numerical Study. Hydrology, 8, 103. https://doi.org/10.3390/hydrology8030103

Ahsan, M.M., Mahmud, M.A.P., Saha, P.K., Gupta, K.D., & Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. Technologies 2021, Vol. 9, Page 52, 9(3), 52. https://doi.org/10.3390/TECHNOLOGIES9030052

Alazaiza, M. Y. D., Ramli, M. H., Copty, N. K., & Ling, M. C. (2021). Assessing the impact of water infiltration on LNAPL mobilization in sand column using simplified image analysis method. Journal of Contaminant Hydrology, 238. https://doi.org/10.1016/j.jconhyd.2021.103769

Alazaiza, M. Y. D., Ramli, M. H., Copty, N. K., Sheng, T. J., & Aburas, M. M. (2020). LNAPL saturation distribution under the influence of water table fluctuations using simplified image analysis method. Bulletin of Engineering Geology and the Environment, 79(3), 1543–1554. https://doi.org/10.1007/S10064-019-01655-3

Allen, L., Cohen, M. J., Abelson, D., & Miller, B. (2012). Fossil Fuels and Water Quality. The World's Water, 73–96. https://doi.org/10.5822/978-1-59726-228-6_4

ALS Global. (n.d.) Petrol Pack Basic: BTEX and TPH [OV-20h] in Water. Retrieved from https://www.alsglobal.no/en/package/environment_1/water_2/combination-packages_8/petrol-pack-basic-btex-and-tph-ov-20h-in-water_44640 (on 28 September 2025)

Amat Rodrigo, J., & Escobar Ortiz, J. (2024). skforecast (v0.17.0). Retrieved from https://skforecast.org/0.17.0/introduction-forecasting/introduction-forecasting (on 10 August 2025)

Amos, R. T., Bekins, B. A., Delin, G. N., Cozzarelli, I. M., Blowes, D. W., & Kirshtein, J. D. (2011). Methane oxidation in a crude oil contaminated aquifer: Delineation of aerobic reactions at the plume fringes. Journal of Contaminant Hydrology, 125(1–4), 13–25. https://doi.org/10.1016/j.jconhyd.2011.04.003

Amos, R.T., Mayer, K.U., Bekins, B.A., Delin, G.N., & Williams, R.L. (2005). Use of dissolved and vapor-phase gases to investigate methanogenic degradation of petroleum hydrocarbon contamination in the subsurface. Water Resources Research, 41(2), 1–15. https://doi.org/10.1029/2004WR003433

Anderson, M.P., Woessner, W.W., & Hunt, R.J. (2015). Applied Groundwater Modeling: Simulation of Flow and Advective Transport. 2nd Edition. ISBN 978-0-12-058103-0. https://doi.org/10.1016/C2009-0-21563-7

Appelo, C.A.J., & Postma, D. (2005). Geochemistry, groundwater and pollution, second edition. Geochemistry, Groundwater and Pollution, Second Edition, 1–649. https://doi.org/10.1201/9781439833544

Arabgol, R., Sartaj, M., & Asghari, K. (2016). Predicting Nitrate Concentration and Its Spatial Distribution in Groundwater Resources Using Support Vector Machines (SVMs) Model. Environmental Modeling and Assessment, 21(1), 71–82. https://doi.org/10.1007/S10666-015-9468-0

Arambarri, I., Lasa, M., Garcia, R., & Millán, E. (2004). Determination of fuel dialkyl ethers and BTEX in water using headspace solid-phase microextraction and gas chromatography–flame ionization detection. Journal of Chromatography A, 1033(2), 193–203. https://doi.org/10.1016/J.CHROMA.2004.01.046

Aronson, D., Citra, M., Shuler, K., Printup, H., Howard, P.H., & Weber, E.J. (1999). Aerobic biodegradation of organic chemicals in environmental media: A summary of field and laboratory studies. Final Report, SRC TR 99-002. Environmental Science Center, Syracuse Research Corporation

Arsenault, R., Martel, J.L., Brunet, F., Brissette, F., & Mai, J. (2023). Continuous streamflow prediction in ungauged basins: Long short-term memory neural networks clearly outperform traditional hydrological models. Hydrology and Earth System Sciences, 27(1), 139–157. https://doi.org/10.5194/hess-27-139-2023

Astrahan, P. (2018). Monocyclic aromatic hydrocarbons (phthalates and BTEX) and aliphatic components in the SE Mediterranean costal Sea-surface microlayer (SML): Origins and phase distribution analysis. Marine Chemistry, 205, 56–69. https://doi.org/10.1016/J.MARCHEM.2018.07.009

Atlas Scientific LLC. (2022). Lab grade pH probe: Gen 2 (Version 4.8). Retrieved from https://files.atlas-scientific.com/l-l-ph-probe.pdf on (15 September 2025)

Atlas Scientific LLC. (2024a). Lab grade gold ORP probe: Gen 3 (Version 1.0). Retrieved from https://files.atlas-scientific.com/ENV-40-ORP-G.pdf (15 September 2025)

Atlas Scientific LLC. (2024b). Mini Conductivity Probe K 1.0: Gen 3 (Version 2.5). Retrieved https://files.atlas-scientific.com/Mini_EC_K_1.0_probe.pdf (15 September 2025)

Atlas Scientific LLC. (2025). Mini lab grade dissolved oxygen probe: Gen 3 (Version 2.0). Retrieved from https://files.atlas-scientific.com/Mini_DO_probe.pdf (15 September 2025)

Ayscough, N.J., Young, W., Whitehouse, P., & Agency., G. Britain. E. (2002). Proposed environmental quality standards for ethylbenzene in water. Environment Agency

Baedecker, J.M., Cozzarelli, I.M., Eganhouse, R.P., & Siegel, D.I. (1993). Crude oil in a shallow sand and gravel aquifer-Ill. Biogeochemical reactions and mass balance modeling in anoxic groundwater. In Applied Geochemistry (Vol. 8)

Baedecker, M.J., Eganhouse, R.P., Bekins, B.A., & Delin, G.N. (2011). Loss of volatile hydrocarbons from an LNAPL oil source. Journal of Contaminant Hydrology, 126(3–4), 140–152. https://doi.org/10.1016/J.JCONHYD.2011.06.006

Bakker, M. (2019). Analytic Solutions for Tidal Propagation in Multilayer Coastal Aquifers. Water Resources Research, 55(4), 3452–3464. https://doi.org/10.1029/2019WR024757

Bakker, M., & Post, V. (2022). Analytical groundwater modeling: Theory and applications using python. Analytical Groundwater Modeling: Theory and Applications Using Python, 1–226. https://doi.org/10.1201/9781315206134

Bakker, M., Post, V., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., & Fienen, M. N. (2016). Scripting MODFLOW Model Development Using Python and FloPy. Groundwater, 54(5), 733–739. https://doi.org/10.1111/GWAT.12413

Banadkooki, F.B., Ehteram, M., Panahi, F., Sh. Sammen, S., Othman, F.B., & EL-Shafie, A. (2020). Estimation of total dissolved solids (TDS) using new hybrid machine learning models. Journal of Hydrology, 587, 124989. https://doi.org/10.1016/J.JHYDROL.2020.124989

Barbaro, J. R., Barker, J.F., Lemon, L.A., & Mayfield, C.I. (1992). Biotransformation of BTEX under anaerobic, denitrifying conditions: Field and laboratory observations. In Journal of Contaminant Hydrology (Vol. 11).

Baste, S., Klotz, D., Espinoza, E.A., Bardossy, A., & Loritz, R. (2025). Unveiling the Limits of Deep Learning Models in Hydrological Extrapolation Tasks. EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-425

Batlle-Aguilar, J., Brouyère, S., Dassargues, A., Morasch, B., Hunkeler, D., Höhener, P., Diels, L., Vanbroekhoven, K., Seuntjens, P., & Halen, H. (2009). Benzene dispersion and natural attenuation in an alluvial aquifer with strong interactions with surface water. Journal of Hydrology, 369(3–4), 305–317. https://doi.org/10.1016/j.jhydrol.2009.02.014

Batlle-Aguilar, J., Morasch, B., Hunkeler, D., & Brouyère, S. (2014). Benzene dynamics and biodegradation in alluvial aquifers affected by river fluctuations. Groundwater, 52(3), 388–398. https://doi.org/10.1111/gwat.12070

Beck, P. & Mann, B. (2010). A technical guide for demonstrating monitored natural attenuation of petroleum hydrocarbons in groundwater. CRC CARE Technical Report no. 15, CRC for Contamination Assessment and Remediation of the Environment, Adelaide, Australia. ISBN: 978-1-921431-25-8

Beckett, G. D., & Lundegard, P. (1997). Practically impractical - the limits of LNAPL recovery and relationship to risk. Ground Water Publishing Co., Westerville, OH (United States).

Bekins, B.A., F.D., Hostettler, W.N., Herkelrath, G.N., Delin, E., Warren, and Essaid, H.I. (2005). Progression of methanogenic degradation of crude oil in the subsurface. Environmental Geosciences 12: 139–152.

Bender, F., Mohler, R., Ricco, A.J., & Josse, F. (2012). Quantification of Benzene in Groundwater Using SH-Surface Acoustic Wave Sensors. https://doi.org/10.5162/IMCS2012/5.4.2

Bennett, P. C., Siegel, D. E., Baedecker, M. J., & Holt, M. F. (1993). Crude oil in a shallow sand and gravel aquifer-l. Hydrogeology and inorganic geochemistry. In Applied Geochemistry (Vol. 8)

Bergstra, J., Yamins, D. & Cox, D. (2013). Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. Proceedings of the 30th International Conference on Machine Learning, in Proceedings of Machine Learning Research 28(1):115-123. Retrieved from https://proceedings.mlr.press/v28/bergstra13.html (on 30 June 2025)

Besaw, L.E., Rizzo, D.M., Bierman, P.R., & Hackett, W.R. (2010). Advances in ungauged streamflow prediction using artificial neural networks. Journal of Hydrology, 386(1–4), 27–37. https://doi.org/10.1016/J.JHYDROL.2010.02.037

Bi, J., Li, Y., Chang, X., Yuan, H., Qiao, J. (2023). Hybrid Water Quality Prediction with Frequency Domain Conversion Enhancement and Seasonal Decomposition. Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, 5200–5205. https://doi.org/10.1109/SMC53992.2023.10394421

Bielefeldt, A.R., & Stensel, H.D. (1999). Modeling competitive inhibition effects during biodegradation of BTEX mixtures. Water Research, Vol 33, Issue 3. https://doi.org/10.1016/S0043-1354(98)00256-5

Bisong, E. (2019). Logistic Regression. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4470-8_20

Bregnard, T.P., Höhener, P., Häner, A., & Zeyer, J. (1996). Degradation of weathered diesel fuel by microorganisms from a contaminated aquifer in aerobic and anaerobic microcosms. Environmental Toxicology and Chemistry, 15(3), 299–307. https://doi.org/10.1002/ETC.5620150312

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (2017). Classification and regression trees. Classification and Regression Trees, 1–358. https://doi.org/10.1201/9781315139470

Briem, G.J., Benediktsson, J.A., & Sveinsson, J.R. (2002). Multiple classifiers applied to multisource remote sensing data. IEEE Transactions on Geoscience and Remote Sensing, 40(10), 2291–2299. https://doi.org/10.1109/TGRS.2002.802476

Brun, A., Engesgaard, P., Christensen, T.H., & Rosbjerg, D. (2002). Modelling of transport and biogeochemical processes in pollution plumes: Vejen landfill, Denmark.

