



Modeling PFAS Transport in Groundwater

Exploring current approaches and evaluating
parameter importance

Clara Eklund

MASTER'S THESIS 30HP

ISRN LUTVDG/(TVT—5180)/1-65/(2023)

ENGINEERING GEOLOGY

Faculty of Engineering

Lunds University



Modeling PFAS transport in groundwater

Exploring current approaches and evaluating
parameter importance

Clara Eklund



LUNDS
UNIVERSITET

MASTER'S THESIS

Submitted to the Division of Engineering Geology, Faculty of Engineering, Lund
University in Partial Fulfilment of the Requirements for the Degree of Master of Science
in Water Resources Engineering.

Lund 2023

Cover picture: Water on a repellent fabric, probably containing PFAS. Photo: Swedish Chemical agency, 2021

Lund University, Faculty of Engineering
Division of Engineering Geology

Modeling PFAS transport in groundwater - Exploring current approaches and evaluating parameter importance

Modellering av PFAS i grundvattnet - Undersökning om vilka metoder som används och parametrarnas betydelse

Author: Eklund, Clara

Supervisors: Lévy, Léa (Engineering Geology, LTH), Erdal, Daniel (Tyréns)

Examiner: Rosberg, Jan-Erik (Engineering Geology, LTH)

ISRN LUTVDG/(TVTG—5180)/1-65/(2023)

Keywords: PFAS; Groundwater modeling; FEFLOW; Python; Machine learning
Language: English

The work is carried out in collaboration with Tyréns

Digital edition Lund 2023

Abstract

PFAS contamination in drinking water is a current problem, and new regulations for drinking water limits were recently implemented in each member country of the EU in January this year. Understanding the spread of PFAS in groundwater is therefore important to prevent it from reaching drinking water sources. Groundwater modeling is a valuable tool for this purpose. However, due to the specific characteristics of PFAS, such as sorption, there is a knowledge gap regarding the optimal implementation of this tool. Therefore, this thesis aims to investigate the currently used modeling strategies and examine the parameters that seem to have the most significant impact on PFAS transport. This research also seeks to understand the importance of identifying the specific PFAS compounds present in the contaminant.

A literature review was conducted, revealing that the most commonly used method involved transport simulations using MODFLOW. As a result, a similar approach using FEFLOW was adopted with the aim of conducting a sensitivity analysis to identify which parameters are most crucial in determining the obtained PFAS concentration in each node. A total of 500 simulations were performed for four different scenarios, considering two different aquifers and the transport of two different PFAS compounds. The results of these simulations were used to train a random forest regression model, which exhibited a high level of accuracy in predicting the resulting concentration at specific nodes. A sensitivity analysis was conducted on the model, revealing that hydraulic conductivity was the most important parameter in steady-state modeling, followed by recharge for sandy aquifers and longitudinal dispersivity for sand and gravel aquifers. Sorption did not have a significant impact on the results in this context. However, in transient modeling, sorption was found to be of great importance, suggesting that the specific PFAS compound may play a significant role in such scenarios. Lastly, the results differed significantly between steady-state and transient-state modeling, indicating that the choice of modeling approach is also of great importance.

Keywords: *PFAS, Groundwater modeling, FEFLOW, Python, Machine learning*

Preface

With this work, I end my studies at the Civil Engineering program in Environmental Engineering at Lund University. The master's thesis corresponds to 30 university credits and has been carried out in the spring/summer of 2023. The work has been carried out at the Department for Engineering Geology, Faculty of Engineering, Lund University, together with Tyréns.

First and foremost I want to thank both my supervisors, Daniel Erdal (Tyréns) and Léa Lévy (LTH), for all their support through out the project and for all the interesting conversations and valuable comments. Also, thank you to my examiner, Jan-Erik Rosberg for your constructive feedback.

I would also like to thank Dan Svenonius for your big support and help through out the work. Lastly, I would like to give a big thank you to my classmates, who made my 5 years in Lund incredibly memorable.

List of Acronyms

AWI	Air-water interface
CEC	Cation exchange capacity
CMC	Critical micelle concentration
Cond	Hydraulic conductivity
C_x	Fluorocarbon chain of length x
DOC	Dissolved organic carbon
FASA	Perfluoroalkyl sulfonamides
FPs	Fluoropolymers
FTOH	Fluorotelomer alcohol
FTSA	Fluorotelomer sulfonic acids
K_d	Soil-water distribution coefficient
K_f	Freundlich sorption capacity parameter
K_{oc}	Organic carbon normalised partition coefficient
L_{dis}	Longitudinal dispersivity
n	Freundlich non-linearity parameter
NOM	Natural organic matter
Mse	Mean square error
PASF	Perfluoroalkane sulfonyl fluoride
PFAA	Per- or polyfluoroalkyl ether acids
PFAI	Perfluoroalkyl iodides
PFAS	Per- or polyfluoroalkyl substances
PFBA	Perfluorobutanoic acid
PFBS	Perfluorobutanesulfonic acid
PFCA	Perfluorinated carboxylate acids
PFDA	Perfluorodecanoic acid
PFDS	Perfluorodecanesulfonic acid
PFDoDA	Perfluorododecanoic acid
PFDoDS	Perfluorododecanesulfonic acid
PFHpA	Perfluoroheptanoic acid
PFHpS	Perfluoroheptanesulfonic acid
PFHxA	Perfluorohexanoic acid
PFHxS	Perfluorohexanesulfonic acid
PFNA	Perfluorononanoic acid
PFNS	Perfluorononanesulfonic acid
PFOA	Perfluorooctanoic acid
PFOS	Perfluorooctanesulfonic acid
PFOSA	Perfluorooctane sulfonamide
PFPA	Pentafluoropropionic anhydride
PFPeA	Perfluoropentanoic acid
PFPE	Perfluoropolyethers
PFPIA	Perfluorohexylperfluorooctylphosphinic acid

PFSA	Perfluorinated sulfonic acids
PFTrDa	Perfluorotridecanoic acid
PFTrDS	Perfluorotridecanesulfonic acid
PFUnDA	Perfluorounddecanoic acid
Poro	Porosity
R²	Coefficient of determination
Rec	Groundwater recharge
Rmse	Root mean square error
SOM	Soil organic matter
Sorp	Sorption
T_{dis}	Transversal dispersivity

Table of Contents

1	Introduction	1
1.1	Objectives	2
1.2	Structure	2
2	Background	3
2.1	Per- and Polyfluorinated Substances	3
2.1.1	Classification	3
2.1.2	Regulations	5
2.1.3	Chemical structure	6
2.1.4	Surfactants	7
2.1.5	Transformation and degradation	7
2.1.6	Solubility	8
2.1.7	Spreading and exposure	8
2.2	Transport and fate of PFAS	9
2.2.1	Advection, Dispersion and Diffusion	9
2.2.2	Sorption	11
2.2.3	Micelle formation	14
2.2.4	Precursors	14
2.3	Modeling Strategies	15
2.3.1	FEFLOW	15
2.3.2	Analysis of modeling results	22
3	Methods for simulating transport of PFAS in the literature	25
3.1	Method	25
3.2	Results	25
4	Methodology	27
4.1	Modeling	27
4.1.1	FEFLOW model	28
4.1.2	Parameters	29
4.1.3	Parameter interval	32
4.1.4	Python	33
4.1.5	FEFLOW simulations	34
4.1.6	Random forest regression model	34
4.1.7	Sensitivity analysis	34
5	Result	36
5.1	Steady state vs Transient state	36
5.2	Shapley values	37
5.3	Random forest importance	39
5.4	Concentration	41
6	Discussion	43
6.1	Methodology	43
6.2	Concentrations	43

6.3	Importance of parameters	44
6.4	Modeling Insights and Uncertainties	45
6.5	Future work	46
7	Conclusion	47
	References	56
	Appendix A - Henry's sorption coefficient	57
	Appendix B - Longitudinal dispersivity	59
	Appendix C - Transversal dispersivity and hydraulic conductivity	61
	Appendix D - Shapley values	63
	Appendix E - Random forest importance	65

List of Figures

1	<i>General classification of per- and polyfluoroalkyl substances (PFAS). Figure modified from Ambaye et al., 2022.</i>	3
2	<i>The left image illustrates the phenomena behind dispersion, due to the different path length possible for a contaminant to travel. The right image shows the effect of dispersion on the concentration plume, where S represents the source, Adv represents the transport that would occur with only advection, D_L represents the dispersion in the longitudinal direction, and D_T represents the transport in the transversal direction.</i>	10
3	<i>A schematic image illustrating how a negative charged PFAS can be sorbed to the soil or sorbent through either electrostatic interactions or hydrophobic interactions. This Figure is modified from Figure 1 and 2 in Du et al. (2014).</i>	12
4	<i>Conceptual image of how the PFAS molecules can sorb to the soil or sorbent due to the formation of different types of micelles. Figure is modified from Figure 5.2 in Interstate Technology Regulatory Council (2022).</i>	15
5	<i>A schematic overview of the basis behind the random forest regression model is shown. The dark grey nodes indicate the path down each decision tree. At each internal node, a decision is made whether to go left or right down the tree. Finally, one of the leaf nodes is reached, which provides a prediction of the resulting concentration. The random forest regression algorithm derives its final prediction from the average value obtained from all the created decision trees.</i>	23
6	<i>Flowchart describing the methodology used for the modeling.</i>	27
7	<i>The 3D model used in FEFLOW is divided into two layers: an upper layer (blue) and an underlying layer (grey). However, for the simulations, these layers were treated as one homogeneous layer. The red area represents the source of contamination. Three observation points are marked by flags: the blue flag represents observation point 1, the red flag indicates observation point 2, and the black flag signifies observation point 3. The edge of the model represents the shoreline.</i>	29
8	<i>The result obtained from one simulation in steady state and one simulation in transient state. The left image shows the result from the steady state model and the right image show the result from the transient state model (running time of 50 years).</i>	36
9	<i>The Shapley values for the four different scenarios modeled in steady state, in observation point 1. Blue color indicates low values of the parameters, while the pink color indicates high values (as the scale to the right present). The points located on the left side of the y-axis indicated a negative contribution to the predicted output concentration, while points on the right side indicates a positive contribution.</i>	38
10	<i>The Shapley values for the four different scenarios modeled in transient state, in observation point 3. Blue color indicates low values of the parameters, while the pink color indicates high values (as the scale to the right present). The points located on the left side of the y-axis indicated a negative contribution to the predicted output concentration, while points on the right side indicates a positive contribution.</i>	39
11	<i>The heat-map for the random forest importance values obtained for the five different models in observation point 3.</i>	40
12	<i>The heat-map for the random forest importance values obtained for the five different models in observation point 1.</i>	41

13	<i>The concentration of PFOS in the observation points for the 50 first simulations of PFOS in a sandy aquifer modeled in steady state.</i>	42
14	<i>The concentration of PFOS in the observation points for the 50 first simulations of PFOS in a sandy aquifer modeled in transient state.</i>	42
A	<i>The Shapley values for the four different scenarios in Observation point 2. Blue color indicates low values of the parameters, while the pink color indicates high values (as the scale to the right present).</i>	63
B	<i>The Shapley values for the four different scenarios in Observation point 3. Blue color indicates low values of the parameters, while the pink color indicates high values (as the scale to the right present).</i>	64
C	<i>The heat-map for the random forest importance values obtained for the five different models in observation point 2.</i>	65

List of Tables

1	<i>Some om the most important groups of PFASs and their chemical structure(ITRC, 2020b). Molecular structures taken from Ahrens, 2011.</i>	4
2	<i>Previous and current threshold values for different groups of PFAS in Sweden (Livsmedelsverket, 2022; Livsmedelsverket et al., 2021).</i>	6
3	<i>Minimum, maximum and average values found for the longitudinal dispersivity in two different types of soil. eq indicates that the values have been calculated by using Equation (10) together with the values of the scale, belonging to the obtained values of α_L. The values of α_L, together with the belonging scale, were obtained from Schulze-Makuch, 2005.</i>	30
4	<i>Minimum, maximum and average values found for the transversal dispersivity in two different types of soil. The data which this is based on can be found in Appendix C.</i>	30
5	<i>Minimum, maximum and average values found for the relationship between longitudinal and transversal dispersivity (α_T / α_L) in two different types of soil. The data which this is based on can be found in Appendix C.</i>	30
6	<i>Dry density, together with minimum and maximum values for porosity for two different types of soil. Values are obtained from: ¹ Yu, Loureiro, et al., 1993, ² Yu, Cheng, et al., 2023 and ³Zhang, Xiao, et al., 2017.</i>	31
7	<i>Maximum, minimum and mean values of Henry’s sorption coefficient for two different PFASs: PFOS and PFOA. Together with 10th and 90th percentile values. The data which this is based on can be found in Appendix A.</i>	31
8	<i>Minimum, maximum and average values found for the hydraulic conductivity in two different kinds of soil. The values where obtained from Gelhar et al., 1992.</i>	32
9	<i>Values for estimated maximum and minimum groundwater recharge in Sweden for three different soil types. Data is obtained from Rodhe et al., 2006.</i>	32
10	<i>The range of the parameter values entered into Python for the simulation of the four different scenarios.</i>	33
11	<i>The R^2 values for all the different created random forest regression models.</i>	37
A	<i>Henry’s sorption coefficient for PFOS collected from several different sources. Density used for calculating K_h from K_d is obtained form ¹ Yu, Loureiro, et al., 1993 and ² Zhang, Xiao, et al., 2017, see table 6.</i>	57

B	<i>Henry's sorption coefficient for PFOA collected from several different sources. Density used for calculating K_h from K_d is obtained from ¹ Yu, Loureiro, et al., 1993 and ² Zhang, Xiao, et al., 2017, see table 6.</i>	58
C	<i>Values for the longitudinal dispersivity in sand found in the literature, together with the belonging scale. Alpha(eq) are the values obtained by inserting the scale together into equation (9).</i>	59
D	<i>Values for the longitudinal dispersivity in sand and gravel found in the literature, together with the belonging scale. Alpha(eq) are the values obtained by inserting the scale together into equation (9).</i>	60
E	<i>Values for the transversal dispersivity, the ration between D_T and D_L as well as the conductivity in sand. All the values were obtained during the literature review.</i>	61
F	<i>Values for the transversal dispersivity, the ration between D_T and D_L as well as the conductivity in sand. All the values were obtained during the literature review.</i>	62

1 Introduction

Per- and Polyfluorinated Substances (PFAS) cover a wide range of approximately 5,000 different chemicals that are produced within industrial environments. The production of PFAS began in the late 1940s, and since then, these substances have been extensively used in various products and chemicals. Examples of such products include cosmetics, textiles treated with water and grease repellents, fire extinguishing foam, non-stick coated cookware, and cleaning agents (Swedish EPA, 2023).

A PFAS molecule typically consists of a hydrophobic part, known as the fluorinated carbon chain, and a hydrophilic part, represented by its functional group (such as carboxylic or sulfonic acid). This unique chemical structure provides PFAS both hydrophobic and oleophobic properties, making them effective at repelling both water and grease (Swedish Chemical agency, 2021). However, the environmental downside of PFAS lies in their stability, attributed to the strong carbon-fluorine bond. This stability prevents their natural degradation, leading to accumulation in soil, water, and biota, as well as long-distance transport (Swedish Chemical agency, 2021). Consequently, PFAS can be detected almost everywhere in the environment, even in locations where no intentional emissions or use have occurred (Swedish EPA, 2023).