Journal of Hydrology, 256(3–4), 228–247. https://doi.org/10.1016/S0022-1694(01)00549-2

Buchan, J. (2006). Our uncomfortable addiction. Financial Review. Retrieved from https://www.afr.com/markets/commodities/our-uncomfortable-addiction-20060721-jfb45 (on 16 October 2025)

Buerck, J., Roth, S., Kraemer, K., Scholz, M., & Klaas, N. (2001). Application of a fiber-optic NIR-EFA sensor system for in situ monitoring of aromatic hydrocarbons in contaminated groundwater. Journal of Hazardous Materials, 83(1–2), 11–28. https://doi.org/10.1016/S0304-3894(00)00335-6

Caldwell, M.E., & Suflita, J.M. (2000). Detection of Phenol and Benzoate as Intermediates of Anaerobic Benzene Biodegradation under Different Terminal Electron-Accepting Conditions. Environmental Science and Technology, 34(7), 1216–1220. https://doi.org/10.1021/ES990849J

Caswell, T.A., Droettboom, M., Lee, A., Hunter, J., Firing, E., Sales De Andrade, E., & Ivanov, P. (2020). matplotlib/matplotlib: Rel: v3. 3.1 [Software]. Zenodo. https://matplotlib.org/

Cavelan, A., Golfier, F., Colombano, S., Davarzani, H., Deparis, J., & Faure, P. (2022). A critical review of the influence of groundwater level fluctuations and temperature on LNAPL contaminations in the context of climate change. In Science of the Total Environment (Vol. 806). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2021.150412

Chang, G. (2014). Robust Kalman filtering based on Mahalanobis distance as outlier judging criterion. J Geod 88, 391–401. https://doi.org/10.1007/s00190-013-0690-8

Chen, K., Zhang, P., You, L., & Sun, J. (2024). Research on Kalman Filter Fusion Navigation Algorithm Assisted by CNN-LSTM Neural Network. Applied Sciences 2024, Vol. 14, Page 5493, 14(13), 5493. https://doi.org/10.3390/APP14135493

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Chiang, W.H., & Kinzelbach, W. (2003). Your First Groundwater Model with PMWIN. 3D-Groundwater Modeling with PMWIN, 7–53. https://doi.org/10.1007/978-3-662-05549-6_2

Chicco, D., Warrens, M.J., Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci, 7:e623. DOI: 10.7717/peerj-cs.623

Chiu, H.Y., Hong, A., Lin, S.L., Surampalli, R.Y., & Kao, C.M. (2013). Application of natural attenuation for the control of petroleum hydrocarbon plume: Mechanisms and effectiveness evaluation. Journal of Hydrology, 505, 126–137. https://doi.org/10.1016/J.JHYDROL.2013.09.027

Chollet, F., & others. (2015). Keras. Retrieved from https://keras.io (on 20 June 2025)

Colombani, N., Mastrocicco, M., Gargini, A., Davis, G.B., & Prommer, H. (2009). Modelling the fate of styrene in a mixed petroleum hydrocarbon plume. Journal of Contaminant Hydrology, 105(1–2), 38–55. https://doi.org/10.1016/j.jconhyd.2008.11.005

Cooper, J.S., Chow, E., Hubble, L.J., Chai, R., & Sosa-Pintos, A. (2020). Using Chemiresistor Sensor Arrays to Test Petrol Station Groundwater Samples for Hydrocarbon Pollutants. ECS Meeting Abstracts, MA2020-01(29), 2204. https://doi.org/10.1149/MA2020-01292204MTGABS

Copeland, A., Lytle, D.A. (2014). Measuring the Oxidation–reduction Potential of Important Oxidants in Drinking Water, 106. American Water Works Association, pp. E10–E20. https://doi.org/10.5942/jawwa.2014.106.0002

Cova, C.M., Rincon, E.M., Espinosa, E., Serrano, L., Zuliani, A., Zuliani, A. (2022). Paving the way for a green transition in the design of sensors and biosensors for the detection of volatile organic compounds (VOCs). Biosensors 12 (2), 51. https://doi.org/10.3390/BIOS12020051

Cozzarelli, I.M., Bekins, B.A., Baedecker, M.J., Aiken, G.R., Eganhouse, R.P., and Tuccillo, M.E. (2001). Progression of natural attenuation processes at a crude-oil spill site: I Geochemical evolution of the plume. Journal of Contaminant Hydrology 53: 369–385

Cozzarelli, I.M., Schreiber, M.E., Erickson, M.L., & Ziegler, B.A. (2016). Arsenic Cycling in Hydrocarbon Plumes: Secondary Effects of Natural Attenuation. https://doi.org/10.1111/gwat.12316

Curtis, G. P. (2003). Comparison of approaches for simulating reactive solute transport involving organic degradation reactions by multiple terminal electron acceptors. Computers and Geosciences, 29(3), 319–329. https://doi.org/10.1016/S0098-3004(03)00008-6

Cvetkovic, V., Dagan, G., & Cheng, H. (1998). Contaminant transport in aquifers with spatially variable hydraulic and sorption properties. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 454(1976), 2173–2207. https://doi.org/10.1098/RSPA.1998.0254

D'Affonseca, F.M., Prommer, H., Finkel, M., Blum, P., & Grathwohl, P. (2011). Modeling the long-term and transient evolution of biogeochemical and isotopic signatures in coal tar-contaminated aquifers. Water Resources Research, 47(5). https://doi.org/10.1029/2010WR009108

de Winter, C., Palleti, V.R., Worm, D., Kooij, R. (2019). Measuring imperfections of water quality sensors in water distribution networks. Meas. Sci. Technol. 30 (9), 095101. https://doi.org/10.1088/1361-6501/AB1EEB

Declercq, I., Cappuyns, V., & Duclos, Y. (2012). Monitored natural attenuation (MNA) of contaminated soils: State of the art in Europe—A critical evaluation. Science of The Total Environment, 426, 393–405. https://doi.org/10.1016/J.SCITOTENV.2012.03.040

Delle Site, A. (2001). Factors Affecting Sorption of Organic Compounds in Natural Sorbent/Water Systems and Sorption Coefficients for Selected Pollutants. A Review. Journal of Physical and Chemical Reference Data, 30(1), 187–439. https://doi.org/10.1063/1.1347984

Dey, P., Nag, K., Pal, T., Pal, N.R. (2018). Regularizing multilayer perceptron for robustness. IEEE Trans Syst Man Cybern Syst 48 (8), 1255–1266. https://doi.org/10.1109/TSMC.2017.2664143

Dillard, L. A., Essaid, H. I., & Herkelrath, W. N. (1997). Multiphase flow modeling of a crude-oil spill site with a bimodal permeability distribution. Water Resources Research, 33(7), 1617–1632. https://doi.org/10.1029/97WR00857

Domenico, P. A., & Schwartz, F. W. (1998). Physical and Chemical Hydrogeology Second Edition. John Wiley & Sons, Inc

Dou, J., Liu, X., & Hu, Z. (2008). Substrate interactions during anaerobic biodegradation of BTEX by the mixed cultures under nitrate reducing conditions. Journal of Hazardous Materials, 158(2–3), 264–272. https://doi.org/10.1016/j.jhazmat.2008.01.075

Duan, X., & Li, Y. (2017). Sources and Fates of BTEX in the General Environment and Its Distribution in Coastal Cities of China. Undefined, 01(02), 86–106. https://doi.org/10.26502/JESPH.9612009

Dupuit, E., Dandrieux, A., Kvapil, P., Ollivier, J., Dusserre, G., & Thomas, O. (2000). UV spectrophotometry for monitoring toxic gases. Analusis, 28(10), 966–972. https://doi.org/10.1051/ANALUSIS:2000163

Eberhardt, C., & Grathwohl, P. (2002). Time scales of organic contaminant dissolution from complex source zones: coal tar pools vs. blobs. Journal of

Contaminant Hydrology, 59(1–2), 45–66. https://doi.org/10.1016/S0169-7722(02)00075-X

Eckert, P., & Appelo, C.A.J. (2002). Hydrogeochemical modeling of enhanced benzene, toluene, ethylbenzene, xylene (BTEX) remediation with nitrate. Water Resources Research, 38(8), 5-1-5–11. https://doi.org/10.1029/2001wr000692

Eckert, P., Wisotzky, F., Obermann, P., Kracht, O., & Strauss, H. (2020). Natural bioattenuation and active remediation in a BTEX contaminated aquifer in Düsseldorf (Germany). Groundwater 2000, 389–390. https://doi.org/10.1201/9781003078593-192

El-Naas, M.H., Acio, J.A., & El Telib, A.E. (2014). Aerobic biodegradation of BTEX: Progresses and Prospects. Journal of Environmental Chemical Engineering, 2(2), 1104–1122. https://doi.org/10.1016/J.JECE.2014.04.009

Energy Institute (2025). Statistical Review of World Energy. 74th Ed., Energy Institute, London. ISBN 978 1 78725 474 9. Retrieved from https://www.energyinst.org/statistical-review/resources-and-data-downloads (on 10 October 2025)

Erickson, T.A., & Lear, K.L. (2014). An integrated optoelectronic chip for sensing aromatic hydrocarbon contaminants in groundwater. Sensors and Actuators B: Chemical, 204, 421–428. https://doi.org/10.1016/J.SNB.2014.06.020

Essaid, H.I., Bekins, B.A., Godsy, E.M., Warren, E., Baedecker, M.J., & Cozzarelli, I.M. (1995). Simulation of aerobic and anaerobic biodegradation processes at a crude oil spill site. Water Resources Research, Vol. 31, Issue 12. https://doi.org/10.1029/95WR02567

Essaid, H.I., Cozzarelli, I.M., Eganhouse, R.P., Herkelrath, W.N., Bekins, B.A., & Delin, G.N. (2003). Inverse modeling of BTEX dissolution and biodegradation at the Bemidji, MN crude-oil spill site. Journal of Contaminant Hydrology, 67(1–4), 269–299. https://doi.org/10.1016/S0169-7722(03)00034-2

European Union (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

Faouzi, N.E.E., & Klein, L.A. (2016). Data Fusion for ITS: Techniques and Research Needs. Transportation Research Procedia, 15, 495–512. https://doi.org/10.1016/J.TRPRO.2016.06.042

Fenner, K., Canonica, S., Wackett, L.P., & Elsner, M. (2013). Evaluating pesticide degradation in the environment: blind spots and emerging opportunities. Science (New York, N.Y.), 341(6147), 752–758. https://doi.org/10.1126/SCIENCE.1236281

Fetter, C.W., Charles W., Boving, T.B., & Kreamer, D.K. (2018). Contaminant hydrogeology. Pages 647, ISBN 13: 978-1-4786-3279-5

Fiore, S., & Zanetti, M.C. (2009). Sorption of Phenols: Influence of Groundwater pH and of Soil Organic Carbon Content. American Journal of Environmental Sciences, 5(4), 546–554

Fraters, B., Boumans, L.J.M., Van Elzakker, B.G., Gast, L.F.L., Griffioen, J., Klaver, G.T., Nelemans, J.A., Velthof, G.L., & Veld, H. (2008). A new compliance checking level for nitrate in groundwater? Feasibility study on monitoring the upper five metres of groundwater. www.rivm.nl

Freitas, J.G., Doulatyari, B., Molson, J.W., & Barker, J.F. (2011). Oxygenated gasoline release in the unsaturated zone, Part 2: Downgradient transport of ethanol and hydrocarbons. Journal of Contaminant Hydrology, 125(1–4), 70–85. https://doi.org/10.1016/j.jconhyd.2011.05.002

Friedt, J.M., Luzet, V., Soumann, V., Nief, N., Segues, B., Pucheu, G., Dehez, J.S., Ordonez-Varela, J.R., & Chérioux, F. (2025). Specific Detection of BTEX Contamination in Water Using a π-Hole-Catching Surface Acoustic Wave Sensor. ACS Omega, 10(11), 10903–10910. https://doi.org/10.1021/ACSOMEGA.4C08826

Gao, H., Tatomir, A.B., Karadimitriou, N.K., Steeb, H., & Sauter, M. (2022). Effect of Pore Space Stagnant Zones on Interphase Mass Transfer in Porous Media, for Two-Phase Flow Conditions. Transport in Porous Media, 146(3), 639–667. https://doi.org/10.1007/S11242-022-01879-0

Garg, S., Newell, C.J., Kulkarni, P.R., King, D.C., Adamson, D.T., Renno, M.I., & Sale, T. (2017). Overview of Natural Source Zone Depletion: Processes, Controlling Factors, and Composition Change. Groundwater Monitoring and Remediation, 37(3), 62–81. https://doi.org/10.1111/gwmr.12219

Geng, X., Boufadel, M.C., & Cui, F. (2017). Numerical modeling of subsurface release and fate of benzene and toluene in coastal aquifers subjected to tides. Journal of Hydrology, 551, 793–803. https://doi.org/10.1016/J.JHYDROL.2016.10.039

George, S., Dixit, A. (2021). A machine learning approach for prioritizing groundwater testing for per-and polyfluoroalkyl substances (PFAS). J. Environ. Manage. 295, 113359. ISSN 0301-4797. https://doi.org/10.1016/j.jenvman.2021.113359

Gharedaghloo, B., & Price, J.S. (2017). Fate and Transport of free-phase and dissolved-phase hydrocarbons in peat and peatlands: developing a conceptual model. Environmental Reviews. 26(1): 55-68. https://doi.org/10.1139/er-2017-0002

Gharedaghloo, B., & Price, J.S. (2021). Assessing benzene and toluene adsorption with peat depth: Implications on their fate and transport. Environmental Pollution, 274. https://doi.org/10.1016/J.ENVPOL.2021.116477

Gleeson, K., Husband, S., & Boxall, J. (2025). Multi-parameter multi-sensor data fusion for drinking water distribution system water quality management. AQUA - Water Infrastructure, Ecosystems and Society, 74(7), 451–465. https://doi.org/10.2166/AQUA.2025.002

Goblirsch, T., Mayer, T., Penzel, S., Rudolph, M., Borsdorf, H. (2023). In Situ Water Quality Monitoring Using an Optical Multiparameter Sensor Probe. Sensors, 23, 9545. https://doi.org/10.3390/s23239545

Godsy, E.M., Warren, E., Cozzarelli, I.M., Bekins, B.A., & Eganhouse, R.P. (1999). Determining BTEX Biodegradation Rates Using In Situ Microcosms at the Bemidji site, Minnesota: Trials and Tribulations

Gomez, D.E., & Alvarez, P.J.J. (2010). Comparing the effects of various fuel alcohols on the natural attenuation of Benzene Plumes using a general substrate interaction model. Journal of Contaminant Hydrology, 113(1–4), 66–76. https://doi.org/10.1016/j.jconhyd.2010.02.002

Goudar, C.T., & Strevett, K.A. (1998). Comparison of relative rates of BTEX biodegradation using respirometry. In Journal of Industrial Microbiology & Biotechnology (Vol. 21). http://www.stockton-press.co.uk/jim

Gualtieri, A.G., Chettri, S., Cromp, R., & Johnson, L.F. (1999). Support Vector Machine Classifiers as Applied to AVIRIS Data

Gupta, P.K., & Yadav, B.K. (2020). Three-Dimensional Laboratory Experiments on Fate and Transport of LNAPL under Varying Groundwater Flow Conditions. Journal of Environmental Engineering, 146(4), 4020010. https://doi.org/10.1061/(asce)ee.1943-7870.0001672

Hadley, P.W., Arulanantham, R., & Gandhi, D. (2015). California’s Low-Threat LUFT Site Closure Policy: Looking Forward. Remediation Journal, 25(2), 9–33. https://doi.org/10.1002/REM.21421

Haider, F.U., Ejaz, M., Cheema, S.A., Khan, M.I., Zhao, B., Liqun, C., Salim, M.A., Naveed, M., Khan, N., Núnez-Delgado, A., Mustafa, A. (2021). Phytotoxicity of petroleum hydrocarbons: sources, impacts and remediation strategies. Environ. Res. 197, 111031. https://doi.org/10.1016/J.ENVRES.2021.111031

Hakimi, O., Liu, H., Abudayyeh, O., Houshyar, A., Almatared, M., & Alhawiti, A. (2023). Data Fusion for Smart Civil Infrastructure Management: A Conceptual

Digital Twin Framework. Buildings 2023, Vol. 13, Page 2725, 13(11), 2725. https://doi.org/10.3390/BUILDINGS13112725

Hakkal, S., Lahcen, A.A., (2024). XGBoost to enhance learner performance prediction. Comput. Educ. 7, 100254. https://doi.org/10.1016/J.CAEAI.2024.100254

Hansen, S.K., & Kueper, B.H. (2014). A new model for coupled multicomponent NAPL dissolution and aqueous-phase transport, with application to creosote dissolution in discrete fractures. Water Resources Research, 50(1), 58–70. https://doi.org/10.1002/2013WR013773

Harbaugh, A.W. (2005). MODFLOW-2005, The U.S. Geological Survey Modular Ground-Water Model-the Ground-Water Flow Process. http://water.usgs.gov/software/ground_water.html/

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T.E. (2020). Array programming with NumPy. Nature, 585(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hasan, R., Chu, H., 2022. Noise in Datasets: What Are the Impacts on Classification Performance? 11th International Conference on Pattern Recognition Applications and Methods. https://doi.org/10.5220/0010782200003122

Healy, R.W., & Cook, P.G. (2002). Using groundwater levels to estimate recharge. Hydrogeology Journal, 10(1), 91–109. https://doi.org/10.1007/S10040-001-0178-0

Helbling, D.E. (2015). Bioremediation of pesticide-contaminated water resources: the challenge of low concentrations. Current Opinion in Biotechnology, 33, 142–148. https://doi.org/10.1016/J.COPBIO.2015.02.012

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/NECO.1997.9.8.1735

Hounslow, A. (1995). Water Quality Data: Analysis and Interpretation. 1st ed., CRC Press. https://doi.org/10.1201/9780203734117

Housego, R., Raubenheimer, B., Elgar, S., Cross, S., Legner, C., and Ryan, D. (2021). Coastal flooding generated by ocean wave- and surge-driven groundwater fluctuations on a sandy barrier island. Journal of Hydrology, 603, Part B. https://doi.org/10.1016/j.jhydrol.2021.126920

Höyng, D., Prommer, H., Blum, P., Grathwohl, P., & Mazo D'Affonseca, F. (2015). Evolution of carbon isotope signatures during reactive transport of hydrocarbons in heterogeneous aquifers. Journal of Contaminant Hydrology, 174, 10–27. https://doi.org/10.1016/J.JCONHYD.2014.12.005

Huang, F., Chuang, M., Wang, G.S., Yeh, H. (2015). Tide-induced groundwater level fluctuation in a U-shaped coastal aquifer. Journal of Hydrology, 530. https://doi.org/10.1016/j.jhydrol.2015.09.032

Huang, Y., Ding, L., Liu, W., Niu, H., Yang, M., Lyu, G., Lin, S., Hu, Q., Huang, Y., Ding, L., Liu, W., Niu, H., Yang, M., Lyu, G., Lin, S., & Hu, Q. (2023). Groundwater Contamination Site Identification Based on Machine Learning: A Case Study of Gas Stations in China. Water 2023, Vol. 15, 15(7). https://doi.org/10.3390/W15071326

Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, doi: 10.1109/MCSE.2007.55

Huntley, D., & Beckett, G.D. (2002). Persistence of LNAPL sources: relationship between risk reduction and LNAPL recovery. Journal of Contaminant Hydrology, 59(1–2), 3–26. https://doi.org/10.1016/S0169-7722(02)00073-6

Huntley, D., Hawk, R.N., & Corley, H.P. (1994). Nonaqueous Phase Hydrocarbon in a Fine-Grained Sandstone: 1. Comparison Between Measured and Predicted Saturations and Mobility. Groundwater, 32(4), 626–634. https://doi.org/10.1111/J.1745-6584.1994.TB00898.X

Hussain, M., O’Nils, M., Lundgren, J., & Mousavirad, S.J. (2024). A Comprehensive Review on Deep Learning-Based Data Fusion. IEEE Access, 12, 180093–180124. https://doi.org/10.1109/ACCESS.2024.3508271

Industrial Economics, Incorporated (2023). High Resolution Site Characterization at Petroleum Underground Storage Tank Release Sites - Applicability, Benefits, and Costs. Final Report. Retrieved from https://www.epa.gov/system/files/documents/2023-04/High%20Resolution%20Site%20Characterization%20Study%20Report.pdf (on 20 August 2025)

In-Situ. (2024). Aqua TROLL Multiparameter Sondes. https://in-situ.com/pub/media/support/documents/Aqua-TROLL-500-600-700-800_Spec-Sheet_ltr_en.pdf

International Labour Organization (2022). The future of work in the oil and gas industry: Opportunities and challenges for a just transition to a future of work that contributes to sustainable development (Geneva, 28 November–2 December 2022), International Labour Office, Sectoral Policies Department, Geneva. ISBN 978-92-2-037558-7

Interstate Technology & Regulatory Council [ITRC] (2009). Evaluating LNAPL Remedial Technologies for Achieving Project Goals. LNAPL-2. Washington, D.C.: Interstate Technology & Regulatory Council, LNAPLs Team. www.itrcweb.org

Irvine, T. (n.d.). Shock and Vibration Response Spectra Course Unit 2A. Sine Vibration Characteristics.