The main pathways through which PFAS enter groundwater include the use of fire extinguishing foam during fire drills and accidents, which directly emits PFAS into the environment. Additionally, PFAS can leach into groundwater from landfills, wastewater treatment plants, and industrial sites (Swedish EPA, 2023). Once PFAS contaminates groundwater, it can be transported over considerable distances, potentially reaching drinking water wells. Furthermore, PFAS can also be airborne and deposit into lakes and rivers, which are important sources of drinking water (Haug et al., 2011). In a study conducted by the Geological Survey of Sweden (SGU) from 2016 to 2017, which examined the presence of environmental toxins in municipal groundwater resources, PFAS was found in two-thirds of the groundwater resources, although at relatively low levels. A more comprehensive assessment conducted under the Swedish Water Framework Directive identified around 250 groundwater resources at a significant risk of PFAS impact (SGU, 2021). Starting from the beginning of 2023, the limit values for PFAS in Swedish drinking water will be lowered to 4 ng/L for PFAS-4 and 100 ng/L for PFAS-21 (Livsmedelsverket, 2022). This highlights the relevance of addressing PFAS contamination in drinking water and emphasizes the importance of understanding its groundwater transport to anticipate and prevent further contamination.

Groundwater modeling serves as an effective tool for tracking the spread of PFAS contamination. However, due to the specific characteristics of PFAS, there is limited knowledge regarding the application of this tool for simulating the transport. Furthermore, there are many uncertainties associated with determining several of the input parameters used in groundwater modeling. Therefore, it would be highly valuable to understand the extent to which these different parameters impact the resulting outcomes.

1.1 Objectives

Due to the aforementioned reasons, it is important to gain an understanding of how groundwater modeling can be used for simulating the transport of PFAS in groundwater. Therefore, the objective of this thesis is to investigate the modeling of PFAS transport in groundwater. The study aims to explore both the current modeling methods used and the choice of parameter values, as well as to assess the importance these parameters have on the resulting concentration in the model.

Thereby, this thesis seeks to address the following research questions:

- What are the commonly used methods for modeling PFAS transport?
- Which parameters significantly influence the simulation of PFAS transport within the groundwater?
- Do the specific characteristics of different PFAS compounds impact the results, or can they simply be modeled as a single compound?

By addressing these questions, the thesis aims to contribute to the understanding and improvement of PFAS transport modeling in the groundwater.

1.2 Structure

Chapter 2 presents the background, starting with an overview of PFAS characteristics and the previous and current regulations. Additionally, it provides an introduction to modeling strategies in both FEFLOW and Python.

Chapter 3 describes the conducted literature review, including the methodology used and the obtained results.

Chapter 4 explains the methodology employed for the modeling and analysis.

Chapter 5 presents the obtained results.

Chapter 6 discusses and analyzes the results, and proposes further areas for investigation.

Finally, Chapter 7 presents the conclusions drawn from the study.

2 Background

In this section, the necessary background is provided. First of all, some theory about PFAS is given. The classification of PFAS molecules, as well as their chemical structure, is investigated. Furthermore, various properties of PFAS, such as surfactants, transformation, degradation, and solubility, are reviewed. The spreading and exposure of PFAS are also presented, along with the current regulations of PFAS in both the European Union and Sweden. Moreover, the transport and fate of PFAS are examined. Lastly, some theory of modeling strategies, both in FEFLOW and in Python, is presented.

2.1 Per- and Polyfluorinated Substances

2.1.1 Classification

PFAS is an abbreviation for Per- and Polyfluorinated Substances, which are characterized by the partial (poly) or complete (per) fluorination of their alkyl chains. This means that some or all of the hydrogen atoms on the carbon chains are replaced with fluorine atoms (Swedish EPA, 2023). PFAS can be classified into two main groups: polymers and non-polymers, each with additional subcategories. Polymers include fluoropolymers (FPs), side-chain fluorinated polymers, and perfluoropolyethers (PFPEs) (OECD, 2015). Non-polymers can be further categorized into perfluoroalkyl acids (PFAAs), perfluorohexylperfluorooctylphosphinic acid (PFPIA), pentafluoropropionic anhydride (PFPA), and perfluoroalkane sulfonyl fluoride (PASF) (Ambaye et al., 2022). Figure 1 provides the classification of per- and polyfluoroalkyl substances.

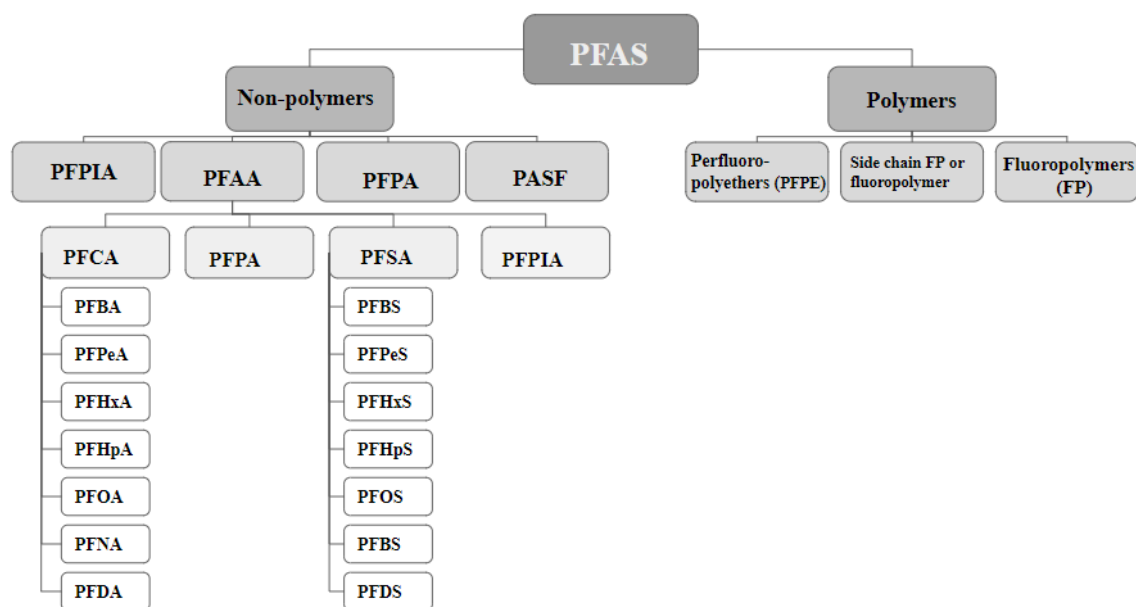
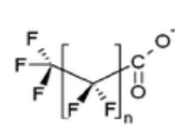
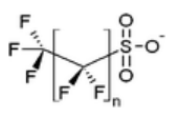
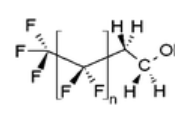
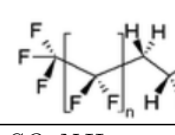
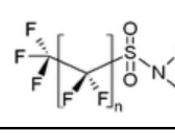


Figure 1: *General classification of per- and polyfluoroalkyl substances (PFAS). Figure modified from Ambaye et al., 2022.*

Polymers are formed by connecting several small, identical molecules together in a repeat-

Table 1: *Some of the most important groups of PFASs and their chemical structure (ITRC, 2020b). Molecular structures taken from Ahrens, 2011.*

Group	Functional group and molecule structure	Substance	Acronym	Chemical formula
Perfluoroalkyl carboxylic acid (PFCA)	$COOH$ 	Perfluorobutanoic acid	PFBA	$C_3F_7CO_2H$
		Perfluoropentanoic acid	PFPeA	$C_4F_9CO_2H$
		Perfluorohexanoic acid	PFHxA	$C_5F_{11}CO_2H$
		Perfluoroheptanoic acid	PFHpA	$C_6F_{13}CO_2H$
		Perfluorooctanoic acid	PFOA	$C_7F_{15}CO_2H$
		Perfluorononanoic acid	PFNA	$C_8F_{17}CO_2H$
		Perfluorodecanoic acid	PFDA	$C_9F_{19}CO_2H$
		Perfluoroundecanoic acid	PFunDA	$C_{10}F_{21}CO_2H$
		Perfluorododecanoic acid	PFDoDA	$C_{11}F_{23}CO_2H$
		Perfluorotridecanoic acid	PFTrDA	$C_{12}F_{25}CO_2H$
Perfluoroalkyl sulfonic acids (PFSA)	SO_3H 	Perfluorobutanesulfonic acid	PFBS	$C_4F_9SO_3H$
		Perfluoropentanesulfonic acid	PFPeS	$C_5F_{11}SO_3H$
		Perfluorohexanesulfonic acid	PFHxS	$C_6F_{13}SO_3H$
		Perfluoroheptanesulfonic acid	PFHpS	$C_7F_{15}SO_3H$
		Perfluorooctanesulfonic acid	PFOS	$C_8F_{17}SO_3H$
		Perfluorononanesulfonic acid	PFNS	$C_9F_{19}SO_3H$
		Perfluorodecanesulfonic acid	PFDS	$C_{10}F_{21}SO_3H$
		Perfluorododecanesulfonic acid	PFDoDS	$C_{12}F_{25}SO_3H$
		Perfluorotridecanesulfonic acid	PFTrDS	$C_{13}F_{27}SO_3H$
Fluorotelomer alcohols (FTOH)	CH_2CH_2OH 	Fluorotelomer alcohol	6:2 FTOH	$C_8H_5F_{13}O$
		Fluorotelomer alcohol	8:2 FTOH	$C_{10}H_5F_{17}O$
		Fluorotelomer alcohol	10:2 FTOH	$C_{12}H_5F_{21}O$
Fluorotelomer sulfonic acids (FTSA)	$CH_2CH_2SO_3H$ 	Fluorotelomer sulfonate	6:2 FTS	$C_8H_4F_{13}SO_3H$
Perfluoroalkane sulfonamides (FASA)	SO_2NH_2 	Perfluorooctane sulfonamide	PFOSA	$C_8F_{17}SO_2NH_2$

ing pattern. Compared to non-polymer PFAS (ITRC, 2020b), polymers are often larger molecules. The classification of polymer PFAS is based on their distinct chemical structures. Fluoropolymers, for instance, have a carbon-only polymer backbone with fluorine atoms directly attached to the carbon (ITRC, 2020b). PFPE, on the other hand, feature a carbon and oxygen polymer backbone with fluorine atoms directly attached to the carbon. This group of PFAS is relatively less known (ITRC, 2020b). In side-chain fluorinated polymers, the backbone is a nonfluorinated polymer, while the fluorinated carbon chains

act as side chains branching out from the backbone. Some of these substances can serve as precursors to PFAA, as they can degrade into PFAA (ITRC, 2020b).

PFAA represent some of the simplest PFAS molecules. Under normal conditions, they are generally non-degradable, hence often referred to as "terminal PFAS" since the degradation of longer PFAS halts once they break down into a PFAA (ITRC, 2020b). Precursors are polyfluoroalkyl substances that can degrade into PFAA (ITRC, 2020b). Furthermore, PFAA can be further classified into two major groups: Perfluoroalkane sulfonic acids (PFSA) and Perfluoroalkyl carboxylic acids (PFCA), distinguished by their functional groups, which is either sulfonic or carboxylic acid (ITRC, 2020b).

Moreover, Table 1 provides an overview of some of the most significant PFAS groups, including their chemical formulas and molecular structures.

2.1.2 Regulations

Until 2021, there were no legally binding limit values for PFAS in drinking water in Sweden or the European Union (EU). However, the updated EU Drinking Water Directive (2020/2184) now includes monitoring requirements for PFAS in drinking water. The directive sets minimum standards, allowing member countries to adopt stronger regulations if desired. The new regulations had to be implemented in each member country by January 2023 at the latest (Livsmedelsverket et al., 2021). The directive includes two limit values: one for the sum of a group of PFAS compounds known as PFAS20, with a threshold value of 100 ng/L, and one for the total amount of PFAS, with a threshold value of 500 ng/L. The specific substances to be included in the "total amount of PFAS" are yet to be decided, and each country can choose whether to implement both or just one of the limit values (Livsmedelsverket et al., 2021).

In Sweden, the Swedish Food Agency (Livsmedelsverket) has had an action limit value for PFAS in drinking water since 2014. Initially, this limit included seven different PFAS compounds (PFAS7), but it was later expanded to include eleven (PFAS11). In 2016, another action limit value was introduced, which included four different PFAS compounds (PFAS4) (Livsmedelsverket et al., 2021). Currently, Sweden has implemented the EU directive by establishing two threshold values: one for PFAS4 and one for PFAS21. PFAS21 includes the 20 PFAS compounds specified in the EU directive, along with the substance 6:2 FTS. The threshold value for PFAS4 is set to 4 ng/L and for PFAS21 to 100 ng/L (Livsmedelsverket, 2022).

Table 2 presents the historical and current threshold values for various groups of PFAS in Sweden. The predominant PFAS compounds found in the environment are PFOS and PFOA, followed by PFHxS and PFNA (Agency, 2019). PFOS has been commonly utilized in fire-fighting foam, cleaning detergents, and as an impregnating agent, whereas PFOA is primarily employed as an auxiliary chemical in the production of polytetrafluoroethylene (PTFE), well-known through trademarks such as Teflon and Gore-Tex (ALS Europe, 2023).

Table 2: Previous and current threshold values for different groups of PFAS in Sweden (Livsmedelsverket, 2022; Livsmedelsverket et al., 2021).

PFAS4	PFAS7	PFAS11	PFAS20/21
PFOA		PFOA	
PFOS		PFOS	
PFH _x S		PFH _x S	
PFNA		PFBS	
4 ng/L		PFPeA	
		PFH _x A	
		PFHpA	
	90 ng/L		PFNA
			6:2 FTOH
			PFBA
			PFDA
		90 ng/L	PFUnDA
			PFDoDA
			PFT _r DA
			PFPeS
			PFHpS
			PFNS
			PFDS
			PFDoDS
			PFT _r DS
			6:2 FTS
			100 ng/L

2.1.3 Chemical structure

The chemical structure of PFAS varies among different types of compounds. In addition to having a partly or fully fluorinated carbon chain (tail), PFAS also consist of a functional group (head) that varies between different PFAS, giving the molecules different properties.

Branched and linear

Depending on the manufacturing process used, different types of PFAS isomers may be produced. The structure of the PFAS molecule can be either branched or linear, which may impact the fate and transport of the molecules (ITRC, 2020b). A linear PFAS isomer has a straight carbon backbone, where the carbon atoms are solely bound to one or two other carbons. This results in only one possible linear isomer for each compound with the same number of carbons (ITRC, 2020b). However, there are several possible variations of branched isomers for each PFAS substance. In a branched isomer, at least one carbon is bonded to two or more carbon atoms, giving it a branched backbone structure (ITRC,

2020b). Studies have found that linear isomers have a greater sorption affinity than branched isomers, suggesting that branched isomers are more mobile (Xiao et al., 2022). Furthermore, the manufacturing process may also influence the types of PFAS formed through the transformation of precursor PFAS (ITRC, 2020b). Therefore, understanding the manufacturing process of emitted PFAS can be important when investigating their environmental fate and transport.

Chain length

PFAS, especially PFAA, can be categorized into short- or long-chain PFAS. The categorization varies among different types of PFAA, where long-chain PFCA contain 8 or more carbons, while long-chain PFSA contain six or more carbons. Short-chain PFCA have seven or fewer carbons, while short-chain PFSA have five or fewer carbons (ITRC, 2020b). The chain length has a significant impact on the molecules' affinity to sorb onto the soil, which will be discussed further in section 2.2.

Charge

Depending on the functional group, different PFAS molecules can possess different charges under natural environmental conditions. PFAS with an acid functional group commonly have a low acid dissociation constant, pK_a , due to the strong electron-withdrawing nature of fluorine atoms (Xiao et al., 2022). PFAA and other anionic PFAS have small pK_a values, resulting in them being deprotonated and negatively charged under normal pH levels in the environment. However, cationic PFAS have a positively charged functional group and remain positively charged under normal environmental conditions. Furthermore, the charge of zwitterionic PFAS strongly depends on the surrounding pH conditions, as their molecular structure encompasses both negative and positive charges, allowing them to exhibit anionic, cationic, or neutral characteristics (Xiao et al., 2022).