Jakobsen, R., & Postma, D. (1999). Redox zoning, rates of sulfate reduction and interactions with Fe-reduction and methanogenesis in a shallow sandy aquifer, Rømø, Denmark. Geochimica et Cosmochimica Acta, 63(1), 137–151. https://doi.org/10.1016/S0016-7037(98)00272-5

Ji, J., Deng, C., Shen, W., & Zhang, X. (2006). Field analysis of benzene, toluene, ethylbenzene and xylene in water by portable gas chromatography–microflame ionization detector combined with headspace solid-phase microextraction. Talanta, 69(4), 894–899. https://doi.org/10.1016/J.TALANTA.2005.11.032

Jiang, X.W., Sun, Z.C., Zhao, K.Y., Shi, F.S., Wan, L., Wang, X.S., & Shi, Z.M. (2017). A method for simultaneous estimation of groundwater evapotranspiration and inflow rates in the discharge area using seasonal water table fluctuations. Journal of Hydrology, 548, 498–507. https://doi.org/10.1016/J.JHYDROL.2017.03.026

Jiang, Y., Xi, B., Li, R., Li, M., Xu, Z., Yang, Y., & Gao, S. (2019). Advances in Fe(III) bioreduction and its application prospect for groundwater remediation: A review. Frontiers of Environmental Science and Engineering, 13(6), 1–11. https://doi.org/10.1007/s11783-019-1173-9

Jo, M.S., Rene, E.R., Kim, S.H., & Park, H.S. (2008). An analysis of synergistic and antagonistic behavior during BTEX removal in batch system using response surface methodology. Journal of Hazardous Materials, 152(3), 1276–1284. https://doi.org/10.1016/j.jhazmat.2007.08.002

Kabadayi, S., Pridgen, A., & Julien, C. (2006). Virtual Sensors: Abstracting Data from Physical Sensors

Kapuściński, R. (1985). Shah of Shahs (W. R. Brand & K. Mroczkowska-Brand, Trans.). Harcourt Brace Jovanovich. (Original work published 1982)

Karami, E., Bui, F.M., & Nguyen, H.H. (2012). Multisensor data fusion for water quality monitoring using wireless sensor networks. 2012 4th International Conference on Communications and Electronics, ICCE 2012, 80–85. https://doi.org/10.1109/CCE.2012.6315875

Karimi, H., Sahour, S., Khanbeyki, M., Gholami, V., Sahour, H., Shahabi-Ghahfarokhi, S., & Mohammadi, M. (2025). Enhancing groundwater quality prediction through ensemble machine learning techniques. Environmental Monitoring and Assessment, 197(1), 1–25. https://doi.org/10.1007/S10661-024-13506-0

Khan, S., Newport, D., & Le Calvé, S. (2020). Low-volume PEEK gas cell for BTEX detection using portable deep-UV absorption spectrophotometry. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 243, 118727. https://doi.org/10.1016/J.SAA.2020.118727

Khatri, N., Tyagi, S. (2015). Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. Front. Life Sci. 8 (1), 23–39. https://doi.org/10.1080/21553769.2014.933716

Koch, W. (2014). Tracking and Sensor Data Fusion: Methodological Framework and Selected Applications (1st ed.). Springer Berlin, Heidelberg

Kohfahl, C., Massmann, G., & Pekdeger, A. (2009). Sources of oxygen flux in groundwater during induced bank filtration at a site in Berlin, Germany. Hydrogeology Journal, 17(3), 571–578. https://doi.org/10.1007/s10040-008-0389-8

Kulkarni, P.R., Newell, C.J., King, D.C., Molofsky, L.J., & Garg, S. (2020). Application of Four Measurement Techniques to Understand Natural Source Zone Depletion Processes at an LNAPL Site. Groundwater Monitoring and Remediation, 40(3), 75–88. https://doi.org/10.1111/GWMR.12398

Lacey, Z.M. (2021). A Modeling Study of the Mobilization and Sequestration of Trace Metals in a Crude-Oil Contaminated Aquifer. Geosciences Student Honors Theses. 22. https://digitalcommons.trinity.edu/geo_honors/22

Larsson, H., & Dasgupta, P.K. (2003). Liquid core waveguide-based optical spectrometry for field estimation of dissolved BTEX compounds in groundwater: A feasibility study. Analytica Chimica Acta, 485(2), 155–167. https://doi.org/10.1016/S0003-2670(03)00423-9

Lee, J.Y., Cheon, J.Y., Lee, K.K., Lee, S.Y., & Lee, M.H. (2001). Factors affecting the distribution of hydrocarbon contaminants and hydrogeochemical parameters in a shallow sand aquifer. In Journal of Contaminant Hydrology (Vol. 50). www.elsevier.comrlocaterjconhyd

Lekmine, G., Bastow, T. P., Johnston, C.D., & Davis, G.B. (2014). Dissolution of multi-component LNAPL gasolines: The effects of weathering and composition. Journal of Contaminant Hydrology, 160, 1–11. https://doi.org/10.1016/j.jconhyd.2014.02.003

Lekmine, G., Sookhak Lari, K., Johnston, C.D., Bastow, T.P., Rayner, J.L., & Davis, G.B. (2017). Evaluating the reliability of equilibrium dissolution assumption from residual gasoline in contact with water saturated sands. Journal of Contaminant Hydrology, 196, 30–42. https://doi.org/10.1016/j.jconhyd.2016.12.003

Lenhard, R.J., Oostrom, M., & Dane, J.H. (2004). A constitutive model for air–NAPL–water flow in the vadose zone accounting for immobile, non-occluded (residual) NAPL in strongly water-wet porous media. Journal of Contaminant Hydrology, 73(1–4), 283–304. https://doi.org/10.1016/J.JCONHYD.2004.07.005

Lenhard, R.J., Sookhak Lari, K., Rayner, J.L., & Davis, G.B. (2018). Evaluating an Analytical Model to Predict Subsurface LNAPL Distributions and Transmissivity from Current and Historic Fluid Levels in Groundwater Wells: Comparing Results to Numerical Simulations. Groundwater Monitoring & Remediation, 38(1), 75–84. https://doi.org/10.1111/GWMR.12254

Li, H., Gu, X., Song, J., Hui, K., Chen, G., Tan, W., Wang, H., Jiang, Y., & Yuan, Y. (2024). Effects of soil-groundwater environmental factors on BTEX transport and transformation: A review. Journal of Environmental Chemical Engineering, 12(5), 113697. https://doi.org/10.1016/J.JECE.2024.113697

Li, H., Son, J.H., Hanif, A., Gu, J., Dhanasekar, A., Carlson, K. (2017). Colorado water watch: real-time groundwater monitoring for possible contamination from oil and gas activities. J. Water Resour. Protect. 9. https://doi.org/10.4236/jwarp.2017.913104

Li, J.J., Tong, X. (2020). Statistical hypothesis testing versus machine learning binary classification: distinctions and guidelines. Patterns 1 (7), 100115. https://doi.org/10.1016/J.PATTER.2020.100115

Li, L., Qiao, J., Yu, G., Wang, L., Li, H.Y., Liao, C., Zhu, Z. (2022). Interpretable tree-based ensemble model for predicting beach water quality. Water Res. 211, 118078. https://doi.org/10.1016/J.WATRES.2022.118078

Li, P., Karunanidhi, D., Subramani, T., & Srinivasamoorthy, K. (2021). Sources and Consequences of Groundwater Contamination. Archives of Environmental Contamination and Toxicology, 80(1), 1–10. https://doi.org/10.1007/S00244-020-00805-Z

Li, Y., Du, Y., Deng, Y., Fan, R., Tao, Y., Ma, T., & Wang, Y. (2022). Predicting the spatial distribution of phosphorus concentration in Quaternary sedimentary aquifers using simple field parameters. Applied Geochemistry, 142, 105349. https://doi.org/10.1016/J.APGEOCHEM.2022.105349

Libera, A., de Barros, F.P.J., Faybishenko, B., Eddy-Dilek, C., Denham, M., Lipnikov, K., Moulton, D., Maco, B., & Wainwright, H. (2019). Climate change impact on residual contaminants under sustainable remediation. Journal of Contaminant Hydrology, 226, 103518. https://doi.org/10.1016/J.JCONHYD.2019.103518

Liggins II, M., Hall, D.L., & Llinas, J. (2017). Handbook of multisensor data fusion : theory and practice (2nd Edition). CRC Press. https://www.oreilly.com/library/view/handbook-of-multisensor/9781351835374/

Lin, S., Zheng, H., Han, B., Li, Y., Han, C., Li, W. (2022). Comparative performance of eight ensemble learning approaches for the development of models of slope stability prediction. Acta Geotech. 17 (4), 1477–1502. https://doi.org/10.1007/s11440-021- 01440-1

Littlejohns, J.V, & Daugulis, A.J. (2008). Kinetics and interactions of BTEX compounds during degradation by a bacterial consortium. Process Biochemistry, 43(10), 1068–1076. https://doi.org/10.1016/j.procbio.2008.05.010

Liu, X.L., Zhao, L., Liu, S.J., & Cui, C.Y. (2010). Remediation of Groundwater Contaminated by BTEX Using a Biological Barrier. Advanced Materials Research, 178, 254–259. https://doi.org/10.4028/www.scientific.net/amr.178.254

Lo Savio, F., Grasso, G.M., Maccarrone, G. (2024). Low-cost UV-induced fluorescence sensor for monitoring aromatic hydrocarbons in coastal marine water, Acta IMEKO, vol. 13, no. 4, pp. 1-9. DOI: 10.21014/actaimeko.v13i4.1749

Lohman, S.W. (1972). Groundwater Hydraulics (p. 708). U.S Geological Survey Professional Paper. https://doi.org/10.3133/pp708

Lu, G., Clement, T.P., Zheng, C., & Wiedemeier, T.H. (1999). Natural attenuation of BTEX compounds: model development and field-scale application. Ground Water, 37(5), 707–717. https://doi.org/10.1111/J.1745-6584.1999.TB01163.X

Lueders, T. (2017). The ecology of anaerobic degraders of BTEX hydrocarbons in aquifers. FEMS Microbiology Ecology, 93(1), 220. https://doi.org/10.1093/FEMSEC/FIW220

Lyons, K.J., Ikonen, J., Hokajärvi, A.M., Räsänen, T., Pitkänen, T., Kauppinen, A., Kujala, K., Rossi, P.M., & Miettinen, I.T. (2023). Monitoring groundwater quality with real-time data, stable water isotopes, and microbial community analysis: A comparison with conventional methods. Science of The Total Environment, 864, 161199. https://doi.org/10.1016/J.SCITOTENV.2022.161199

Mao, B., Liu, Z., Liu, S., Zhang, M., & Lu, T. (2020). Investigation of relative permeability, saturation and capillary pressure relations of NAPL-contaminated sands. Journal of Soils and Sediments, 20(3), 1609–1620. https://doi.org/10.1007/S11368-019-02506-0

Martin, D., Kühl, N., & Satzger, G. (2021). Virtual Sensors. Business & Information Systems Engineering, 63. https://doi.org/10.1007/s12599-021-00689-w

Mastrocicco, M., Colombani, N., & Gargini, A. (2012). Characterization and modeling of a BTEX plume originated by a sulphur rich NAPL source. Characterization and modeling of a BTEX plume originated by a sulphur rich NAPL source. https://www.researchgate.net/publication/280647793

Maurer, M., & Rittmann, B.E. (2004). Modeling intrinsic bioremediation for interpret observable biogeochemical footprints of BTEX biodegradation: the need for fermentation and abiotic chemical processes. In Biodegradation (Vol. 15)

McGuire, J.T., Cozzarelli, I.M., Bekins, B.A., Link, H., & Martinović-Weigelt, D. (2018). Toxicity Assessment of Groundwater Contaminated by Petroleum Hydrocarbons at a Well-Characterized, Aged, Crude Oil Release Site. Environmental Science and Technology, 52(21), 12172–12178. https://doi.org/10.1021/ACS.EST.8B03657

McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 56–61. https://doi.org/10.25080/MAJORA-92BF1922-00A

McKnight, U.S., Funder, S.G., Rasmussen, J.J., Finkel, M., Binning, P.J., Bjerg, P.L. (2010). An integrated model for assessing the risk of TCE groundwater contamination to human receptors and surface water ecosystems. Ecol. Eng. 36 (9), 1126–1137. https://doi.org/10.1016/J.ECOLENG.2010.01.004

McMahon, P.B., Chapelle, F.H., & Bradley, P.M. (2011). Evolution of redox processes in groundwater. ACS Symposium Series, 1071, 581–597. https://doi.org/10.1021/bk-2011-1071.ch026

Menz, C. (2016). Oxygen delivering processes in groundwater and their relevance for iron-related well clogging processes-a case study on the quaternary aquifers of Berlin. http://dx.doi.org/10.17169/refubium-13888

Miles, B., Peter, A., & Teutsch, G. (2008). Multicomponent simulations of contrasting redox environments at an LNAPL site. Ground Water, 46(5), 727–742. https://doi.org/10.1111/j.1745-6584.2008.00457.x

Mitchell, H.B. (2007). Multi-sensor data fusion: An introduction. In Multi-Sensor Data Fusion: An Introduction. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71559-7

Mitra, S., & Roy, P. (2011). BTEX : A Serious Ground-water Contaminant. Research Journal of Environmental Sciences 5(5):394-398. DOI:10.3923/rjes.2011.394.398

Molins, S., Mayer, K. U., Amos, R. T., & Bekins, B. A. (2010). Vadose zone attenuation of organic compounds at a crude oil spill site - Interactions between biogeochemical reactions and multicomponent gas transport. Journal of Contaminant Hydrology, 112(1–4), 15–29. https://doi.org/10.1016/j.jconhyd.2009.09.002

Molson, J.W., Barker, J.F., Frind, E.O., & Schirmer, M. (2002). Modeling the impact of ethanol on the persistence of benzene in gasoline-contaminated groundwater. Water Resources Research, 38(1), 4-1-4–12. https://doi.org/10.1029/2001wr000589

Molson, J.W.P., & Eng, I. (2011). BIONAPL/3D: A 3D Model for Groundwater Flow, and Multi-Component NAPL Dissolution with Dissolved-Phase Advective-Dispersive Transport and Biodegradation in Porous and Fractured Porous Media User Guide

Morris, B.E.L., Henneberger, R., Huber, H., & Moissl-Eichinger, C. (2013). Microbial syntrophy: interaction for the common good. FEMS Microbiology Reviews, 37(3), 384–406. https://doi.org/10.1111/1574-6976.12019

Muhammad, G., Alshehri, F., Karray, F., Saddik, A. El, Alsulaiman, M., & Falk, T.H. (2021). A comprehensive survey on multimodal medical signals fusion for smart healthcare systems. Information Fusion, 76, 355–375. https://doi.org/10.1016/J.INFFUS.2021.06.007

Muneron Mello, J.M., Brandão, H.L., Valério, A., de Souza, A.A.U., de Oliveira, D., da Silva, A., & de Souza, S.M. A.G.U. (2019). Biodegradation of BTEX compounds from petrochemical wastewater: Kinetic and toxicity. Journal of Water Process Engineering, 32. https://doi.org/10.1016/j.jwpe.2019.100914

Narkhede, M.V., Bartakke, P.P., & Sutaone, M.S. (2022). A review on weight initialization strategies for neural networks. Artificial Intelligence Review, 55(1), 291–322. https://doi.org/10.1007/S10462-021-10033-Z

Naval Facilities Engineering Command (2010). LNAPL Site Management Handbook. Retrieved from https://exwc.navfac.navy.mil/Portals/88/Documents/EXWC/Restoration/er_pdfs/l/navfac-ev-hdbk-lnapl-mgmt-20101130.pdf?ver=NzA6R9OJkpLkkVl4_04jDA%3D%3D (on 10 October 2025)

Nayak, S., Das, M., Pradhan, B., & Kundu, S.N. (2019). Environmental Impacts of Fossil Fuels. Sustainable Energy and Environment, 245–270. https://doi.org/10.1201/9780429430107-13

Neto, D.C., Chang, H.K., & van Genuchten, M.T. (2016). A Mathematical View of Water Table Fluctuations in a Shallow Aquifer in Brazil. Groundwater, 54(1), 82–91. https://doi.org/10.1111/GWAT.12329

Neumann, A., Sander, M., & Hofstetter, T.B. (2011). Redox properties of structural Fe in smectite clay minerals. ACS Symposium Series, 1071, 361–379. https://doi.org/10.1021/bk-2011-1071.ch017

Ng, G.H.C., Bekins, B.A., Cozzarelli, I.M., Baedecker, M.J., Bennett, P.C., & Amos, R.T. (2014). A mass balance approach to investigating geochemical controls on secondary water quality impacts at a crude oil spill site near Bemidji, MN. Journal of Contaminant Hydrology, 164, 1–15. https://doi.org/10.1016/J.JCONHYD.2014.04.006

Ng, G.H.C., Bekins, B.A., Cozzarelli, I.M., Baedecker, M.J., Bennett, P.C., Amos, R.T., & Herkelrath, W.N. (2015). Reactive transport modeling of geochemical controls on secondary water quality impacts at a crude oil spill site near Bemidji, MN. Water Resources Research, 51(6), 4156–4183. https://doi.org/10.1002/2015WR016964

Njobuenwu, D.O., Amadi, S.A., & Ukpaka, P.C. (2005). Dissolution rate of BTEX contaminants in water. Canadian Journal of Chemical Engineering, 83(6), 985–989. https://doi.org/10.1002/CJCE.5450830608

Noble, P., & Morgan, P. (2002). The Effects of Contaminant Concentration on the Potential for Natural Attenuation. R&D Technical Report P2-228/TR. Environment Agency, ISBN: 1 85705 599 3

Nordstrom, D., Plummer, L., Wigley, T., Wolery, T., Ball, J., Jenne, E., Bassett, R., Crerar, D., Florence, T., Fritz, B., Hoffman, M., Holdren, G., Lafon, G., Mattigod, S., Mcduff, R., Morel, F., Reddy, M., Sposito, G., & Thrailkill, J. (1979). A Comparison of Computerized Chemical Models for Equilibrium Calculations in Aqueous Systems 1. https://doi.org/10.1021/bk-1979-0093.ch038

Ololade, I.A., Arogunrerin, I.A., Oladoja, N.A., Ololade, O.O., & Alabi, A.B. (2021). Concentrations and Toxic Equivalency of Polycyclic Aromatic Hydrocarbons (PAHs) and Polychlorinated Biphenyl (PCB) Congeners in Groundwater Around Waste Dumpsites in South-West Nigeria. Archives of Environmental Contamination and Toxicology 2021 80:1, 80(1), 134–143. https://doi.org/10.1007/S00244-020-00790-3

Österreicher-Cunha, P., Vargas, E.A., Guimarães, J.R.D., Lago, G.P., Antunes, F.S., & da Silva, M.I.P. (2009). Effect of ethanol on the biodegradation of gasoline in an unsaturated tropical soil. International Biodeterioration and Biodegradation, 63(2), 208–216. https://doi.org/10.1016/j.ibiod.2008.09.004

Pandolfo, E., Caracciolo, A.B., & Rolando, L. (2023). Recent advances in bacterial degradation of hydrocarbons. Water, 15(2), 375. https://doi.org/10.3390/w15020375

Pant, D., Keesari, T., Rishi, M., Jaryal, A., Sharma, D.A., Thakur, N., Singh, G., Kamble, S.N., Sangwan, P., Sinha, U.K., & Tripathi, R.M. (2019). Quality and Quantity of Groundwater in Highly Exploited Aquifers of Northwest India. Journal of Hazardous, Toxic, and Radioactive Waste, 24(2), 05019009. https://doi.org/10.1061/(ASCE)HZ.2153-5515.0000483

Park, E., & Parker, J.C. (2008). A simple model for water table fluctuations in response to precipitation. Journal of Hydrology, 356(3–4), 344–349. https://doi.org/10.1016/J.JHYDROL.2008.04.022

Parkhurst, D.L., & Appelo, C.A.J. (2000). User's Guide to PHREEQC (Version 2)--A Computer Program for Speciation, Batch-Reaction, One-Dimensional Transport, and Inverse Geochemical Calculations. https://wwwbrr.cr.usgs.gov/projects/GWC_coupled/phreeqc/html/final.html

Parmar, A., Katariya, R., & Patel, V. (2019). A Review on Random Forest: An Ensemble Classifier. 10.1007/978-3-030-03146-6_86

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830

Peter, O. (2010). Biological Remediation of Hydrocarbon and Heavy Metals Contaminated Soil. www.intechopen.com

Picone, S. (2012.). Transport and biodegradation of volatile organic compounds: influence on vapor intrusion into buildings. https://doi.org/10.18174/212664

Pishgar, R., Hettiaratchi, J.P., & Chu, A. (2022). Natural Source Zone Depletion (NSZD) Quantification Techniques: Innovations and Future Directions. Sustainability 2022, Vol. 14, Page 7027, 14(12), 7027. https://doi.org/10.3390/SU14127027

Pollock, D.W. (2017). MODPATH v7.2.01: A particle-tracking model for MODFLOW: U.S. Geological Survey Software Release, http://dx.doi.org/10.5066/F70P0X5X

Popescu, M.C., Balas, V.E., Perescu-Popescu, L., & Mastorakis, N. (2009). Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems. https://doi.org/10.5555/1639537.1639542

Postma, D., & Jakobsen, R. (1996). Redox zonation: Equilibrium constraints on the Fe(III)/SO4-reduction interface. Geochimica et Cosmochimica Acta, 60(17), 3169–3175. https://doi.org/10.1016/0016-7037(96)00156-1

Powers, S.E., Hunt, C.S., Heermann, S.E., Corseuil, H.X., Rice, D., & Alvarez, P.J.J. (2001). The Transport and Fate of Ethanol and BTEX in Groundwater Contaminated by Gasohol, 31(1), 79–123. https://doi.org/10.1080/20016491089181

Prinčič, A., Mahne, I., Megušar, F., Paul, E.A., & Tiedje, J.M. (1998). Effects of pH and oxygen and ammonium concentrations on the community structure of nitrifying bacteria from wastewater. Applied and Environmental Microbiology, 64(10), 3584–3590. DOI: 10.1128/aem.64.10.3584-3590.1998

Prommer, H., & Post, V. (2010). A Reactive Multicomponent Transport Model for Saturated Porous Media. http://www.pht3d.org

Prommer, H., Anneser, B., Rolle, M., Einsiedl, F., & Griebler, C. (2009). Biogeochemical and isotopic gradients in a BTEX/PAH contaminant plume: Model-based interpretation of a high-resolution field data set. Environmental Science and Technology, 43(21), 8206–8212. https://doi.org/10.1021/es901142a

Prommer, H., Barry, D.A., & Davis, G.B. (2002). Modelling of physical and reactive processes during biodegradation of a hydrocarbon plume under transient groundwater flow conditions. www.elsevier.com/locate/jconhyd

Prommer, H., Sun, J., & Kocar, B.D. (2019). Using Reactive Transport Models to Quantify and Predict Groundwater Quality. Elements, 15(2), 87–92. https://doi.org/10.2138/GSELEMENTS.15.2.87