2.1.4 Surfactants

Another notable property of PFAS molecules is that they contain both a hydrophilic and a hydrophobic part. This characteristic makes PFAS surfactants, meaning they are often found in the surface layer between a solid surface and a liquid, or between a fluid and the air (Swedish Chemical agency, 2021). As a result, PFAS molecules have the ability to repel both water and greases, which is desirable for applications such as impregnating textiles and paper. This property is also the reason why PFAS is commonly used in products like non-stick pans. Additionally, PFAS can create a film between two different surfaces, making them useful in products like fire extinguishing foams and cosmetics, where they act as a barrier between two surfaces. However, this property also makes it challenging to determine the relationship between octanol and water (K_{ow}) for PFAS substances, which is an important parameter for describing the solubility of organic compounds (Swedish Chemical agency, 2021).

2.1.5 Transformation and degradation

The carbon-fluorine bond in PFAS is the strongest bond in organic chemistry, with an average bond energy of approximately 104.6 kJ/mol (Xiao et al., 2022). Due to the pres-

ence of multiple carbon-fluorine bonds, PFAS is highly persistent and does not naturally degrade. The only possible decomposition is the formation of more stable variants of PFAS molecules. This persistence contributes to the widespread use of PFAS, as they can withstand high temperatures, oxidation, and low pH values (Swedish Chemical agency, 2021). However, it is also a major environmental concern, as persistent molecules do not degrade in nature and instead accumulate in soil, water, and living organisms (Swedish Chemical agency, 2021). PFAS molecules have half-lives ranging from hundreds to thousands of years, depending on whether they are precursors or not (Raschke et al., 2022). Furthermore, their persistence allows them to be transported over long distances through water flows or by wind dispersion (Swedish Chemical agency, 2021).

2.1.6 Solubility

The solubility of different PFAS molecules can vary greatly due to their diverse molecular structures (Swedish Chemical agency, 2021). However, compared to other organic compounds, PFAS is generally more soluble. Solubility depends mainly on two factors: the length of the hydrophobic carbon chain and the properties of the hydrophilic functional group. In general, longer carbon chains (longer hydrophobic parts) lead to lower water solubility (Swedish Chemical agency, 2021). The ability of PFAS to dissolve in water, combined with their persistence, allows them to be transported over long distances in water environments such as rivers, streams, and groundwater, eventually reaching drinking water sources (Swedish Chemical agency, 2021). However, due to their hydrophilic and hydrophobic properties, a significant amount of PFAS compounds accumulate at the interface between air and water, resulting in lower concentrations in soil pore water (Lyu et al., 2021).

At room temperature, PFAS molecules are typically in a solid state. However, there are several short-chain PFAS molecules that are in a liquid phase at these temperatures and are easily dissolved in water (Xiao et al., 2022). Although long-chain PFAS compounds are solid, they can still dissolve in water to a limited degree (USEPA, 2014). Generally, PFAS with shorter chains are more soluble in water compared to those with longer carbon chains. The solubility of PFAS is still a subject of uncertainty, but according to USEPA, 2014, two of the most common PFAS compounds, PFOS and PFOA, have solubilities of 550-570 mg/L and 9500 mg/L, respectively. Under normal conditions of room temperature and normal pressure, PFAS are thermally and chemically stable. However, there are exceptions where certain compounds may be sensitive to light or heat or can interact with the air (Lyu et al., 2021).

2.1.7 Spreading and exposure

PFAS can enter the environment through primary, secondary, or diffuse sources. Primary sources include businesses that handle and store firefighting foams or use PFAS in the manufacturing process (Swedish Chemical agency, 2021). Secondary sources are businesses that handle products or waste containing PFAS. Diffuse sources arise from the use and wear of PFAS-containing products (Swedish Chemical agency, 2021). In surface and groundwater, the main source of PFAS contamination is firefighting training sites where direct emissions to the environment have occurred due to the use of PFAS-containing

firefighting foams (SGU, 2021). Due to the chemical properties of PFAS, groundwater is an important pathway for their spreading. Local hydrogeology and substance-specific properties play a significant role in controlling the spread of PFAS (SGU, 2021; Swedish Chemical agency, 2021).

The most significant sources of human exposure to PFAS are contaminated seafood, drinking water, inhalation of contaminated indoor air, and contact with contaminated media (Sunderland et al., 2019). The health effects of PFAS exposure in humans are not yet fully understood. However, multiple studies suggest that high levels of PFAS may be associated with various health issues, including an increased risk of kidney or testicular cancer, elevated cholesterol levels, liver damage, and thyroid disease. Additionally, several health risks are linked to pregnancy and the developing fetus, such as delayed mammary gland development, reduced response to vaccines, lower birth weight, and an increased risk of pregnancy-induced preeclampsia or high blood pressure (ATSDR, 2022; European Environment Agency, 2022; Swedish Chemical agency, 2021).

2.2 Transport and fate of PFAS

2.2.1 Advection, Dispersion and Diffusion

Advection is the transport of substances which are entrained in the water flow. Consequently, the substrate is moving due to the fact that the water is moving, according to the hydraulic gradient. However, since the distribution of the water flow velocity is not uniform, advection also tends to spread out or disperse the substrate along the way (Fitts, 2013). The movement with the water is not the only process that affects the contaminant moving, thus using the advection method alone would most often be insufficient. Other processes that may need to be taken into consideration are dispersion, diffusion, and sorption. These processes may also affect the groundwater flow, which in turn affects the contaminant transport (Postigo et al., 2018).

Diffusion is the transport of a contaminant along the chemical gradient, moving from a high gradient to a low, in order to equalize the differences in concentration (Postigo et al., 2018). The diffusion through a cross section can be described by Fick's first law, Equation (1) (Larsson, 2003).

$$q_M = -D_M \frac{dc}{dx} \quad (1)$$

where q_M is the solutes mass flow [$\text{kg m}^{-2}\text{s}^{-1}$],
 D_M is the molecular diffusion coefficient [m^2s^{-1}],
 $\frac{dc}{dx}$ is the concentration gradient [$\text{kg m}^{-3} \text{m}^{-1}$].

If it were possible to exactly define the three-dimensional flow pattern in the aquifer, advection and diffusion would be enough to describe the transport of solute particles in the groundwater. However, since this is not possible, a new transport process, dispersion, has to be introduced to account for the uncertainties in groundwater velocity through the aquifer (Pickens and Grisak, 1981). Dispersivity is incorporated to capture the variations within the aquifer, which can occur at both microscale and macroscale levels. At the microscale, variations may include misaligned pore directions, differences in pore size, and

friction against pore walls. On the macroscale, unknown variations within the aquifer play a prominent role in influencing dispersivity (DHI, 2023a). As a result, dispersion occurs in the x , y , and z directions, leading to increased spreading of PFAS molecules within the aquifer (ITRC, 2020a).

Contaminant spreading to the clean areas of the aquifer, together with an overall decreased concentration of contaminants, are both results of hydrodynamic dispersion (Postigo et al., 2018). Figure 2 shows the phenomena behind dispersion as well as the effect dispersion has on the concentration plume.

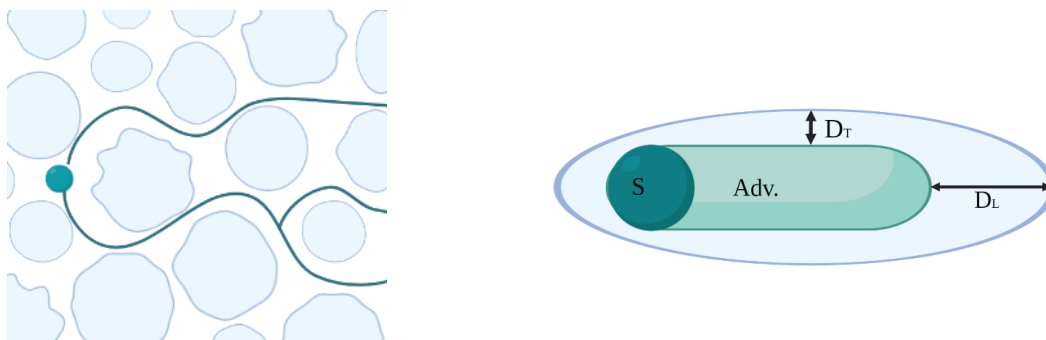


Figure 2: *The left image illustrates the phenomena behind dispersion, due to the different path length possible for a contaminant to travel. The right image shows the effect of dispersion on the concentration plume, where S represents the source, $Adv.$ represents the transport that would occur with only advection, D_L represents the dispersion in the longitudinal direction, and D_T represents the transport in the transversal direction.*

To conclude, the spreading of PFAS is strongly influenced by all the aforementioned processes. However, due to the small diffusion rate compared to the advection rate, diffusion is often neglected in groundwater modeling (ITRC, 2020a). Furthermore, in reality, advection is identified as the primary process driving the transport of contaminants in groundwater. However, when modeling contaminant transport in groundwater, relying solely on advection would be insufficient, as it does not account for the uncertainties in groundwater velocity through the aquifer. To simulate a more realistic spreading within the aquifer, dispersion is incorporated to consider the effects of spatial variability and to ensure a more accurate representation of the transport process. By combining advection and dispersion, the groundwater modeling approach achieves a more comprehensive and realistic representation of a contaminant movement in the aquifer.

Due to the characteristics of PFAS, it is essential to understand the interactions between PFAS and the surrounding media when modeling the transport of these substances. Advection is based only on the properties of the media and does not account for the properties of the transported molecule. Therefore, other processes, such as sorption, needs to be taken into consideration when modeling the transport of PFAS. Sorption is a process that depends on both the characteristics of the transported contaminant and the characteristics of the surrounding media (ITRC, 2020a).

2.2.2 Sorption

Sorption is the process in which a molecule is adsorbed to the solid surface of the aquifer. One example is positively charged ions being adsorbed to the solid surfaces of certain minerals in the aquifer. This can happen due to the oxidation of the solid surface of organic matter, leaving behind negatively charged organic matter, causing the positive ions in the water to be adsorbed to the walls primarily through electrostatic forces, and sometimes even through covalent bonding. This is an important process in contaminant hydrology since it decreases the concentration of contaminants in the water (Postigo et al., 2018).

The distribution or partition coefficient K_d describes the affinity of a pollutant to sorb onto the solids in the aquifer. The expression for this coefficient is described by Equation (2):

$$K_d = \frac{C_s}{C_w} \quad (2)$$

where C_s is the contaminant adsorbed concentration [mg/kg],
 C_w is the contaminant dissolved concentration [mg/L].

A high value of K_d implies that a large amount of the pollutant is adsorbed to the solid surface in the aquifer, while a low value indicates that the pollutant prefers to remain in the water (Postigo et al., 2018).

However, adsorption is a reversible process, meaning that the pollutant can revert into the aqueous phase. For that reason, the sorption/desorption processes trends to slow down the transport of pollutant (Postigo et al., 2018). The retardation factor describes this effect and is expressed as seen in Equation (3) (Postigo et al., 2018):

$$R = \frac{V_x}{V_{\text{cont}}} = 1 + \left(K_d \cdot \frac{\rho}{\theta} \right) \quad (3)$$

where V_{cont} is the advective velocity of the contaminant [m/s],
 V_x is the average linear velocity of the groundwater [m/s],
 K_d is the soil-water distribution coefficient of the contaminant [L/kg],
 ρ is the soil bulk density [kg/m³],
 θ is the volumetric water content [m³/m³].

When examining the equation, it can be seen that a large value of K_d means that the contamination is moving slower than the groundwater, while a small value of K_d indicates transport with the same speed as the groundwater.

Sorption is a key factor when it comes to understanding the transport and fate of PFAS in the soil-water environment. To predict the presence and mobility of PFAS, it is important to understand their soil-water partitioning (Nguyen et al., 2020). The value of K_d varies for different PFAS due to the previously mentioned differences in their structure and charge characteristics. In addition to the properties of PFAS, soil properties and environmental conditions are also important factors that affect sorption (Nguyen et al., 2020). PFAS can be sorbed to the sorbent through several different interactions, such

as electrostatic, hydrophobic, ligand and ion exchange, hydrogen bonding, as well as diffusion. Electrostatic and hydrophobic interactions are the two most common sorption mechanisms for PFAS (Nguyen et al., 2020), and they can be further examined in Figure 3.

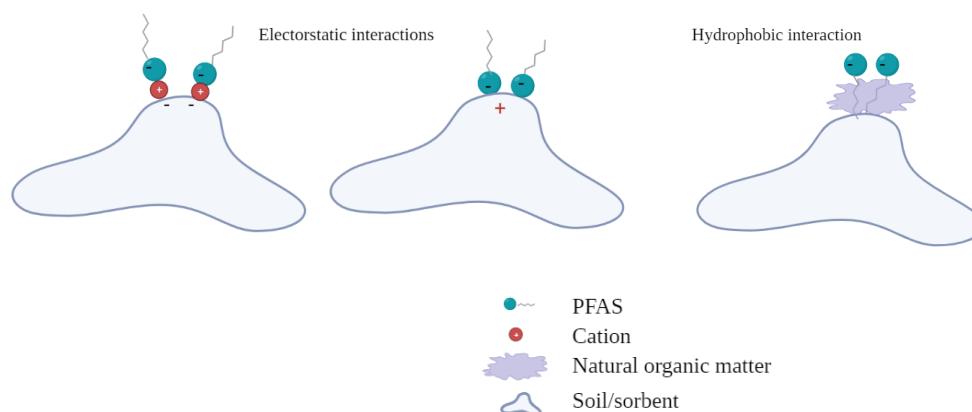


Figure 3: A schematic image illustrating how a negative charged PFAS can be sorbed to the soil or sorbent through either electrostatic interactions or hydrophobic interactions. This Figure is modified from Figure 1 and 2 in Du et al. (2014).

Chemical properties of PFAS

In a study conducted by Nguyen et al. (2020), a relationship was found between K_d and carbon chain length, where an increase in the carbon chain length resulted in an increased value of K_d . This implies that short-chain PFAS are predominantly found in the water phase, while longer-chain PFAS tend to be absorbed into the soil phase. Consequently, short-chain PFAS are more mobile in the environment. Furthermore, Nguyen et al. (2020) observed that PFAS molecules with longer fluorinated carbon chains ($\geq C6$) exhibited more fluctuations in their K_d values when the surrounding properties changed, indicating that they are more influenced by the surrounding soil. Therefore, the fate of long-chain PFAS is dependent on the specific characteristics of the site.

Similar findings were reported by Milinovic et al. (2015), who investigated the sorption of three different PFAS: PFOS, PFOA, and PFBS. The results indicated that PFOS had the greatest affinity for soils, while PFBS exhibited the lowest affinity. This supports the results obtained by Nguyen et al. (2020), as PFBS is the compound with the shortest carbon chain. The hydrophobic interaction between the organic matter in the soil and the hydrophobic fluorinated carbon chain is considered to be the primary sorption process for PFAS, explaining why a longer carbon chain increases sorption.

However, it should be noted that PFOS, despite having a shorter carbon chain than PFOA, exhibits a greater affinity for sorption onto the soil. This indicates that the functional group also plays a significant role in sorption (Milinovic et al., 2015). Additionally, Li, Fang, et al. (2019) found that PFSA adsorb more strongly compared to PFCA, even

when they have the same carbon chain length, further emphasizing the importance of the functional group. Furthermore, the charge of the molecules affects their sorption affinity to the soil, making it an important factor to consider when examining the fate and transport of PFAS (Li, Fang, et al., 2019). Moreover, the effect of the charge also depends on the properties of the surrounding soil. Carboxylic and sulfonic acids are two common functional groups that deprotonate and acquire a negative charge. This deprotonation reaction is favored in basic or alkaline conditions, where there is an abundance of hydroxide ions (OH^-) or other species capable of accepting protons. (Raschke et al., 2022). However, the different functional groups in PFOS and PFOA result in varying anionic strength due to the distribution of the negative charge. Carboxylic acids can distribute the negative charge across two oxygen atoms, while sulfonic acids can distribute it over three oxygen atoms. As a consequence, PFOS is more easily adsorbed to acidic soils, where soil particles tend to have a positive charge, compared to PFOA. This difference can be attributed to PFOS having a slightly higher negative charge due to the presence of the sulfonic acid group. (Raschke et al., 2022).