Qiao, F., Wang, J., Song, J., Chen, Z., Kwaw, A.K., Zhao, Y., & Zheng, S. (2025). The spatiotemporal evolution of dissolved-phase NAPL plumes revealed by the integrated groundwater quality and machine learning models. Water Research, 280. https://doi.org/10.1016/j.watres.2025.123535

Rad, M., Abtahi, A., Berndtsson, R., McKnight, U. S., & Aminifar, A. (2024). Interpretable machine learning for predicting the fate and transport of pentachlorophenol in groundwater. Environmental Pollution, 345, 123449. https://doi.org/10.1016/J.ENVPOL.2024.123449

Rama, F., Ramos, D. T., Müller, J. B., Corseuil, H. X., & Miotliński, K. (2019). Flow field dynamics and high ethanol content in gasohol blends enhance BTEX migration and biodegradation in groundwater. Journal of Contaminant Hydrology, 222, 17–30. https://doi.org/10.1016/j.jconhyd.2019.01.003

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. Encyclopedia of Database Systems, 532–538. https://doi.org/10.1007/978-0-387-39940-9_565

Renu, V., & Kumar, G.S. (2016). Numerical Modeling on Benzene Dissolution into Groundwater and Transport of Dissolved Benzene in a Saturated Fracture-Matrix

System. Environmental Processes, 3(4), 781–802. https://doi.org/10.1007/s40710-016-0166-y

Richard-Cerda, J.C., Bockstiegel, M., Muñoz-Vega, E., Knöller, K., Schüth, C., & Schulz, S. (2024). High-Resolution Monitoring and Redox-Potential-Based Solute Transport Modeling to Partition Denitrification Pathways at an Agricultural Site. ACS ES&T Water, 4(11), 4917–4931. https://doi.org/10.1021/ACSESTWATER.4C00540

Richter, K., Schicklberger, M., & Gescher, J. (2011). Dissimilatory reduction of extracellular electron acceptors in anaerobic respiration. In Applied and Environmental Microbiology (Vol. 78, Issue 4, pp. 913–921). https://doi.org/10.1128/AEM.06803-11

Rixey, W.G., and Joshi, S. (2000). Dissolution of MTBE from a residually trapped gasoline source. A summary of research results from API's Soil & Groundwater Technical Task Force. American Petroleum Institute. https://www.api.org/~/media/files/ehs/clean_water/bulletins/13_bull.pdf

Robinson, C., Brovelli, A., Barry, D.A., & Li, L. (2009). Tidal influence on BTEX biodegradation in sandy coastal aquifers. Advances in Water Resources, 32(1), 16–28. https://doi.org/10.1016/J.ADVWATRES.2008.09.008

Ross, J. (2013). Groundwater Monitoring Program Newcastle Gas Storage Facility Project i Groundwater Monitoring Program Newcastle Gas Storage Facility Document Revision History F Nicola Fry for HWC and NOW 6 monthly AGL-HWC-NOW meeting

Rotzoll, K. & El-Kadi, A.I. (2008). Estimating hydraulic properties of coastal aquifers using wave setup. Journal of Hydrology, 353 (1–2), 201-213. https://doi.org/10.1016/j.jhydrol.2008.02.005

Rupok, Q.S.S., Suman, K.H., Sakib, M.N., & Agarwal, J. (2020). Sensor Data Fusion for Monitoring Water Quality Toward Sustainable Freshwater Fisheries. ETCCE 2020 - International Conference on Emerging Technology in Computing, Communication and Electronics. https://doi.org/10.1109/ETCCE51779.2020.9350876

Rusydi, A.F., Onodera, S.I., Saito, M., Ioka, S., Maria, R., Ridwansyah, I., & Delinom, R.M. (2021). Vulnerability of groundwater to iron and manganese contamination in the coastal alluvial plain of a developing Indonesian city. SN Applied Sciences, 3(4), 1–12. https://doi.org/10.1007/S42452-021-04385-Y

Rynkiewicz, J. (2019). On overfitting of multilayer perceptrons for classification. In J. Rynkiewicz (ed.), Computational Intelligence and Machine Learning. http://www.i6doc.com/en/

Salanitro, J.P. (1993). The Role of Bioattenuation in the Management of Aromatic Hydrocarbon Plumes in Aquifers. Groundwater Monitoring & Remediation, 13(4), 150–161. https://doi.org/10.1111/J.1745-6592.1993.TB00459.X

Sale, T., Gallo, S., Askarani, K.K., Irianni-Renno, M., Lyverse, M., Hopkins, H., Blotevogel, J., & Burge, S. (2021). Real-time soil and groundwater monitoring via spatial and temporal resolution of biogeochemical potentials. Journal of Hazardous Materials, 408, 124403. https://doi.org/10.1016/J.JHAZMAT.2020.124403

Salowsky, H., Schäfer, W., Schneider, A.L., Müller, A., Dreher, C., & Tiehm, A. (2021). Beneficial effects of dynamic groundwater flow and redox conditions on Natural Attenuation of mono-, poly-, and NSO-heterocyclic hydrocarbons. Journal of Contaminant Hydrology, 243. https://doi.org/10.1016/j.jconhyd.2021.103883

Sani Gaya, M., Isah Abba, S., Abdu, A.M., Tukur, A.I., Saleh, M.A., Esmaili, P., Wahab, N.A., & Gaya, M.S. (2020). Estimation of water quality index using artificial intelligence approaches and multi-linear regression. IAES International Journal of Artificial Intelligence (IJ-AI, 9(1), 126–134. https://doi.org/10.11591/ijai.v9.i1.pp126-134

Schmidt, F., Wainwright, H.M., Faybishenko, B., Denham, M., & Eddy-Dilek, C. (2018). In Situ Monitoring of Groundwater Contamination Using the Kalman Filter. Environmental Science & Technology, 52(13), 7418–7425. https://doi.org/10.1021/ACS.EST.8B00017

Schmidt, L., Santurkar, S., Tsipras, D., Talwar, K., Madry, A. (2018). Adversarially robust generalization requires more data. Adv. Neural Inf. Proces. Syst. 31. https://doi.org/ 10.48550/arXiv.1804.11285

Schooltink, W.T. (2020). Testing the Sensitivity of Machine Learning Classifiers to Attribute Noise in Training Data. http://deeplearning.net/datasets/

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., & Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecological Modelling, 406, 109–120. https://doi.org/10.1016/J.ECOLMODEL.2019.06.002

Schreiber, M.E., & Bahr, J.M. (2002). Nitrate-enhanced bioremediation of BTEX-contaminated groundwater: parameter estimation from natural-gradient tracer experiments. In Journal of Contaminant Hydrology (Vol. 55). www.elsevier.com/locate/jconhyd

Schreiber, M.E., Carey, G.R., Feinstein, D.T., & Bahr, J.M. (2004). Mechanisms of electron acceptor utilization: Implications for simulating anaerobic biodegradation. Journal of Contaminant Hydrology, 73(1–4), 99–127. https://doi.org/10.1016/j.jconhyd.2004.01.004

Seagren, E.A., & Becker, J.G. (2002). Review of Natural Attenuation of BTEX and MTBE in Groundwater. Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management, 6(3), 156–172. https://doi.org/10.1061/(ASCE)1090-025X(2002)6:3(156)

Sen, P.C., Hajra, M., Ghosh, M. (2020). Supervised classification algorithms in machine learning: a survey and review. Adv. Intell. Syst. Comput. 937, 99–111. https://doi. org/10.1007/978-981-13-7403-6_11

Shafer, G. (1990). Perspectives on the theory and practice of belief functions. International Journal of Approximate Reasoning, 4(5–6), 323–362. https://doi.org/10.1016/0888-613X(90)90012-Q

Sheskin, D.J. (2020). Handbook of Parametric and Nonparametric Statistical Procedures, Fifth Edition. Handbook of Parametric and Nonparametric Statistical Procedures. https://doi.org/10.1201/9780429186196

Shores, A., Laituri, M., & Butters, G. (2017). Produced Water Surface Spills and the Risk for BTEX and Naphthalene Groundwater Contamination. Water, Air, and Soil Pollution, 228(11). https://doi.org/10.1007/s11270-017-3618-8

Sihota, N.J., & Mayer, K.U. (2012). Characterizing Vadose Zone Hydrocarbon Biodegradation Using Carbon Dioxide Effluxes, Isotopes, and Reactive Transport Modeling. Vadose Zone Journal, 11(4), vzj2011.0204. https://doi.org/10.2136/vzj2011.0204

Simon, D. (2006). Optimal state estimation: Kalman, H∞, and nonlinear approaches. In D. Simon (Ed.), Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches. Wiley Blackwell. https://doi.org/10.1002/0470045345

Singh, S.K., Shirzadi, A., & Pham, B.T. (2021). Application of Artificial Intelligence in Predicting Groundwater Contaminants. Water Pollution and Management Practices, 71–105. https://doi.org/10.1007/978-981-15-8358-2_4

Sookhak Lari, K., Davis, G.B., & Johnston, C.D. (2016). Incorporating hysteresis in a multi-phase multi-component NAPL modelling framework; a multi-component LNAPL gasoline example. Advances in Water Resources, 96, 190–201. https://doi.org/10.1016/j.advwatres.2016.07.012

Sookhak Lari, K., Davis, G.B., & Rayner, J.L. (2025). A roadmap to understanding key knowledge gaps in natural source zone depletion. Nature Water 2025 3:5, 3(5), 537–549. https://doi.org/10.1038/s44221-025-00436-5

Sorenson, R.P. (2020). Pre-Drake Oil and Gas Wells. Petroleum History Institute, Oil Industry History, vol 21. https://archives.datapages.com/data/phi/v21-2020/sorenson.htm

Spence, M.J., Bottrell, S.H., Thornton, S.F., Richnow, H.H., & Spence, K.H. (2005). Hydrochemical and isotopic effects associated with petroleum fuel biodegradation pathways in a chalk aquifer. Journal of Contaminant Hydrology, 79(1–2), 67–88. https://doi.org/10.1016/j.jconhyd.2005.06.003

Starbuck, C. (2023). Logistic Regression. The Fundamentals of People Analytics, 223–238. https://doi.org/10.1007/978-3-031-28674-2_12

Steefel, C.I., Appelo, C.A.J., Arora, B., Jacques, D., Kalbacher, T., Kolditz, O., Lagneau, V., Lichtner, P.C., Mayer, K.U., Meeussen, J.C.L., Molins, S., Moulton, D., Shao, H., Šimůnek, J., Spycher, N., Yabusaki, S.B., & Yeh, G.T. (2015). Reactive transport codes for subsurface environmental simulation. Computational Geosciences, 19(3), 445–478. https://doi.org/10.1007/S10596-014-9443-X

Stellman, C.M., Ewing, K.J., Bucholtz, F., & Aggarwal, I.D. (1998). Determination of BTEX contaminants in water via a long-pathlength fiber-optic Raman 'dip-stick.' Sensors and Actuators B: Chemical, 53(3), 173–178. https://doi.org/10.1016/S0925-4005(98)00337-2

Suarez, M.P., & Rifai, H.S. (2004). Modeling Natural Attenuation of Total BTEX and Benzene Plumes with Different Kinetics. Groundwater Monitoring & Remediation, 24(3), 53–68. https://doi.org/10.1111/J.1745-6592.2004.TB01292.X

Sulaymon, A.H., & Gzar, H.A. (2011). Experimental investigation and numerical modeling of light nonaqueous phase liquid dissolution and transport in a saturated zone of the soil. Journal of Hazardous Materials, 186(2–3), 1601–1614. https://doi.org/10.1016/j.jhazmat.2010.12.035

Suthersan, S., Koons, B., & Schnobrich, M. (2015). Advances in Remediation Solutions Contemporary Management of Sites with Petroleum LNAPL Presence. In NGWA.org Groundwater Monitoring & Remediation (Vol. 35, Issue 1).