Soil properties

Soil properties such as the cation exchange capacity (CEC), soil texture, and soil organic carbon (SOC) also have an impact on the sorption of PFAS. Nguyen et al. (2020) found that the sorption of a particular PFAS compound is better described by considering multiple soil properties rather than a single one. Moreover, different PFAS compounds are influenced by different soil properties due to their distinct chemical structures (Nguyen et al., 2020). For instance, natural aquifer and soil surfaces are typically negatively charged, and since anionic PFAS (such as PFAA) also carry a negative charge under normal environmental conditions, the repelling force between them enhances the transport of such PFAS in the subsurface environment (Xiao et al., 2022). However, there are aquifers or types of soil that possess a positively charged surface, leading to increased sorption of anionic PFAS in such environments (Xiao et al., 2022).

Furthermore, several studies have discovered a relationship between PFAS sorption and pH (Campos Pereira et al., 2018; Nguyen et al., 2020), where an increase in pH value leads to decreased sorption. Although PFAS such as PFOS and PFOA are deprotonated in high pH conditions, some soil minerals and organic matter may become less protonated at higher pH levels, reducing their overall positive charge. This change in charge on the solid phase can also decrease the attraction between the negatively charged PFOS/PFOA molecules and the sorbent surfaces, further decreasing sorption. It's important to note that the relationship between pH and sorption can be complex and may vary depending on the specific sorbent material and the properties of the contaminants involved.

The study emphasizes the importance of pH and soil organic matter (SOM) net charge in influencing PFAS sorption behavior. Long-chain PFAS are primarily influenced by pH, while short-chain PFAS are more sensitive to SOM net charge. These differences in sorption behavior indicate varying binding preferences within the soil matrix. Longer PFAS preferentially bind to humin, a stable component of SOM, due to their larger size and complex structures. Shorter PFAS, being smaller and simpler in composition, tend to interact more with other SOM components like fulvic and humic acids (Campos Pereira

et al., 2018).

The sorption of PFAS is reduced in the presence of other organic compounds, such as humic acids, natural organic matter (NOM), and dissolved organic carbon (DOC), due to strong competition for sorption sites between these compounds (Zhang, Liang, et al., 2019). The presence of other surfactants also affects the sorption of PFAS. However, the effect of coexisting surfactants depends on their concentration in the solvent (Du et al., 2014). The sorption of PFAS is decreased at high concentrations of surfactants since the surfactants then form micelles, which increases the solubility of PFAS. On the other hand, a low concentration of surfactants enhances the sorption because the surfactants are able to capture the PFAS from the water when they are already adsorbed onto the sorbent (Du et al., 2014).

Inorganic ions are another important factor affecting the adsorption of PFAS. The effects of coexisting inorganic ions on the adsorption of PFAS are rather complex (Zhang, Liang, et al., 2019). The presence of inorganic ions has several consequences, including competitive adsorption, neutralization of surface area, and bridge formation between PFAS anions and negatively charged groups (Du et al., 2014).

2.2.3 Micelle formation

Amphiphilic molecules, which consist of a hydrophilic and a hydrophobic part, have the ability to form spherical aggregates where their hydrophobic parts are grouped together in the core and the hydrophilic parts are facing outwards to the solvent. This creates a stable colloidal system and the structure is called micelles (Aguilar, 2013). Due to PFAS containing both a hydrophilic and a hydrophobic part, they may form micelles or hemimicelles, which are partial or incomplete micelles, when dissolved in water (Xiao et al., 2022). This occurs when the PFAS molecules gather in clusters, with their hydrophobic fluorinated carbon chain in the center and their functional group facing outwards towards the water, creating a micelle. Hemicelles and admicelles are created when either their carbonic chain or their functional group is attached to a particle surface, as can be seen in Figure 4 (Xiao et al., 2022).

The formation of micelles can both enhance and, in some cases, reduce the sorption of PFAS (Interstate Technology Regulatory Council, 2022). What is unique for PFAS is that they can form micelles even at concentrations lower than the critical micelle concentration (CMC). This ability is possible due to the fluorine atoms (Xiao et al., 2022). PFOA and other long-chain PFAS are able to form hemimicelles at concentrations as low as 0.01 to 0.03 of their CMC (Xu, Liu, et al., 2020).

2.2.4 Precursors

A precursor is a chemical substance that, through a chemical process, can be converted to another persistent substance. There are several PFAS molecules, both perfluorinated and polyfluorinated, that can be converted to PFAA (perfluoroalkyl acids) in the environment, including biota. These PFAS are therefore considered precursors to PFAAs. In certain contaminated sites, the majority of the PFAS compounds found may consist of precursors

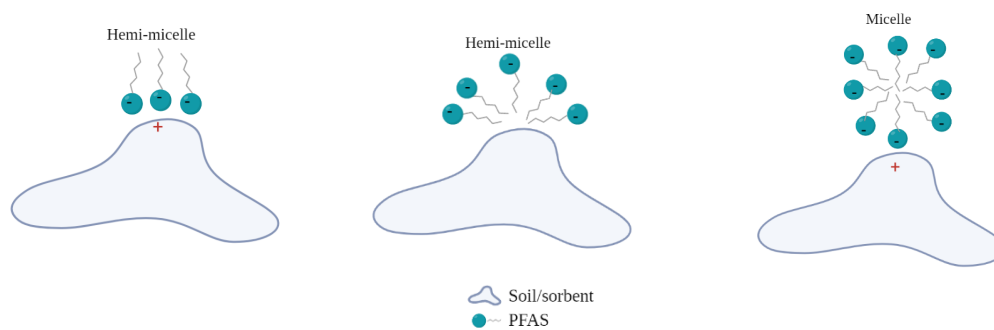


Figure 4: *Conceptual image of how the PFAS molecules can sorb to the soil or sorbent due to the formation of different types of micelles. Figure is modified from Figure 5.2 in Interstate Technology Regulatory Council (2022).*

rather than primary PFAAs (Swedish Chemical agency, 2021). This can complicate the assessment of the contaminant site, as the precursors may degrade to PFAAs later on. Moreover, the presence of precursors could positively impact the longevity of the PFAS plume in groundwater, leading to a spread of PFAS over an extended time period (Gefell et al., 2022). Additionally, there are several PFAA precursors that can be transported by wind in the atmosphere, ending up in locations far away from the original source. Once deposited, they can be converted to PFAAs, resulting in the presence of PFAS even in locations distant from the original source. Considering these facts, it becomes essential to account for precursors when examining the fate and transport of PFAS at a contaminated site (Swedish Chemical agency, 2021).

2.3 Modeling Strategies

2.3.1 FEFLOW

FEFLOW is an acronym for Finite Element subsurface FLOW and transport system. It is an advanced 3D Finite-Element groundwater modeling system that can be used for simulating heat transfer, mass transfer, and groundwater flow in both fractured and porous media (Trefry and Muffels, 2007).

Flow modeling principles

Groundwater models based on physical processes follow two fundamental flow modeling principles. The first principle is the conservation of mass, which states that water cannot be created or destroyed, and any changes in water storage are a result of water entering or leaving the system. The second principle is Darcy's law, which assumes that flow occurs from areas of high hydraulic potential to areas of low hydraulic potential. (Andersson et al., 2015).

The conservation of mass can be described by Equation (4), which represents the balance of water within the system:

$$\Delta\text{Storage} = Q_{\text{in}} - Q_{\text{out}} \quad (4)$$

Groundwater recharge, which refers to the portion of precipitation that infiltrates and reaches the groundwater, plays a significant role in water input to the system (Hartmann, 2022). Recharge is therefore an important parameter in groundwater modeling. Furthermore, porosity is another crucial parameter that represents the amount of water that soil or rock can hold, and therefore affects the total amount of water within the system (Earle, 2015).

The second principle, Darcy's law, further describes how groundwater flow is influenced by three key factors: hydraulic conductivity, hydraulic gradient, and cross-sectional area, as shown in Equation (5) (Ge, 2003).

$$Q = -KA \frac{dh}{dl} \left[\frac{m^3}{s} \right] \quad (5)$$

where K is the hydraulic conductivity [m/s],
 A is the cross-sectional area perpendicular to the flow direction [m²],
 $\frac{dh}{dl}$ represents the hydraulic gradient [m/m].

The primary driving force for groundwater flow is the hydraulic gradient, and the negative sign in the Equation indicates that flow occurs from areas of high potential to areas of low potential (Ge, 2003). Hydraulic conductivity characterizes the speed with which water can move through the pore space of an aquifer. A higher hydraulic conductivity value indicates easier flow. The hydraulic conductivity value varies depending on soil properties and can differ significantly between different materials. However, even within the same material, the hydraulic conductivity can exhibit significant variation, making it a complex parameter to consider (Lu, 2015).

Transport and Dispersivity

As already mentioned, advection and dispersion is two of the most important processes to use when it comes to modeling the transport of a contaminant in the groundwater. Where the concept of dispersivity is introduced in groundwater modeling in order to account for the inherent spatial variations within the aquifer. The transport Equation describing the advective flux can found below, see Equation (6) (Phillips and Castro, 2014).

$$J_{adv} = C \bar{K} \nabla h = C \bar{q} \quad (6)$$

where J_{adv} is the advective flux [mol/m²s],
 C the volume concentration [mol/m³],
 \bar{K} the hydraulic conductivity tensor [m/s],
 ∇h the hydraulic gradient [m/m],
 \bar{q} the specific discharge [m³/m²s].

To describe the dispersive flux, the linear Fickian relationship is often used. Equation (7) states Fick's law, which establishes a linear relationship between the concentration gradient ∇C and the dispersive flux J (Hassanizadeh, 1996).

$$J_{dis} = -D \cdot \nabla C \quad (7)$$

where J_{dis} is the dispersive flux [mol/m²s],
 D is the dispersion tensor [m²/s],
 ∇C is the concentration gradient [mol /m³m].

By combining advection and dispersion together the advection-dispersion Equation could be obtained, see Equation (8).

$$\frac{\partial C}{\partial t} = D_x \frac{\partial^2 C}{\partial x^2} + D_y \frac{\partial^2 C}{\partial y^2} + D_z \frac{\partial^2 C}{\partial z^2} - v \frac{\partial C}{\partial x} \quad (8)$$

where D_x, D_y, D_z are hydrodynamic dispersion coefficients [m²/s],
 C is the concentration [mol /m³],
 t is the time [s],
 v the advective transport in the x direction [m/s].

In FEFLOW, the default approach for describing macroscale dispersion is through the linear Fickian relationship. This relationship divides dispersion into longitudinal dispersion (along the flow direction) and transverse dispersion (perpendicular to the flow direction) (DHI, 2023a). Generally, longitudinal dispersion is more significant, while transverse dispersion is relatively less important (Jonasson et al., 2007).

The determination of dispersivity values is a complex task, and the underlying theory is not well understood. However, Schulze-Makuch (2005) discovered a relationship between the scale of the plume and the longitudinal dispersivity, which can be utilized to estimate dispersivity for a given aquifer. This suggests that dispersivity depends on the scale of the flow distance. In other words, if the flow distance is long, greater dispersion is expected. Equation (9) represents the relationship identified by Schulze-Makuch (2005). Compared to the dispersion tensor, D , α only describes the dispersion in one direction.

$$\alpha = cL^m \quad (9)$$

where α is the longitudinal dispersivity [m],
 c is a parameter characteristic of the geological medium [m],
 L is the flow distance [m],
 m is the scaling exponent [-].

The scaling exponent was found to range between 0.4 and 0.94, with a mean value of 0.5. Additionally, the parameter c is dependent on the geological media. For example, sandstone and unconsolidated media had a value of approximately 0.01 m, while carbonate rocks had a value of 0.8 m. This is a reasonable finding since carbonate rocks can be extremely heterogeneous on small scales (Schulze-Makuch, 2005). As demonstrated in Equation (9), the longitudinal dispersivity increases with a greater flow distance, indicating that it grows with an increasing measurement scale (Schulze-Makuch, 2005). Dispersivity is not a constant value; instead, it increases with the length of flow transport due to the heterogeneity of the aquifer. Dispersivity can vary between 10⁻² and 10⁴ m (Jonasson et al., 2007).

Additionally, another expression for longitudinal dispersivity was proposed by Xu and Eckstein, 1995. This expression also describes the dependency of longitudinal dispersivity on the plume length and is shown in Equation (10).

$$\alpha_L = 0.83(\log L_s)^{2.414} \quad (10)$$

where α_L is the longitudinal dispersivity [m],
 L_s is the plume length [m].

Sorption

As previously mentioned, sorption is a crucial process to consider when modeling the transport of PFAS, given the specific characteristics of PFAS compounds. Common methods used to quantify the sorption of PFAS in nature include the Freundlich isotherm, the linear isotherm, and the Langmuir isotherm (Sima and Jaffé, 2021).

The Freundlich isotherm is an empirical relationship that is ideal for expressing non-ideal sorption. It can be expressed as (Sparks, 2003):

$$C_s = K_f \cdot C^n \quad (11)$$

where C_s is the concentration in the solid phase [mol/kg],
 C is the concentration in the fluid phase [mol/m³],
 K_f is the Freundlich adsorption constant [m³/kg],
 n is the Freundlich exponent [-].

The Freundlich isotherm describes the nonlinear sorption of a substance and is suitable for describing PFAS sorption. This is because the sorption of PFAS is typically nonlinear, with the adsorption constant decreasing as the concentration of PFAS increases. The value of K_f is greatly influenced by both the specific PFAS compounds and the properties in the surrounding soil (Sima and Jaffé, 2021).

By assuming the Freundlich exponent (n) to be equal to 1, we obtain a specific case of the Freundlich isotherm known as the linear isotherm. The linear isotherm sorption can be described by Equation (12):

$$C_s = K_d \cdot C \quad (12)$$

where C_s is the concentration in the solid phase [mol/kg],
 C is the concentration in the fluid phase [mol/m³],
 K_d is the equilibrium distribution coefficient [m³/kg].

The value of K_d is influenced by various factors, including the chemical properties of the specific PFAS, soil characteristics, and environmental factors such as pH (Sima and Jaffé, 2021).

The Langmuir isotherm is another nonlinear model frequently used to describe the sorption of PFAS. Unlike the previous isotherms, the Langmuir isotherm incorporates the

concept of sorption capacity. The sorption capacity can vary significantly depending on the soil characteristics and the interactions between PFAS and the soil (Sima and Jaffé, 2021).

The Langmuir isotherm is based on the idea that sorption can only occur at specific homogeneous sites within the sorbent. Once a site is occupied by a compound, no other compound can sorb to that same site. Additionally, there are no interactions between neighboring adsorbed compounds (Ho et al., 2002). The driving force for sorption in this model is the concentration of compounds in the solution (Ho et al., 2002).

The Langmuir isotherm is expressed as follows (Sima and Jaffé, 2021):

$$C_s = \frac{b_L S_m C}{1 + b_L C} \quad (13)$$

where b_L is a constant [-],
 C_s is the concentration in the solid phase [mol/kg],
 C is the concentration in the fluid phase [mol/m³],
 S_m is the maximum sorption capacity [μ g/g].

Due to the limited number of sorption sites in the Langmuir isotherm model, the number of adsorbed molecules at equilibrium is equal to the number of molecules leaving the adsorption sites (Sahu and Singh, 2019). In FEFLOW, the product of b_L and S_m is expressed as k_1 , and b_L is represented by k_2 (DHI, 2023a). The sorption capacity, S_m , is influenced by solution conditions such as pH and temperature, as well as the characteristics of both the PFAS and the adsorbent (Sima and Jaffé, 2021).

One advantage of using the Langmuir isotherm over the Freundlich isotherm is its ability to account for the sorption capacity when multiple PFAS coexist. In such cases, S_m represents the total sorption capacity for all the present PFAS (Sima and Jaffé, 2021). However, it is not possible to conclusively determine whether the Langmuir or Freundlich isotherm is more accurate for simulating PFAS adsorption (Zhang, Liang, et al., 2019).

However, in FEFLOW, a different sorption isotherm is used, called the linear Henry sorption isotherm. Henry's sorption isotherm has a similar expression as Equation 12 except that the constant K_d is replaced with the dimensionless Henry's constant K_H . To obtain the Henry's constant for a substance, the K_d value is multiplied by the density of the solid (DHI, 2023a). Equation (14) shows the expression for the linear Henry sorption isotherm.

$$C_s = K_H \cdot C \quad (14)$$

where C_s is the concentration in the solid phase [mol/m³],
 C is the concentration in the fluid phase [mol/m³],
 K_H is Henry's constant [-].

Oscillation Damping

One common issue in transport processes described by the advection-dispersion Equation (Equation (8)) is the occurrence of negative concentrations in the solution, which arises due to false numerical oscillations in the discretization scheme (Wendland and Schmid, 2000). This problem can occur when the time steps are too long or when the mesh is too coarse. The sharp fronts of concentration plumes, which are not smooth and continuous, contribute to this issue.

To address this problem, FEFLOW offers several methods that can help smoothen out the concentration gradient. These methods include mesh refinement, upwinding techniques, and increasing the dispersivity values (DHI, 2023a). By applying these methods, the numerical oscillations can be mitigated, resulting in more accurate and physically realistic solutions for concentration distribution in groundwater flow simulations. These approaches aim to mitigate or eliminate the false numerical oscillations and prevent the presence of negative concentrations in the solution.

A simple solution to the problem is to refine the mesh, which leads to fewer numerical oscillations. However, a fine mesh implies a more complex model, and therefore this option may not be applicable for all models (DHI, 2023a).

A way to investigate the stability of the advection-dispersion transport process is to examine the Péclet number. The Péclet number describes the relationship between diffusive and convective mass transport and is defined as (DHI, 2023a; Rapp, 2017b):

$$Pe = \frac{v \cdot L_{char}}{D} = \frac{\text{convection transport}}{\text{diffusion transport}} \quad (15)$$

where D include both the mechanical dispersion and the molecular diffusion [m^2/s],
 L_{char} is the characteristic length of the system [m],
 v is the velocity of the flow field [m/s].

A high Péclet number indicates that mass transport is dominated by convection, while a low number indicates the dominance of diffusion transport (Rapp, 2017a). If L_{char} is set as the characteristic length of the element size in the direction of flow, the grid Péclet number can be obtained. A rough rule of thumb often used is that the value of the grid Péclet number should not be higher than 2 (DHI, 2023a).

Another method used to smoothen the concentration gradient is to use upwinding options, which not only reduce the risk of oscillation but also improve numerical stability. Upwinding is a technique used to solve hyperbolic partial differential Equations. It involves estimating derivatives in a flow field by using a set of data points biased to be more "upwind" of the query point with respect to the direction of the flow. This approach ensures that the derivatives are calculated based on information flowing towards the point of interest, enhancing the accuracy of the solution (Wikipedia, 2023b). However, the use of upwinding techniques introduce additional numerical (artificial) dispersivity to the transport model (DHI, 2023a).

FEFLOW offers various upwinding options, including:

1. No upwinding: This is the default setting in FEFLOW, where no upwinding is applied.
2. Streamline upwinding: This technique smoothes the concentration gradient primarily in the flow direction, helping to mitigate oscillations.
3. Full upwinding: Full upwinding smoothes the concentration gradient in all directions.
4. Shock capturing: This technique selectively applies upwinding only where necessary, targeting areas with sharp gradients.
5. Least-squares upwinding: This method uses a least-squares approach to determine the optimal amount of upwinding in each direction.

It's important to note that when using upwinding techniques, artificial dispersivity is added to the model. This can potentially result in a smoother representation of the contaminant plume compared to the actual physical behavior. Therefore, the selection and application of upwinding techniques should be done carefully in order to find the balance between capturing the true behaviour of the plume but at the same time to avoid diffusion or over-smoothing which may lead to incorrect results. (DHI, 2023a).

A third way to solve the problem is simply by increasing the dispersivity values used in the model (DHI, 2023a).

Transport Equations

In FEFLOW, it is possible to choose between two different formulations of the transport Equation. The chosen transport Equation has a significant impact on the interpretation of transport boundary conditions for well types and on the flux. The default formulation used in FEFLOW is the convective form, where dispersion is the driving force for the assigned maximum flux. In an attempt to achieve the assigned mass flux, FEFLOW will adjust the concentration along boundary sections, resulting in a lower concentration at these sections. Consequently, additional mass enters the model through advective transport, leading to a discrepancy between the assigned boundary condition and the obtained mass flow budget (DHI, 2023c).

The second option is the divergence form, where both advective and dispersive flux are represented in the assigned mass flux. By using this transport Equation, a lower and more realistic concentration is obtained at the boundary sections. This results in a lower mass influx, which in turn ensures that the mass flow budget and the assigned boundary conditions are identical (DHI, 2023c).

State

In FEFLOW, simulations can be conducted in two different states: steady state or transient. In a steady state simulation, the obtained solution represents the system's state when it has been exposed to constant material properties and boundary conditions for an infinitely long time (DHI, 2023b). On the other hand, in a transient simulation, the simulation starts from an initial condition and covers a specified time period, rather than an infinite time duration. Both transient and steady state simulations can be used for flow and transport simulations, either by having the same state for both flow and transport or by having a steady state flow with transient transport (DHI, 2023b).

2.3.2 Analysis of modeling results

One approach to gaining insight into the importance of each parameter in groundwater modeling is by using a machine learning algorithm known as random forest regression. This algorithm can be trained using the data obtained from the simulations, which includes various input parameters and the corresponding resulting concentration. If the algorithm achieves accurate predictions, it becomes valuable to analyze the factors it relies on for making those predictions. This analysis can provide a deeper understanding of the relative importance of each parameter in influencing the model's outcomes. All this can be done in Python and the theory behind this will be presented in more detail in the following sections.

Random Forest

Random Forest is an algorithm mainly used for regressions or classification problems (Sruthi, 2023). A simple schematic overview over how the algorithm works can be seen in Figure 5. In the Figure, three decision trees are presented. Each decision tree consists of a root node, internal nodes (also known as decision nodes), and leaf nodes. The decision nodes make choices about whether to go left or right based on available features. For each decision tree, a subset of data points (in this case, concentrations of PFAS in each node) and a subset of features (the input parameter values that generate the given concentration) are selected (Sruthi, 2023).

Each decision tree in the random forest regression algorithm starts with a root node and forms a tree-like structure through multiple feature-based splits, ultimately reaching the leaves. These leaves represent the possible outcomes resulting from a series of decisions. It is important to note that in Figure 5 each decision tree takes different routes down the tree structure, as each tree is unique and makes decisions based on different features (Saini, 2023).

The random forest regression algorithm builds up its decision trees using the training data provided. The training data is divided into multiple subsets, and for each subset, a decision tree is created. At each node of the tree, the algorithm selects the best feature to split the data based on a specific criterion, such as minimizing the mean square error. Once the algorithm reaches a leaf node in the decision tree, it has made a predic-

tion for the target variable, which, in this case, is the concentration of PFAS. (Saini, 2023).

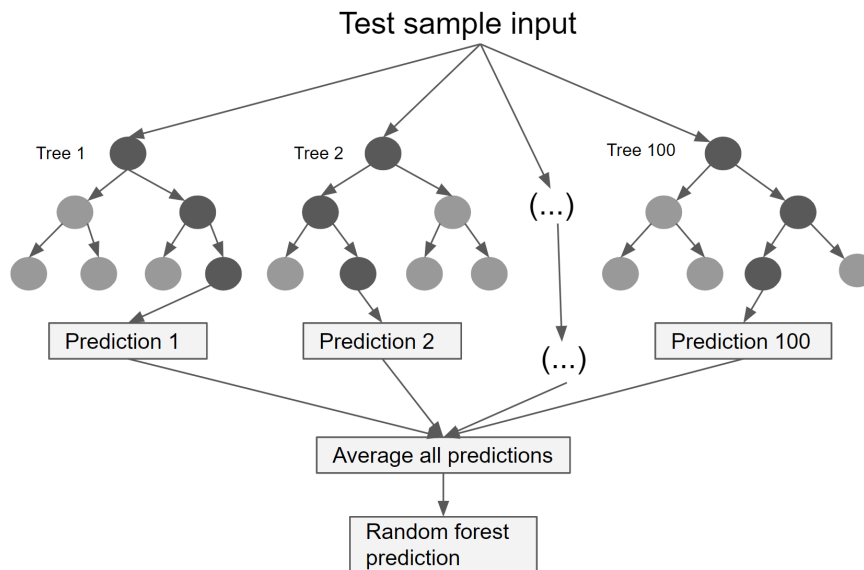


Figure 5: A schematic overview of the basis behind the random forest regression model is shown. The dark grey nodes indicate the path down each decision tree. At each internal node, a decision is made whether to go left or right down the tree. Finally, one of the leaf nodes is reached, which provides a prediction of the resulting concentration. The random forest regression algorithm derives its final prediction from the average value obtained from all the created decision trees.

The random forest algorithm works by creating multiple decision trees based on the provided dataset. Once all the decision trees are created, the Random Forest algorithm examines the predicted values from each decision tree in order to do the final prediction. If the algorithm is used for classification the value of the majority is used, however, if the algorithm is used for regression the average values of all predictions is used (Sruthi, 2023). In this case, when dealing with concentrations, the regression model is employed to predict the resulting concentration.

A way to validate the accuracy of the created Random Forest model is to examine the mean squared error (mse), the root mean squared error (rmse) and the coefficient of determination (R^2) values. The mse value describes the average square difference between the value predicted by the model and the actual value obtained from the simulations in FEFLOW (Wikipedia, 2022). The rmse measures the averages difference between the actual values and the predicted ones. A lower rmse indicates a better model since the predicted value then are very close to the actual ones (Wikipedia, 2023a). Lastly, the R^2 value is a parameter used for regression models in order to describe to which extent the variance of one variable can explain the variance of another, second variable. In other words, a R^2 value of 0.5 means that the model's input can describe about half of the observed variations (Fernando, 2023). The R^2 value can range between 0 and 1, with values between 0.85 and 1 indicating that the model performs relative in line with the actual values.

Shapley values

In machine learning, Shapley values can be used to analyse the importance of the different inserted features or how much the input features will impact the model's predicted output (Bagheri, 2022). Shapley values originate from game theory which models complex human behaviours mathematically, with the aim of being able to understand and predict it (Choudhary, 2019). More specifically, Shapley values originates from cooperative game theory. A cooperative game is a game where the competition is between groups of players rather than individuals.

The game in this case is to predict the output value (concentration of PFAS in a certain node) and the players are the different input parameters (such as hydraulic conductivity, sorption, porosity, etc.) (Choudhary, 2019). In machine learning, the predicted values often differ from the expected ones. The contribution that each input parameter has on creating this difference is presented by the Shapley values, where a high value indicates a greater importance. In other word, the Shapley values for each output concentration aim to determine the appropriate weights for the model parameters so that when all the Shapley values are combined, they add up to the difference between the expected (or average) value and the predicted value (Choudhary, 2019). In this way it is possible to get an indication of which parameter that has the greatest impact on the output concentration. Which could be of interest since it can tell which parameters that are important to use/have a correct value of when simulating the transport of PFAS in the groundwater.

Random Forest Importance

The random forest regression algorithm use a three-base strategy which naturally rank how well each parameter improve the accuracy of the obtained result (Gupta, 2020). Parameters that cause a great decrease in the accuracy of the results will end up in the start of the trees. Furthermore, parameters that have a less impact on the accuracy of the result will end up at the end of the tree. Due to this structure, it is easy to obtain the most important parameters according to the created Random Forest model (Gupta, 2020).

3 Methods for simulating transport of PFAS in the literature

3.1 Method

A comprehensive literature review was conducted to explore previously employed methods for simulating the transport of PFAS in groundwater. The main objective of the review was to identify documented modeling approaches. Additionally, the study aimed to assess whether the parameter values used in these models were supported by scientific literature or based on assumptions.

The literature review encompassed scientific articles, reports, and Master's Theses. The search was conducted using databases such as LUBSearch and Google Scholar, employing keywords such as PFAS, Modeling, Transport, MODFLOW, FEFLOW, PFOS, and PFOA.

In addition to the literature review, experts in the field were consulted via email to gain insights into their preferred modeling methodologies. The following research questions guided the literature review:

- What modeling tools and methods were employed in previous studies?
- Which transport processes were considered in these models?
- What parameter values were utilized for PFAS-specific properties?
- Were the chosen parameter values based on scientific findings?

3.2 Results

The overall result from the literature review was limited; not many articles and reports were found where the modeling was described in depth. Instead, most of them simply mentioned that modeling was conducted without providing specific details about the chosen model or parameter values. However, some information was still found, and several different modeling approaches were discovered. Brusseau (2020) used a Multiprocess rate-limited mass transfer (MPMT) model, while Silva, Guelfo, et al. (2022) and Silva, Šimůnek, et al. (2020) utilized a modified version of the HYDRUS model, which, among other things, included the air-water interface (AWI) adsorption. Furthermore, Wallis et al. (2022) employed an extended version of the LEACHM (Leaching estimation and chemistry model) numerical model, which incorporated the sorption of PFAS compounds to the AWI. However, these methods were all used for groundwater modeling in the unsaturated zone, which was not the focus of this study since the emphasis was on the saturated zone.

Regarding the saturated zone, the most commonly used modeling tool was MODFLOW. In an email received from Prommer (2023), he described using MODFLOW/MT3DMS tools to simulate PFAS fate in saturated groundwater zones. He also mentioned that they only considered equilibrium sorption as the reactive process, neglecting other reactive

processes for simplicity.

Similarly, Pettersson (2020) used MODFLOW/MT3DMS to model the transport of PFOS, PFPeA, and PFBA. Sorption was included in the modeling, with the following K_d values: 15 L/kg for PFOS, 0.49 L/kg for PFPeA, and 0.01 L/kg for PFBA. The K_d values for PFOS and PFPeA were obtained from literature, while the value for PFBA was assumed.

Moreover, MODFLOW was also used by Persson and Andersson (2016) for the simulation of PFOS transport. Due to the lack of data regarding PFOS chemical properties, they used the default setting for very parameter value in MODFLOW, except for sorption where they adjusted the K_d value, which was set to 17 L/kg based on values found in the literature.

Boonraksasat (2019) also employed MODFLOW for the simulation of PFAS transport, specifically investigating PFOS, PFOA, 6:2 FTS, and PFPeA. Similar to other studies, the only reactive process considered in their modeling was sorption, and the K_d values used were based on literature findings: PFOS - 15 L/kg, PFOA - 1.1 L/kg, 6:2 FTS - 0.65 L/kg, and PFPeA - 0.49 L/kg.

Edvinsson (2015) utilized MODFLOW with the MT3D99 tool to simulate the transport of PFOS in groundwater. They conducted simulations with and without sorption, setting the K_d value to 0.44 L/kg for sorption based on a value found in the literature.