Szomolányi, O., & Clement, A. (2023). Use of random forest for assessing the effect of water quality parameters on the biological status of surface waters. GEM - International Journal on Geomathematics, 14(1), 1–29. https://doi.org/10.1007/S13137-023-00229-6/TABLES/3

Teramoto, E.H., & Chang, H.K. (2019). Geochemical conceptual model of BTEX biodegradation in an iron-rich aquifer. Applied Geochemistry, 100, 293–304. https://doi.org/10.1016/j.apgeochem.2018.11.019

Teramoto, E.H., & Chang, H.K. (2020). A screening model to predict entrapped LNAPL depletion. Water (Switzerland), 12(2). https://doi.org/10.3390/w12020334

Teramoto, E.H., Vogt, C., Martins Baessa, M.P., Polese, L., Soriano, A. U., Chang, H.K., & Richnow, H.H. (2020). Dynamics of hydrocarbon mineralization characterized by isotopic analysis at a jet-fuel-contaminated site in subtropical climate. Journal of Contaminant Hydrology, 234. https://doi.org/10.1016/j.jconhyd.2020.103684

Thorn, K.A., & Aiken, G.R. (1998). Biodegradation of crude oil into nonvolatile organic acids in a contaminated aquifer near Bemidji, Minnesota. Organic Geochemistry, 29(4), 909–931. https://doi.org/10.1016/S0146-6380(98)00167-3

Thouement, H. A. A., & Van Breukelen, B. M. (2020). Virtual experiments to assess opportunities and pitfalls of CSIA in physical-chemical heterogeneous aquifers. https://doi.org/10.1016/j.jconhyd.2020.103638

Todd, D. K. (1980). Groundwater Hydrology (2nd ed.). John Wiley & Sons

Trigueros, D.E.G., Módenes, A.N., Kroumov, A.D., & Espinoza-Quiñones, F.R. (2010). Modeling of biodegradation process of BTEX compounds: Kinetic parameters estimation by using Particle Swarm Global Optimizer. Process Biochemistry, 45(8), 1355–1361. https://doi.org/10.1016/J.PROCBIO.2010.05.007

U.S. Environmental Protection Agency [US EPA] (2003). A Review of Emerging Sensor Technologies for Facilitating Long-Term Ground Water Monitoring of Volatile Organic Compounds

U.S. Environmental Protection Agency [US EPA] (2024). National Primary Drinking Water Regulations. https://www.epa.gov/ground-water-and-drinking-water/national-pri mary-drinking-water-regulations

U.S. Geological Survey. (n.d.). USGS Water Data for the Nation. U.S. Department of the Interior. Retrieved from https://waterdata.usgs.gov/nwis (on 6 August 2024)

Uçankuş, T., & Ünlü, K. (2008). The effect of aquifer heterogeneity on natural attenuation rate of BTEX. Environmental Geology, 54(4), 759–776. https://doi.org/10.1007/s00254-007-0861-0

Vaezihir, A., Zare, M., Raeisi, E., Molson, J., & Barker, J. (2012). Field-scale modeling of benzene, toluene, ethylbenzene, and xylenes (BTEX) released from multiple source zones. Bioremediation Journal, 16(3), 156–176. https://doi.org/10.1080/10889868.2012.687415

Valsala, R., & Govindarajan, S.K. (2018a). Mathematical Modeling on Mobility and Spreading of BTEX in a Discretely Fractured Aquifer System Under the Coupled Effect of Dissolution, Sorption, and Biodegradation. Transport in Porous Media, 123(2), 421–452. https://doi.org/10.1007/s11242-018-1049-7

Valsala, R., & Govindarajan, S.K. (2018b). Interaction of dissolution, sorption and biodegradation on transport of BTEX in a saturated groundwater system: Numerical modeling and spatial moment analysis. Journal of Earth System Science, 127(4). https://doi.org/10.1007/s12040-018-0950-3

Valsala, R., & Govindarajan, S.K. (2019). Co-colloidal BTEX and Microbial Transport in a Saturated Porous System: Numerical Modeling and Sensitivity Analysis. Transport in Porous Media, 127(2), 269–294. https://doi.org/10.1007/s11242-018-1191-2

Van Breukelen, B.M., Griffioen, J., Röling, W.F.M., & Van Verseveld, H.W. (2004). Reactive transport modelling of biogeochemical processes and carbon isotope geochemistry inside a landfill leachate plume. Journal of Contaminant Hydrology, 70(3–4), 249–269. https://doi.org/10.1016/J.JCONHYD.2003.09.003

Van Essen Instruments (2025a). TD-Diver. Retrieved from https://www.vanessen.com/products/data-loggers/td-diver/ (on 14 October 2025)

Van Essen Instruments (2025b). Baro-Diver. Retrieved from https://www.vanessen.com/products/data-loggers/baro-diver/ (on 14 October 2025)

Van Groeningen, N., Thomasarrigo, L.K., Byrne, J.M., Kappler, A., Christl, I., & Kretzschmar, R. (2020). Interactions of ferrous iron with clay mineral surfaces during sorption and subsequent oxidation. Environmental Science: Processes & Impacts, 22(6), 1355–1367. https://doi.org/10.1039/D0EM00063A

Vapnik, V. (1982) Estimation of Dependences Based on Empirical Data (Springer Series in Statistics). Springer-Verlag, New York. ISBN 3-540-90733-5

Varjani, S.J., & Upasani, V.N. (2017). A new look on factors affecting microbial degradation of petroleum hydrocarbon pollutants. In International Biodeterioration and Biodegradation (Vol. 120, pp. 71–83). Elsevier Ltd. https://doi.org/10.1016/j.ibiod.2017.02.006

Varjani, S.J., Gnansounou, E., & Pandey, A. (2017). Comprehensive review on toxicity of persistent organic pollutants from petroleum refinery waste and their degradation by microorganisms. Chemosphere, 188, 280–291. https://doi.org/10.1016/j.chemosphere.2017.09.005

Vasudevan, M., Johnston, C.D., Bastow, T.P., Lekmine, G., Rayner, J.L., Nambi, I.M., Suresh Kumar, G., Ravi Krishna, R., & Davis, G.B. (2016). Effect of compositional heterogeneity on dissolution of non-ideal LNAPL mixtures. Journal of Contaminant Hydrology, 194, 10–16. https://doi.org/10.1016/j.jconhyd.2016.09.006

Vasudevan, M., Suresh Kumar, G., & Nambi, I.M. (2014). Numerical modelling of multicomponent LNAPL dissolution kinetics at residual saturation in a saturated subsurface system. Sadhana 39, 1387–1408 (2014). https://doi.org/10.1007/s12046-014-0282-1

Veil, J.A., Puder, M.G., Elcock, D., & Redweik Jr., R.J. (2004). A white paper describing produced water from production of crude oil, natural gas, and coal bed methane. https://doi.org/10.2172/821666

Vencelides, Z., Sracek, O., & Prommer, H. (2007). Modelling of iron cycling and its impact on the electron balance at a petroleum hydrocarbon contaminated site in Hnevice, Czech Republic. Journal of Contaminant Hydrology, 89(3–4), 270–294. https://doi.org/10.1016/j.jconhyd.2006.09.003

Villatoro-Monzón, W.R., Mesta-Howard, A.M., & Razo-Flores, E. (2003). Anaerobic biodegradation of BTEX using Mn(IV) and Fe(III) as alternative electron acceptors. https://iwaponline.com/wst/article-pdf/48/6/125/423510/125.pdf

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., …, and SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17(3), 261-272

Vitens (2021). GitHub - Vitens/phreeqpython: Object-oriented python wrapper for the VIPhreeqc module. https://github.com/Vitens/phreeqpython/tree/master

Vroblesky, D.A., & Chapelle, F.H. (1994). Temporal and spatial changes of terminal electron-accepting processes in a petroleum hydrocarbon-contaminated aquifer and the significance for contaminant biodegradation. Water Resources Research, Vol. 30, Issue 5. https://doi.org/10.1029/94WR00067

Wang, J., Shen, X., (2006). Estimation of generalization error: random and fixed inputs. Stat. Sin. 16, 569–588.

Watanabe, S. (2023). Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance. https://arxiv.org/abs/2304.11127v3

Węglarczyk, S. (2018). Kernel density estimation and its application. In: ITM Web of Conferences, 23. https://doi.org/10.1051/ITMCONF/20182300037

Williams, M.D., & Oostrom, M. (2000). Oxygenation of anoxic water in a fluctuating water table system: an experimental and numerical study. Journal of Hydrology, Vol 230, Issues 1–2. https://doi.org/10.1016/S0022-1694(00)00172-4

Wilson, S., Zhang, H., Burwell, K., Samantapudi, A., Dalemarre, L., Jiang, C., Rice, L., Williams, E., Naney, C. (2013). Leaking Underground Storage Tanks and Environmental Injustice: Is There a Hidden and Unequal Threat to Public Health in South Carolina? Environ Justice. 6(5):175-182. doi: 10.1089/env.2013.0019

Wisotzky, F. (2000). Redox reactions, multi-component stability diagrams and isotopic investigations in sulfur and iron-dominated groundwater systems. Redox Fundamentals, Processes and Applications, pp. 152–160, Springer-Verlag, New York

Wong, K.T., Jang, S.B., Yoon, S. Y., Ryu, B., Valiyaveettil Basheer, R., Abd Rahman, N., Choong, C.E., Jung, J., & Jang, M. (2024). Comparative Analysis of Reliability in On-Time Monitoring Data for NO3-N, BTEX, and TOC: Commercialized Sensors versus Spectroscopic Methods. ACS ES and T Water. https://doi.org/10.1021/acsestwater.4c00487

Wongbunmak, A., Khiawjan, S., Suphantharika, M., & Pongtharangkul, T. (2020). BTEX biodegradation by Bacillus amyloliquefaciens subsp. plantarum W1 and its proposed BTEX biodegradation pathways. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-74570-3

Wu, J., Chen, X.Y., Zhang, H., Xiong, L.D., Lei, H., & Deng, S.H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. Journal of Electronic Science and Technology, 17(1), 26–40. DOI: 10.11989/JEST.1674-862X.80904120

Wu, P., Comte, J.C., Zhang, L., Wang, S., & Chang, B. (2021). Effect of Surface Water Level Fluctuations on the Performance of Near-Bank Managed Aquifer Recharge from Injection Wells. Water 2021, Vol. 13, Page 3013, 13(21), 3013. https://doi.org/10.3390/W13213013

Wu, Q., Zhang, X., & Zhang, Q. (2017). Current situation and control measures of groundwater pollution in gas station. IOP Conference Series: Earth and Environmental Science, 94(1). https://doi.org/10.1088/1755-1315/94/1/012005

Wuebbles, D.J., & Jain, A.K. (2001). Concerns about climate change and the role of fossil fuel use. Fuel Processing Technology, 71(1–3), 99–119. https://doi.org/10.1016/S0378-3820(01)00139-4

Xiao, Z., Gang, W., Yuan, J., Chen, Z., Li, J., Wang, X., Feng, X. (2022). Impacts of data preprocessing and selection on energy consumption prediction model of HVAC systems based on deep learning. Energ. Buildings 258, 111832. https://doi.org/10.1016/J.ENBUILD.2022.111832

Xu, Y., Goodacre, R. (2018). On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. J. Anal. Test. 2, 249–262. https://doi.org/10.1007/s41664-018-0068-2.