Gefell et al. (2022) conducted a more comprehensive simulation, examining both PFOA and its precursors. They also used MODFLOW/MT3D with linear sorption and desorption. The K_d values were calculated by multiplying the organic-carbon-based partition coefficient (K_{oc}) with the fraction of organic carbon in the soil (f_{oc}), both obtained from the literature. For PFOA, the K_d values ranged from 0.0108 to 0.108 L/kg, while for the precursors, the range was 1.349 to 13.49 L/kg. Unlike previous MODFLOW modelings, Gefell et al. (2022) assigned the K_d values on a cell-by-cell basis.

Overall, the literature review did not yield many well-described modeling approaches. It seems, from the lack of documentation and the extensive use of simplifying assumptions, that the field of modeling PFAS transport is still relatively unexplored. Furthermore, sorption is commonly included as the only reactive process, with the sorption coefficient often based on literature findings.

Because the literature review provided limited findings, the initial modeling approach, which aimed to test various methods for simulating PFAS transport, could not be executed effectively. As a result, an alternative modeling approach was adopted, and its details will be further discussed in the subsequent section of this report.

4 Methodology

This section presents the methodology used for modeling and sensitivity analysis.

4.1 Modeling

As mentioned before, the literature review did not yield satisfactory results, necessitating a different modeling approach. Since the literature review did not provide detailed descriptions of how the transport of PFAS has been previously modeled, it was not possible to use the approach of comparing different methods and evaluating their performance. Consequently, the focus shifted to identifying the parameters that have the greatest impact on the resulting concentrations. Additionally, the impact of the modeling state on the results was investigated. Two different modeling strategies were employed: steady state modeling and transient modeling. The methodology used was the same for both strategies, with differences only in the FEFLOW model setup and the number of simulations conducted. Four different scenarios were modeled using the steady state approach: PFOS in a sandy aquifer, PFOA in a sandy aquifer, PFOS in a sand-gravel aquifer, and PFOA in a sand and gravel aquifer. Due to time constraints, only the scenario of PFOS in a sandy aquifer was modeled in the transient state.

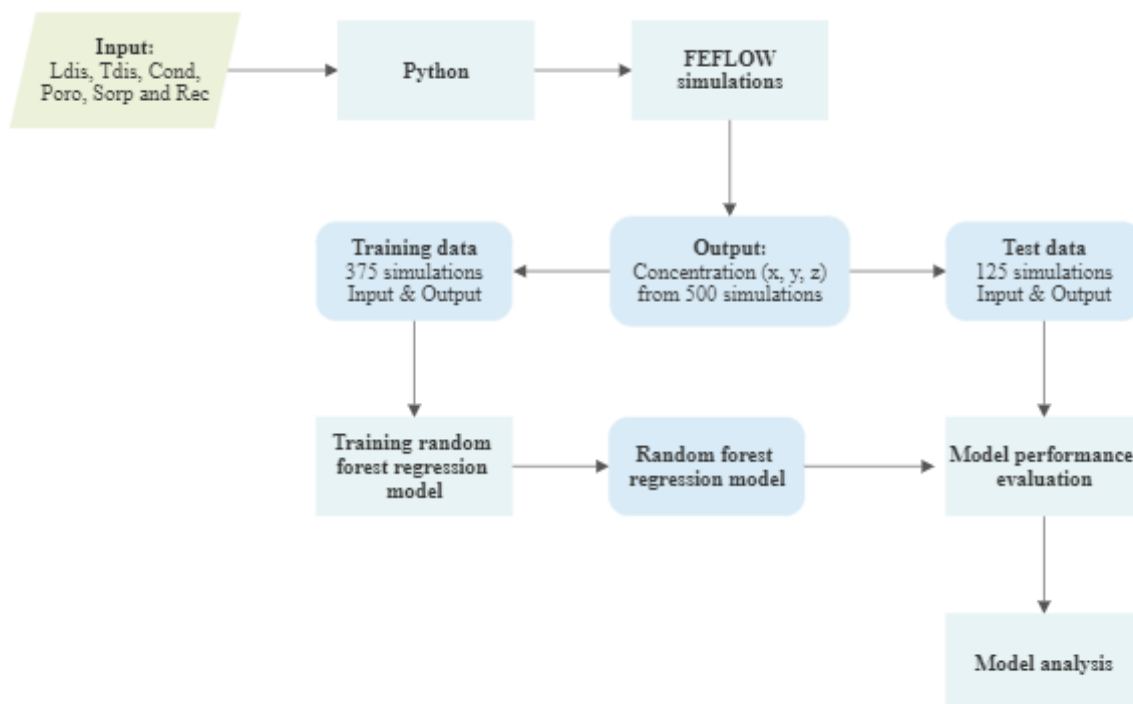


Figure 6: *Flowchart describing the methodology used for the modeling.*

Figure 6 provides an overview of the methodology employed. The method begins by defining the intervals for the six input parameters. The final intervals used for each of

the four scenarios can be found in Table 10. Based on these intervals, an input file is created and fed into Python. Python then conducts the simulations by inputting parameter values into FEFLOW, where the simulations are executed. A total of 500 simulations are performed for each of the four scenarios in steady state, and 200 simulations for the transient modeling. The output consists of the concentrations of PFAS obtained for each node after a single simulation. After conducting 500 simulations, a total of 500 different output files were generated. The output files were divided into training and test datasets. The training data is used to build a random forest regression model, which can predict the output concentration in a given node based on specific input parameter values. The random forest regression model is evaluated by testing it on the test data and comparing its predictions with the output obtained from the simulations. Finally, a model analysis is conducted on the random forest regression model. Each of these steps will be described in more detail in the following sections.

4.1.1 FEFLOW model

A 3D model of the examined area was constructed in FEFLOW. The model included a source of PFAS and three observation wells, where PFAS concentrations were measured. The model was divided into two layers: an upper layer and a lower layer. Additionally, each layer was replicated to accommodate changes made to different parameters during the simulation, resulting in separate upper and lower layers for each parameter. This was done in order to be able to change several parameters at once during the simulation. The model and its two layers can be seen in Figure 7. In the final analysis, for simplicity, the two layers were not distinguished from each other; instead, they were assumed to consist of exactly the same material.

Additionally, the visible edge in Figure 7 represents the shoreline and is treated as a hydraulic constant head boundary condition in the model. Both steady-state and transient simulations were conducted using the same model. The transport Equation was formulated in the divergence form, and oscillation damping was implemented through full upwinding to stabilize the model. Based on the findings from the literature review, where it was observed that in all cases sorption was the only reactive process considered, the same assumption was adopted in this study as well.

Isotropic conditions were applied to simplify the model. However, there were differences between the two modeling strategies. In the steady-state model, a PFAS source concentration of 1 mg/L was utilized, while in the transient model, a higher concentration of 10 mg/L was used. The higher concentration in the transient model was chosen to observe transport phenomena within a reasonable timeframe without the need for an excessively long simulation period. The total run time of the transient model was set to 18,250 days, corresponding to 50 years, with an initial simulation time of 0 days and an initial time step length of 100 days.

Several assumptions were made when the model was created. Firstly, the aquifer was assumed to be isotropic, meaning that the hydraulic conductivity was considered to be the same in all three directions in both of the layers. Secondly, the aquifer was assumed

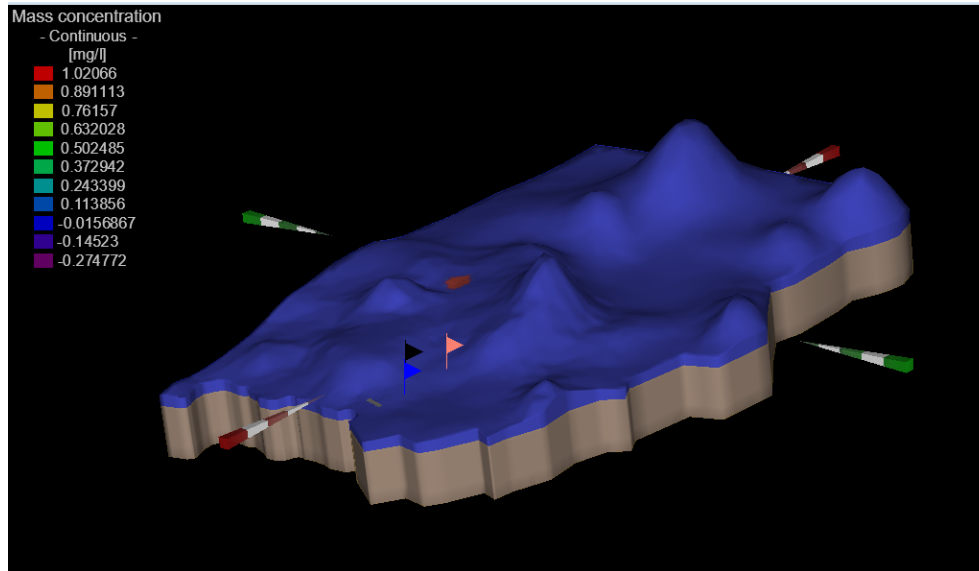


Figure 7: The 3D model used in FEFLOW is divided into two layers: an upper layer (blue) and an underlying layer (grey). However, for the simulations, these layers were treated as one homogeneous layer. The red area represents the source of contamination. Three observation points are marked by flags: the blue flag represents observation point 1, the red flag indicates observation point 2, and the black flag signifies observation point 3. The edge of the model represents the shoreline.

to be homogeneous, implying that the material properties, such as porosity and hydraulic conductivity, were assumed to be uniform and consistent throughout the two layers in the entire aquifer. These are commonly used assumptions in groundwater modeling; however, they are a significant simplification of the reality where aquifers in general are anisotropic and heterogeneous.

4.1.2 Parameters

In the simulations made, eight parameters were alternated: longitudinal and transversal dispersivity, hydraulic conductivity in all three directions, Henry's sorption coefficient, porosity, and groundwater recharge. All the parameters used in the simulation can be found in the Tables presented below. The data for which the parameters are based on come from the literature study conducted and can all be obtained in the Appendix A, B and C.

Dispersivity

The values for the longitudinal dispersivity can be found in Table 3. The "eq" label indicates that these values were calculated using the obtained scale, which represents the length of the plume, in conjunction with Equation (10). This Equation describes the relationship between the plume length and the longitudinal dispersivity. The values in Table 3 represent the minimum, maximum, and average values of the collected and calculated values, which can all be found in Appendix B. The minimum and maximum values were used in the modeling to create an interval from which the numbers were randomized.

Values for the transversal dispersivity can be found in both Table 4 and Table 5. Table

Table 3: *Minimum, maximum and average values found for the longitudinal dispersivity in two different types of soil. eq indicates that the values have been calculated by using Equation (10) together with the values of the scale, belonging to the obtained values of α_L . The values of α_L , together with the belonging scale, were obtained from Schulze-Makuch, 2005.*

Longitudinal Dispersivity (m)						
Soil type	α_L (avg.)	α_L (avg. (eq))	α_L (min)	α_L (min (eq))	α_L (max)	α_L (max (eq))
Sand	0.187	2.30	0.047	0.007	0.550	7.04
Sand and Gravel	5.57	2.83	0.600	0.732	350	7.91

4 displays the minimum, maximum, and average values of the transversal dispersivity for two different types of soil. The values on which this is based can be found in Appendix C. Table 5 shows the ratio between transversal and longitudinal dispersivity, also based on the collected values found in Appendix C. Typically, the transversal dispersivity is never larger than the longitudinal dispersivity. Therefore, to prevent this from happening in the simulations, the transversal dispersivity was set to 1/10 of the longitudinal dispersivity, as this is a commonly used fraction.

Table 4: *Minimum, maximum and average values found for the transversal dispersivity in two different types of soil. The data which this is based on can be found in Appendix C.*

Transversal dispersivity, α_T (m)			
Soil type	Min	Max	Average
Sand	0.005	0.760	0.158
Sand and Gravel	0.020	27.4	6.90

Table 5: *Minimum, maximum and average values found for the relationship between longitudinal and transversal dispersivity (α_T / α_L) in two different types of soil. The data which this is based on can be found in Appendix C.*

Transversal dispersivity (α_T / α_L)			
Soil type	Min	Max	Average
Sand	0.040	0.500	0.234
Sand and Gravel	0.019	0.600	0.286

Porosity

Table 6 shows the dry density, along with the maximum and minimum porosity values for two different soils. The porosity values were used as input in FEFLOW and were different for each simulation, while the density values were used for calculating Henry's sorption coefficient from K_d values (see Appendix B) and remained constant throughout the simulations.

Table 6: *Dry density, together with minimum and maximum values for porosity for two different types of soil. Values are obtained from: ¹ Yu, Loureiro, et al., 1993, ² Yu, Cheng, et al., 2023 and ³ Zhang, Xiao, et al., 2017.*

Soil Density and Porosity			
Soil type	Dry density g/cm ³	Porosity (min)	Porosity (max)
Sand	1.54 ¹	0.250 ²	0.500 ²
Sand and Gravel	2.1 ³	0.250 ²	0.500 ²

Sorption

Values for Henry's sorption coefficient for two different PFAS can be found in Table 7. The values represent the calculated minimum, maximum, average, and median based on collected data, which can be further examined in Appendix A Tables A and B. Additionally, the 10th and 90th percentile values were calculated. These values defined the upper and lower bounds of the interval used to randomize the sorption coefficient for each simulation.

Table 7: *Maximum, minimum and mean values of Henry's sorption coefficient for two different PFASs: PFOS and PFOA. Together with 10th and 90th percentile values. The data which this is based on can be found in Appendix A.*

Henry's sorption coefficient					
	K_H (min)	K_H (max)	K_H (mean)	K_H (10 th percentile)	K_H (90 th percentile)
PFOS	7.60E-10	1.90E+03	8.66E+01	2.60E-03	2.99E+01
PFOA	8.30E-09	3.70E+02	1.67E+01	4.69E-03	2.81E+00

Hydraulic conductivity

To simplify the calculations, the aquifer was assumed to be isotropic, meaning that the conductivity is the same in all directions ($K_X = K_Y = K_Z$). The values used in the modeling can be found in Table 8. These values represent the minimum, maximum, and average values based on collected data, which can be found in Appendix C. The minimum and maximum values were used in the modeling to create an interval from which the numbers were randomized. The default unit of hydraulic conductivity in FEFLOW is m/day, so the values in Table 8 need to be converted to the same unit before they are inserted into FEFLOW.

Table 8: *Minimum, maximum and average values found for the hydraulic conductivity in two different kinds of soil. The values were obtained from Gelhar et al., 1992.*

Hydraulic conductivity (m/s)			
Soil type	Min	Max	Average
Sand	$1.00 \cdot 10^{-6}$	$2.00 \cdot 10^{-4}$	$7.34 \cdot 10^{-6}$
Sand and Gravel	$8.10 \cdot 10^{-5}$	$3.00 \cdot 10^{-2}$	$5.25 \cdot 10^{-3}$

Groundwater recharge

Values for the groundwater recharge for three different soil types can be found in Table 9. In FEFLOW, groundwater recharge is inserted as inflow/outflow on the top/bottom. The standard unit for groundwater recharge in FEFLOW is 10^{-4} m/d, so the values in Table 9 had to be converted to match the unit. The minimum and maximum values were used as inputs in the Python script to create the input file. For a sandy aquifer, the value for till was used, and for the aquifer consisting of sand and gravel, the value for coarse soil was used.

Table 9: *Values for estimated maximum and minimum groundwater recharge in Sweden for three different soil types. Data is obtained from Rodhe et al., 2006.*

Groundwater recharge (mm/year)			
Soil type	Min	Max	Average
Coarse	300	600	450
Till	225	650	438
Fine-grained	150	650	400

4.1.3 Parameter interval

The parameters inserted into Python for the simulations of the transport of the two different PFAS in the two different types of aquifers can be found in Table 10. From these

intervals, a file was created containing 10,000 different combinations of the parameters ranging between their maximum and minimum values. Each row in the file represents one simulation, and due to time constraints, only the first 500 rows were utilized to conduct 500 simulations.

Table 10: *The range of the parameter values entered into Python for the simulation of the four different scenarios.*

	Simulations							
	PFOS in Sand		PFOA in sand		PFOS in sand and gravel		PFOA in sand and gravel	
	<i>Min</i>	<i>Max</i>	<i>Min</i>	<i>Max</i>	<i>Min</i>	<i>Max</i>	<i>Min</i>	<i>Max</i>
Hydraulic Conductivity (m/s)	1.00E-6	2.00E-4	1.00E-6	2.00E-4	8.10E-5	3.00E-2	8.10E-5	3.00E-2
Porosity (-)	2.50E-1	5.00E-1	2.50E-1	5.00E-1	2.50E-1	5.00E-1	2.50E-1	5.00E-1
Longitudinal dispersivity (m)	4.70E-2	5.50E-1	4.70E-2	5.50E-1	6.00E-1	3.50E+2	6.00E-1	3.50E+2
Transversal dispersivity (m)	4.70E-3	5.50E-2	4.70E-3	5.50E-2	6.00E-2	3.50E+1	6.00E-2	3.50E+1
Sorption (-)	2.60E-3	2.99E+1	4.69E-3	2.81E+0	2.60E-3	2.99E+1	4.69E-3	2.81E+0
Recharge (mm/yr)	2.25E+2	6.50E+2	2.25E+2	6.50E+2	3.00E+2	6.00E+2	3.00E+2	6.00E+2

4.1.4 Python

In Python, coding was done to execute multiple simulations with small variations in the input parameters. The previously created file containing the parameter values was used as an input in Python. Furthermore, code was written to easily perform multiple simulations. For steady state modeling, 500 simulations were conducted for each of the four scenarios, whereas for transient modeling, 200 simulations were performed. The four scenarios explored in the steady-state modeling were PFOS in a sandy aquifer, PFOA in a sandy aquifer, PFOS in a sand and gravel aquifer, and PFOA in a sand and gravel aquifer. For the transient modeling, only the scenario of PFOS in a sandy aquifer was simulated due to time constraints.

4.1.5 FEFLOW simulations

The simulations were conducted using FEFLOW in conjunction with Python. This part presented some challenges and proved time-consuming, as ensuring the proper functionality of the Python code to achieve the desired results was complex. Once the code began functioning, each simulation also required a considerable amount of time to execute, particularly in the case of transient state simulations. Consequently, only 500 simulations were carried out for each steady-state scenario, and a total of 200 simulations were completed for the transient state. The outcomes of each simulation were obtained as files containing the resulting concentrations in every node of the model. This meant that, for 500 simulations, a corresponding 500 output files were generated.

4.1.6 Random forest regression model

As mentioned earlier, 500 different output files were obtained from the simulations for each scenario in the case of steady state simulations. To create a random forest regression model, the data was processed to create a new file containing all 500 different concentrations of PFAS obtained at a certain observation point. This file, along with the file containing the values of the input parameters, was then used to train the random forest regression model. The model was trained on 375 data points (training dataset), which included both the input data and the output concentrations. The remaining 125 data points (test dataset) were used to evaluate the model. This was achieved by providing the model with the input parameters and allowing it to predict the output concentration. The model's predictions were then compared with the actual values. All of this was implemented in Python using the `sklearn` package. In the scenario of transient modeling, where 200 simulations were performed, the training dataset comprised 150 data points, while the remaining 50 constituted the test dataset used for evaluation purposes.

To assess the model's performance, metrics such as mean squared error (MSE), root mean squared error (RMSE), and R^2 value were examined using the same package, `sklearn`. The MSE was calculated using the `mean_squared_error` method, which takes the predicted values and the actual values as inputs. The RMSE was obtained by taking the square root of the MSE. Finally, the R^2 value was calculated using the `r2_score` method, with both the predicted and actual concentrations as inputs. All of these assessments were conducted on the test data set.

4.1.7 Sensitivity analysis

A sensitivity analysis is a technique used to determine how changes in independent variables will impact a target variable (EduPristine, 2020). In the context of this study, a sensitivity analysis is conducted to identify which input parameters have a significant impact on the resulting concentrations obtained in the simulations. The sensitivity analysis is performed using the random forest regression model. To achieve this, the algorithm is trained on 75% of the obtained results from the FEFLOW modeling and then tested on the remaining 25% to assess its performance.

By carefully examining the model's performance on the test data, a comprehensive sensitivity analysis can be conducted to understand the basis of its predictions. This analysis helps us determine the importance of each parameter according to the regression model, shedding light on the significant factors that influence the resulting concentrations.

Shapley values

The Shapley values were calculated in Python using the **SHAP** package. The input data included the previously trained random forest regression model and the input parameters from the test dataset. The result was obtained in a Shapley plot which visualize each feature's impact on a specific prediction. The obtained Shapley plots and an explanation of how to read them will be provided in the result section.

Random Forest Importance

The Random Forest Importance was calculated and plotted using the `feature_importances` method on the trained random forest regression model. The output was obtained as a plot, illustrating the relative importance of each parameter as well as the percentages associated with their importance.

5 Result

In this section, the results from both steady state and transient state simulations will be presented.

5.1 Steady state vs Transient state

Figure 8 displays the results obtained from a steady state simulation and a transient state simulation. The parameter values were the same for both simulations, with the only difference being the state in which the simulation was conducted. The transient state simulation was carried out over a duration of 50 years, with an initial time step of 100 days.

The obtained results show similarities, but there is a more distinct contaminant plume visible in the right image, indicating the temporal evolution of the contaminant distribution.

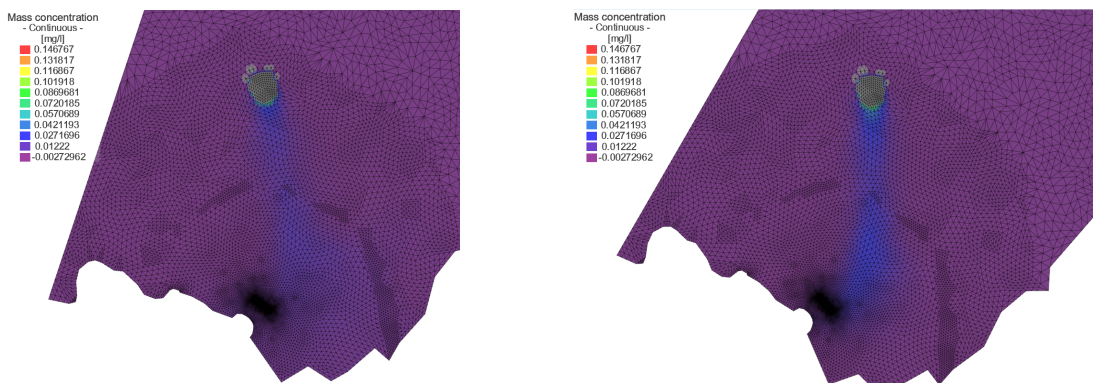


Figure 8: *The result obtained from one simulation in steady state and one simulation in transient state. The left image shows the result from the steady state model and the right image show the result from the transient state model (running time of 50 years).*

The validation of the created random forest regression models can be done by examining the R^2 values, which are a measure of the model's performance. The R^2 values for the models are presented in Table 11.

From the table, it can be observed that the R^2 values for the steady state models are generally very high, with all of the values exceeding 0.94. This indicates that the random forest regression models perform very well in predicting the concentration in a given node based on the input parameter values, based on the results obtained from the steady state modeling.

However, the random forest regression model, based on the results from the transient modeling, demonstrates only moderate performance. The obtained R^2 values for observation points 1 and 2 are significantly lower compared to the other values and also exhibit substantial differences from each other. A low R^2 value of 0.502 suggests poor predictive

performance, while a value of 0.767 indicates relatively decent performance. The relatively poorer performance in the transient state can be attributed to the limited amount of data available for the algorithm to train on. With fewer data points, the random forest regression model may struggle to capture complex patterns and relationships, leading to less accurate predictions compared to when it has a more extensive dataset to train on.

Table 11: *The R^2 values for all the different created random forest regression models.*

	Obs. 1	Obs. 2	Obs 3
Scenario	R^2	R^2	R^2
PFOS - Sand	0.978	0.974	0.973
PFOA - Sand	0.940	0.940	0.925
PFOS - Sand/Gravel	0.951	0.968	0.979
PFOA - Sand/Gravel	0.976	0.978	0.983
Transient PFOS - Sand	0.502	0.767	0.958

Since the models perform so well, it is interesting to see what they general base their predictions on. This will be investigated further by examining the Shapley values and the random forest importance values.

5.2 Shapley values

The obtained Figures for the Shapley values at the three different observation points were rather similar; therefore, only the ones obtained for point 1 will be presented. The Shapley values for observation point 2 and 3 can be found in Appendix D.

Figure 9 shows the Shapley values for the six different parameters in observation point 1. These values are based on the Random Forest model that was made for the 500 simulations of each of the four scenarios. Several observations can be made from the Figure. Each point in the Figure represents a Shapley value for a parameter and one simulation. The position of the points on the y-axis is determined by the parameter, while their position on the x-axis is determined by the Shapley value. The points are color-coded to indicate whether the parameter value is low or high. The pink color indicates a parameter value higher than the average, while the blue color indicates a value lower than the average. Due to the color-coding, it becomes possible to visualize how the parameters impact the results based on whether they have a low or high value. However, in some cases, there is no clear pattern. Instead, the color-coded points may be spread on both sides of the Figure, making it challenging to draw any definitive conclusions about how the parameter values impact the resulting concentration. If multiple points are located at the same position, they are spread out around the y-axis to show the distribution of Shapley values for each parameter. The parameters are ordered according to their importance, with the most important parameters at the top of the Figure. Points located on the left side of

the y-axis indicate a negative contribution to the predicted output concentration, while points on the right side indicate a positive contribution.

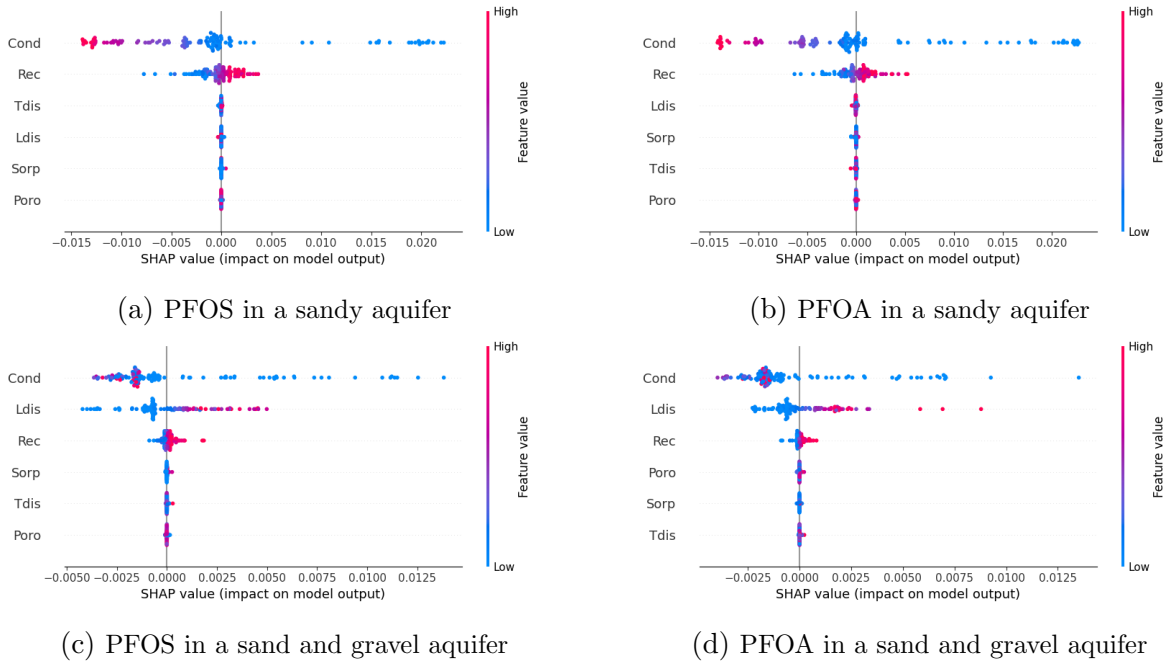


Figure 9: The Shapley values for the four different scenarios modeled in steady state, in observation point 1. Blue color indicates low values of the parameters, while the pink color indicates high values (as the scale to the right present). The points located on the left side of the y-axis indicated a negative contribution to the predicted output concentration, while points on the right side indicates a positive contribution.

As can be seen in all four Figures, the parameter with the greatest importance seems to be the hydraulic conductivity. For the two scenarios with a sandy aquifer, high values of the hydraulic conductivity (pink dots) end up far on the left side, which indicates that high values have a high negative contribution to the predicted concentration. It can also be seen that some low values of the hydraulic conductivity (blue dots) end up far to the right in the Figure, which indicates that low values have a high positive contribution to the predicted concentration. When examining the Shapley values for the hydraulic conductivity in a sand and gravel aquifer, not much can be said about the low values of the hydraulic conductivity since they are present on both sides of the y-axis. However, the high values are all located on the left side, and therefore it is possible to state that high values of the hydraulic conductivity have a negative contribution to the prediction also in a sand and gravel aquifer.

Furthermore, for the two scenarios with the sandy aquifer, the recharge is clearly the second most important parameter according to the Shapley values. Here it is also clear that low values have a negative contribution and high values have a positive contribution. The Shapley values for the remaining four parameters are all located close to or on the y-axis and therefore do not seem to contribute to the prediction at all.

For the sand and gravel aquifers, the second most important parameter, according to the Shapley values, is the longitudinal dispersivity. High values for the longitudinal dispersivity consistently appear on the right side of the y-axis, indicating a positive contribution to the prediction. On the other hand, low values tend to be located on the left side or close to the y-axis on the right side, suggesting a negative contribution.

Additionally, recharge also demonstrates some level of importance, although its contributions appear relatively small. High values of recharge show a slight positive contribution, while low values exhibit a small negative contribution.

Furthermore, the Shapley values for the transient modeling of PFOS in a sandy aquifer can be seen in Figure 10. These values are obtained from the model based on observation point 3, which exhibited the best performance and thus holds the greatest interest for analyzing Shapley values. The Shapley values for observation point 1 and 2 can be found in Appendix D.

Compared to Figure 9, a notable difference is the importance of sorption. In the transient modeling of observation point 3, sorption emerges as the most influential parameter. High values of sorption make a significant positive contribution to the prediction, while low values have a substantial negative contribution. Hydraulic conductivity also retains its importance in this case, with low values exhibiting a positive contribution and high values showing a negative contribution to the predicted concentrations. The remaining parameters appear to have negligible contributions to the prediction.

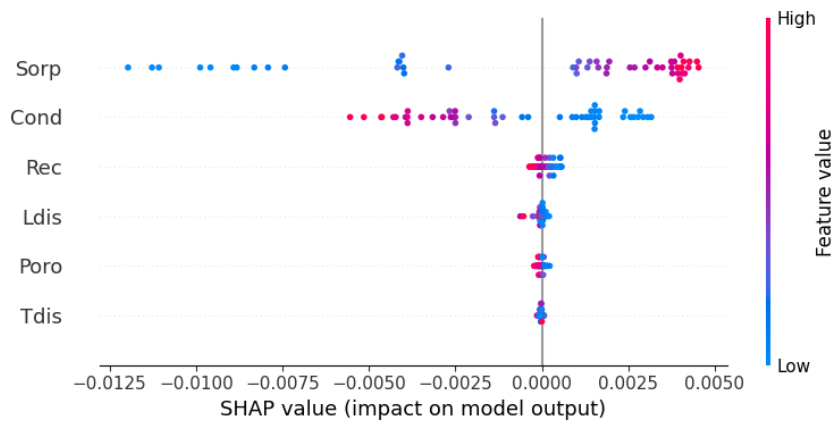


Figure 10: *The Shapley values for the four different scenarios modeled in transient state, in observation point 3. Blue color indicates low values of the parameters, while the pink color indicates high values (as the scale to the right present). The points located on the left side of the y-axis indicated a negative contribution to the predicted output concentration, while points on the right side indicates a positive contribution.*

5.3 Random forest importance

The random forest importance provides valuable insights into the parameters considered most important by the created random forest regression model. Figure 11 presents a heat

map of the random forest importance for five different modelings conducted in observation point 3. The heat map is structured as a matrix, where each row represents a different input parameter and each column represents one of the five models. The sum of each column is equal to 1. The colors of the squares in the heat map indicate the range of values, with dark red indicating high values and dark blue indicating low values. A high value signifies a parameter of great importance, while a small value suggests a parameter of low importance.