Xu, Z., Chai, J., Wu, Y., & Qin, R. (2015). Transport and biodegradation modeling of gasoline spills in soil–aquifer system. Environmental Earth Sciences, 74(4), 2871–2882. https://doi.org/10.1007/s12665-015-4311-0

Xu, Z., Serata, R., Wainwright, H., Denham, M., Molins, S., Gonzalez-Raymat, H., Lipnikov, K., Moulton, J.D., & Eddy-Dilek, C. (2022). Reactive transport modeling for supporting climate resilience at groundwater contamination sites. Hydrology and Earth System Sciences, 26(3), 755–773. https://doi.org/10.5194/HESS-26-755-2022

Yao, L. & Ge, Z. (2021). Dynamic Features Incorporated Locally Weighted Deep Learning Model for Soft Sensor Development. IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-11, 2021, Art no. 2508011. DOI: 10.1109/TIM.2021.3073702

Yao, Y., Tu, C., Hu, G., Zhang, Y., Cao, H., Wang, W., Wang, W. (2024). Groundwater hydrochemistry and recharge process impacted by human activities in an Oasis–Desert in Central Asia. Water 16 (5), 763. https://doi.org/10.3390/ W16050763

You, J., Du, M., Chen, H., Zhang, X., Zhang, S., Chen, J., Cheng, Z., Chen, D., & Ye, J. (2018). BTEX degradation by a newly isolated bacterium: Performance, kinetics, and mechanism. International Biodeterioration and Biodegradation, 129, 202–208. https://doi.org/10.1016/j.ibiod.2018.02.012

Zammouri, M., & Ribeiro, L. (2017). Analyzing the effect of transmissivity uncertainty on the reliability of a model of the northwestern Sahara aquifer system. Journal of African Earth Sciences, 129, 910–922. https://doi.org/10.1016/J.JAFREARSCI.2017.02.034

Zanello, V., Scherger, L.E., & Lexow, C. (2021). Assessment of groundwater contamination risk by BTEX from residual fuel soil phase. SN Applied Sciences, 3(3), 1–20. https://doi.org/10.1007/S42452-021-04325-W

Zhang, A., Lipton, Z.C., Li, M., & Smola, A.J. (2023). Dive into deep learning. Cambridge University Press. https://d2l.ai

Zhang, T., Tremblay, P.L., Chaurasia, A.K., Smith, J.A., Bain, T.S., & Lovley, D.R. (2013). Anaerobic Benzene Oxidation via Phenol in Geobacter metallireducens. Applied and Environmental Microbiology, 79(24), 7800. https://doi.org/10.1128/AEM.03134-13

Zhang, Z. (2025). Enhancing distributed machine learning through data shuffling: techniques, challenges, and implications. In: ITM Web of Conferences, 73. https:// doi.org/10.1051/ITMCONF/20257303018, 03018

Zheng, C., Hill, M.C., Cao, G., & Ma, R. (2012). MT3DMS: Model use, calibration, and validation. Transactions of the ASABE, 55(4), 1549–1559. https://doi.org/10.13031/2013.42263

Zhu, X., Wu, X. (2004). Class noise vs. attribute noise: a quantitative study. Artif. Intell. Rev. 22 (3), 177–210. https://doi.org/10.1007/S10462-004-0751-8.

Ziegler, B.A., Schreiber, M.E., & Cozzarelli, I.M. (2017). The role of alluvial aquifer sediments in attenuating a dissolved arsenic plume. Journal of Contaminant Hydrology, 204, 90–101. https://doi.org/10.1016/J.JCONHYD.2017.04.009

Zyphur, M.J., & Oswald, F.L. (2015). Bayesian Estimation and Inference. Journal of Management, 41(2), 390–420. https://doi.org/10.1177/0149206313501200

Zytner, R.G. (1994). Sorption of benzene, toluene, ethylbenzene and xylenes to various media. Journal of Hazardous Materials, 38(1), 113–126. https://doi.org/10.1016/0304-3894(94)00027-1

# Acknowledgements

A friend once blurted out that the majority of people who chance upon a PhD manuscript tend to read solely the acknowledgements section. With this burden in mind, I now ought to write something worth reading. Perhaps something far more appealing to the public than the book itself might appear, at least to those less inclined toward its subject.

To bask in the glory of the dying sun is to remember that ephemeral moments, such as the culmination of years of quiet labor, make the long winters worthwhile.

Indeed, we were told that the path less traveled might lead to a unique and fulfilling life. But seldom is it said that the untrodden path is full of thorns, nor that the greater the ambition, the greater the sacrifice.

One could say that the willingness to brave such thorns separates a simple journey from a grand adventure.

To those who aided me through thorny roads, who shouldered my sacrifices when the burden was too heavy, and who encouraged me when the leap into the unknown demanded more courage than I could muster, yet less than my ambitions required, I dedicate this to you.

Boris, Martijn, and Luuk, the best combination of promotors and supervisors any candidate could ask for: you have my sincerest, complete, and unending gratitude, even when at times it felt as though you were the ones placing the heaviest stones along my path. You might not have always been the light that guided me through the seemingly unending dark tunnel, but you were certainly the ones who taught me not only how to seek the light strategically, but also how to create it. For that, and for everything else, none of this would have been possible without you.

From scrambled words and full-on ramblings to meaningful sentences, Nadia, your help in giving form to a disorganized writer's thoughts was nothing short of a miracle. A million and twenty-one thanks!

For helping bring this work together in one way or another, Karel and Mateo, your contributions remain woven into every paragraph.

At times, the couch seemed better company than the words on paper, accompanied by the voices of unfinished work inside my head. On these days, when I felt like I was drowning and constantly gasping for air to breath, my Wetsus mates were my limbs, allowing me not only to swim but to thrive.

Cheers to my officemates who made my first days at Wetsus feel like a lovely stroll through a park. You were the people I knew I could lean on throughout my journey, and who brought a breath of fresh air to the workplace. May the dream you hold dearest in your hearts come true.

They say that ‘good fences make good neighbors.’ My gratitude goes to the neighboring offices who have respected my fences yet never withheld their table from me. Our interactions may span from a day to several years, but the warmth of spring is comparable to your presence; may it always remain with you.

Without a heart, the human body cannot survive. To the hearts of Wetsus who kept it alive, from the laboratory to the reception, from HR to the canteen with river of soups, you have my deepest respect and gratitude.

Other days, the storm seems never-ending, bringing nothing but darkness and destruction along its path. But alas, there are friends who take your hand, pull you out into the rain, and dance with you until the cold wind and dark clouds turn into a warm breeze on a cloudless summer afternoon. This, too, I dedicate to you.

Thank you for the early morning talks and late-night conversations, for being with me at the depths of the sea to the heights of the mountains, both literally and figuratively. We may be apart, but friendship removes all distance: Raya, Tin, Benet, and Janine. Let your aspirations and dreams run freer and farther than your imagination can reach. We will always strive to be better than we were.

How strange my world would be if family did not exist; like a bird trying to soar without feathers, like a waterfall without its flow, or a flower without colors. They give meaning to the humming of bees, the rustle of autumn leaves, and the quiet drifting of clouds in the hollow evening. To my parents and my siblings, how strange my world would be without you, indeed.

How often can a heart be broken before it can no longer be mended? To those who joined me in this journey and continue to seek the answer, may the answer remain forever unknown. Your contribution to this work extends beyond what I can acknowledge.

Some words are better left unsaid, as the greatest compliment cannot be chained by mere human language. Thus, to those of you whose roles were not mentioned but who have contributed to the realization of this book, however great or minuscule the effort may have been, I carry your contribution with more than words can express.

And to you readers, who have yet to meet me but have come to know me through my research, I share my hope that one day this work may serve as a cog in the machinery of sustainability and accessibility. May all the little, seemingly insignificant things bring you significant joy and wisdom.

Above all, it was God’s guidance that carried me through paths I could not have walked alone.

# About the Author

A native of Cabanatuan City, Nueva Ecija, Philippines, Chen Lester Reñon Wu was born on 21 September 1994. He began his bachelor's degree in Agricultural and Biosystems Engineering in 2011 at the University of the Philippines Los Baños, Laguna. In 2015, he participated in a short exchange program at the University of Tsukuba in Japan, where he studied wastewater treatment processes.

He subsequently worked as a researcher at the same university, developing a wireless sensor network-based water information system for efficient irrigation water management in the Philippines.

Lester later pursued his master's degree in Water Science and Engineering at IHE Delft Institute for Water Education in the Netherlands. With his thesis entitled "Investigating the Occurrence and Hydrochemistry of Fresh-Saline Waters as a Function of Recharge and Hydrogeological Setting on Delft Island – A Semi-Arid Coral Reef Island in Sri Lanka", he obtained his MSc diploma in 2020.

In 2021, he became a PhD candidate at Wetsus, European Centre of Excellence for Sustainable Water Technology in Leeuwarden, in collaboration with Delft University of Technology, the Netherlands. There, he researched the remote monitoring of petroleum hydrocarbons in groundwater through sensor data fusion using reactive transport modeling and machine learning. The results of this work are presented in this thesis.

# List of Publications

Wu, C.L.R., Wagterveld, R.M., Rietveld, L.C., van Breukelen, B.M. (under review). Real-time Monitoring of Petroleum Hydrocarbons in Groundwater using Hybrid Machine Learning Architectures. Hydrology and Earth System Sciences. https://doi.org/10.5194/egusphere-2025-5842 [preprint]

Wu, C.L.R., Wagterveld, R.M., Rietveld, L.C., van Breukelen, B.M. (2025). Machine learning-based in-situ detection of toxic petroleum hydrocarbons in groundwater. Journal of Contaminant Hydrology. 104771, ISSN 0169-7722. https://doi.org/10.1016/j.jconhyd.2025.104771

Wu, C. L. R., Wagterveld, R. M., & van Breukelen, B. M. (2024). Reactive transport modeling for exploring the potential of water quality sensors to estimate hydrocarbon levels in groundwater. Water Resources Research, 60, e2023WR036644. https://doi.org/10.1029/2023WR036644

The work described in this thesis was performed in the cooperation framework of Wetsus, European Centre of Excellence for Sustainable Water Technology (*www.wetsus.nl*). Wetsus is co-funded by the European Union (Horizon Europe, LIFE, Interreg and EDRF), the Province of Fryslân and the Dutch Government: Ministry of Economic Affairs (TTT, SBO & PPS-I/TKI Water Technology), Ministry of Education, Culture and Science (TTT & SBO) and Ministry of Infrastructure and Water Management (National Growth Fund - UPPWATER). This research received funding from the Netherlands Organisation for Scientific Research (NWO) in the framework of the "Sustainable water technology Partnership Programme NWO-Wetsus on the Impact of Water Research on Energy, Industry, Health, Economy and Environment 2019. The authors like to thank the participants of the research theme "Monitoring & Quality" for the fruitful discussions and their financial support.