Upon examining the Figure, one notable observation is that hydraulic conductivity holds significant importance in all five models. However, the magnitude of its importance varies across the models, with the highest importance observed in the steady state models of sandy aquifers. In the steady state models of sand and gravel aquifers, hydraulic conductivity remains the most important parameter, although its importance is slightly lower.

In contrast, the column representing the transient model stands out from the others. While hydraulic conductivity remains an important parameter, sorption emerges as the parameter with the highest importance, indicating its significant role in predicting the concentrations. Additionally, across all five models, both porosity and transversal dispersivity exhibit importance values close to zero, indicating that these parameters have little influence on the predictions made by the random forest regression model.

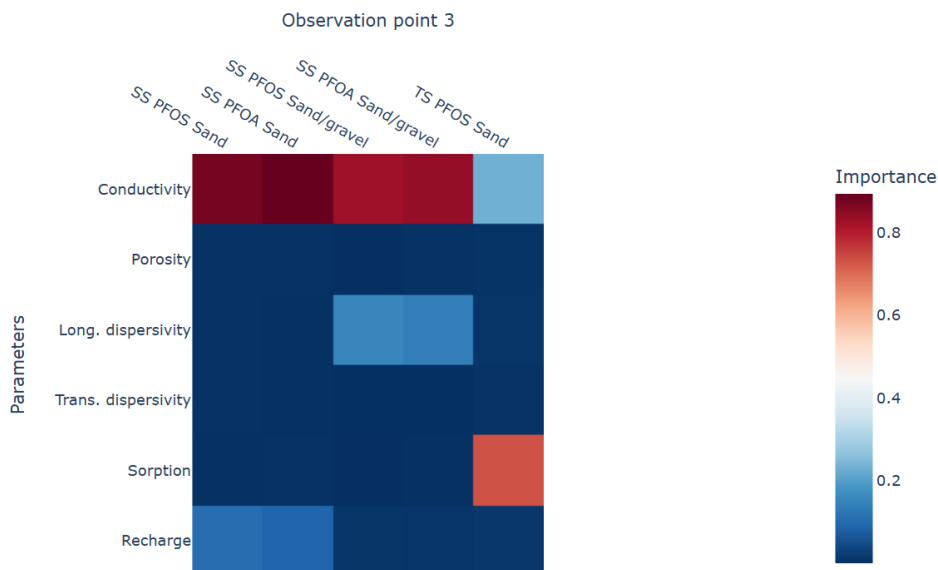


Figure 11: *The heat-map for the random forest importance values obtained for the five different models in observation point 3.*

Figure 12 presents the heat-map of the random forest importance values for observation point 1. The heat-maps for both observation points exhibit similarities, but there are notable differences as well. In the column representing the transient model, sorption is of less importance at observation point 1 compared to observation point 3. Another difference is observed in the importance values of hydraulic conductivity for the steady state models of a sand and gravel aquifer, which are higher in point 3 compared to point

1. Furthermore, it can be seen that the longitudinal dispersivity is more important at point 1 compared to point 3. The differences observed between Figure 11 and 12 for the simulations conducted in steady state indicate that the importance of certain parameters may vary depending on the observation point and its location relative to the source.

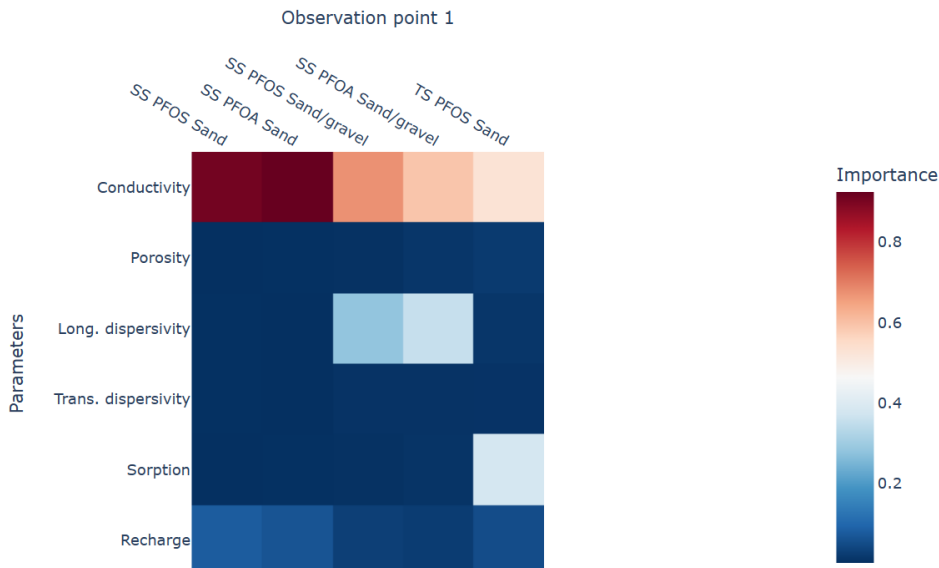


Figure 12: *The heat-map for the random forest importance values obtained for the five different models in observation point 1.*

The heat-map for observation point 2 are quite similar to others and can be found in Appendix E.

5.4 Concentration

Figure 13 and 14 illustrate the concentrations obtained at the three observation points during the first 50 simulations of PFOS transport in a sandy aquifer. Figure 13 corresponds to the steady state modeling, while Figure 14 represents the transient state modeling.

A notable distinction between the two Figures is that, in the case of steady state modeling, the concentrations at the observation points exhibit more similarity compared to the transient state modeling. Additionally, in the transient state modeling, the concentrations are consistently highest at point 3 and lowest at point 1. Conversely, in the steady state modeling, point 2 tends to have the highest concentration levels when concentrations are generally high, while point 3 dominates when concentrations are generally low.

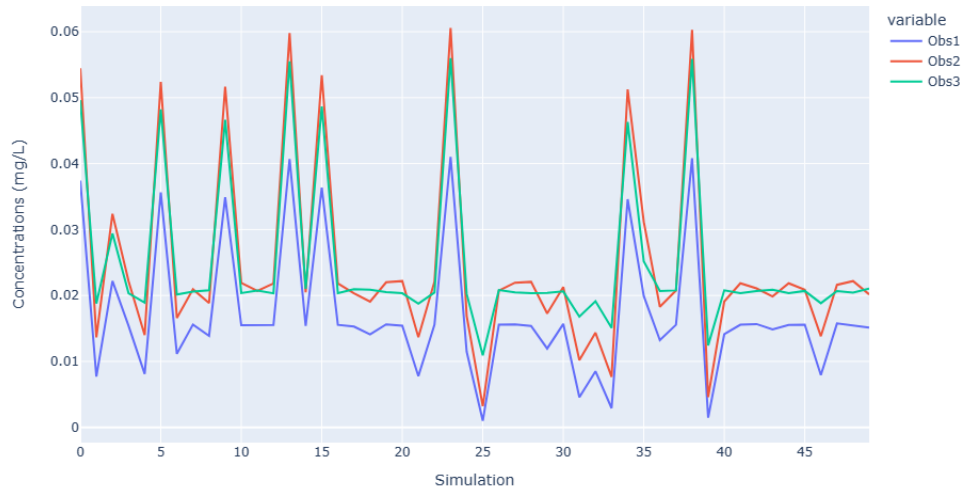


Figure 13: *The concentration of PFOS in the observation points for the 50 first simulations of PFOS in a sandy aquifer modeled in steady state.*

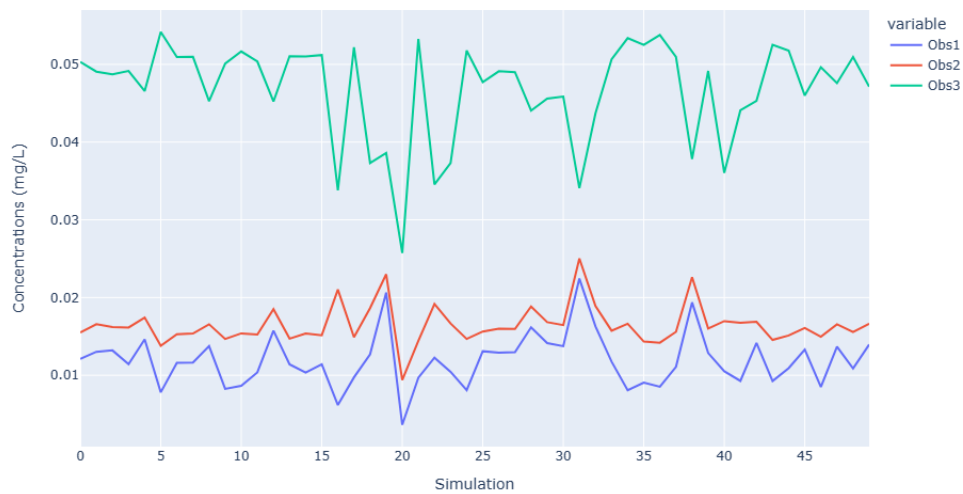


Figure 14: *The concentration of PFOS in the observation points for the 50 first simulations of PFOS in a sandy aquifer modeled in transient state.*

6 Discussion

6.1 Methodology

Since the literature review did not yield the desired results in evaluating various modeling strategies for PFAS, a change in the methodology was necessary. Initially, the intention was to compare different modeling approaches employed for PFAS modeling. However, it became apparent that most models followed a similar approach, limiting the scope for meaningful comparisons. As a result, the focus shifted towards identifying the input parameters with the greatest influence on the resulting concentration.

To achieve this, various methods could have been employed, but the selected approach involved the use of a machine learning algorithm trained to predict concentrations. By training the algorithm to predict the concentration in a specific node, it became possible to analyze the algorithm and gain insights into which parameters have the most significant impact on the resulting concentration. This approach was chosen due to its potential to provide valuable information regarding parameter importance, utilizing random forest importance and the SHAP library, otherwise unavailable in conventional sensitivity analyses.

The random forest regression models based on steady-state simulations exhibited excellent performance, with R^2 values consistently above 0.9. This observation is meaningful as it indicates the feasibility of utilizing machine learning algorithms to predict concentrations in a FEM model. Given the strong performance of the random forest regression models, it became of great interest to investigate the factors underlying their predictions.

However, the R^2 values obtained from the random forest regression models based on transient modeling were not as favorable, particularly for observation points 1 and 2. The relatively limited number of data points available for training these models may have contributed to their lower performance. In general, a larger dataset enables the algorithm to perform better. Due to time constraints, in the transient modeling, only 200 data points were obtained, with only 150 of them used for training. This is significantly lower than the steady-state scenario, where 500 data points were obtained, with 375 used for training.

6.2 Concentrations

Steady-state and transient simulations yield different results, as shown in Figures 8, 13, and 14. A steady-state simulation assumes an infinite time frame, whereas the transient modeling in this case was conducted over a period of 50 years. Consequently, it is expected that the results between these two models would differ.

In the steady-state model, where time is set to infinity, the concentrations observed in the observation points are relatively similar. This can be explained by the diminished significance of distance from the original source over an infinite time frame.

On the other hand, the concentrations observed in the observation points of the transient

model exhibit noticeable differences. Particularly, the concentrations at observation point 3 are significantly higher compared to the other two points. This variation can be explained by the location of the observation points within the model. In the transient model, the finite time frame makes the position compared to the original source more influential in determining the resulting concentrations.

6.3 Importance of parameters

The great performance of the random forest regression model makes it interesting to identify the key parameters on which the predictions are based. Figure 9 illustrates that hydraulic conductivity significantly impacts the resulting concentration in the steady-state model. The same observation is evident in Figure 10, depicting the Shapley values for transient modeling. While not the most critical parameter in this case, hydraulic conductivity remains of considerable importance. In the case of a sandy aquifer (both in steady state and transient state modeling), higher values of hydraulic conductivity are associated with lower concentrations of PFAS, while lower values lead to increased concentrations. This can be explained by two main effects of a high value of hydraulic conductivity in a steady-state model: dilution and faster transport, both of which may have a decreasing impact on the obtained concentration at observation points.

Higher hydraulic conductivity results in more water flowing through the system, leading to the dilution of PFAS concentrations and, consequently, lower observed concentrations at the observation points. Additionally, the increased flow velocity through the aquifer facilitates faster transport of the PFAS plume from the source to the downstream observation points. As a result, the PFAS spends less time in the subsurface, leading to lower concentrations at the observation points compared to when the flow velocity is lower.

Based on the observed Shapley values for hydraulic conductivity in sand and gravel aquifers (see Figure 9, it can be concluded that this parameter significantly affects the resulting concentration. However, due to the distribution of data points on both sides of the Figure, with low values (blue dots) appearing on both sides, it is challenging to analyze the impact of hydraulic conductivity on the results in these scenarios. However, it can be observed that high values (pink dots) are only found on the left side of the Figure. This implies that high hydraulic conductivity values tend to lead to decreased concentrations of PFAS. This conclusion aligns with the findings from the scenarios based on a sandy aquifer.

Figure 9 suggests that sorption does not have a significant impact on the resulting concentration in the case of steady state, as the Shapley values associated with sorption are relatively low compared to other parameters. This observation is consistent with the comparison of Figures, where changes in the composition of the aquifer appear to have a greater influence on the resulting concentration than the variations in the specific PFAS compound being modeled. However, it is important to note that these conclusions are based on the specific model and data analyzed, and further investigations and analysis may be required to fully understand the role of sorption in the transport and fate of PFAS compounds in groundwater.

In contrast to the steady-state model, the transient model reveals that sorption plays a crucial role in determining the resulting concentration of PFAS, as indicated by the Shapley values obtained from observation point 3. The higher the sorption values, the higher the concentrations observed at point 3, while lower sorption values lead to decreased concentrations. These findings are further supported by the results from the random forest importance analysis, which highlight sorption as the most influential parameter for predicting the concentration, as shown in Figure 11. Sorption is the processes where a fraction of the contaminant present in the water adsorb to the soil. This, in turn, affects the transport, since some PFAS molecules become trapped in the soil along their journey, leading to increased PFAS concentrations in those areas. Moreover, adsorbed PFAS can also be desorbed, introducing a delay in the transport process. With all these factors considered, it's reasonable that sorption emerges as an important parameter for predicting concentrations.

These results underscore the significance of sorption and the importance of specifying the specific PFAS compound being modeled in order to obtain accurate results. As mentioned in the background, sorption is one of the unique characteristics of PFAS. Therefore, it is not surprising that this characteristic plays a major role in simulating PFAS transport. However, it is important to note that these results are based on one specific node in the FEM model and only 200 simulations. Further investigations in this area are required to draw more definitive conclusions.

6.4 Modeling Insights and Uncertainties

In general, steady-state models do not perform as well as transient models in simulations of contaminant transport. This conclusion can also be drawn from the results, as the sorption factor does not seem to impact the outcome at all in the steady-state simulation, which is an unexpected finding. Consequently, it can be concluded that transient modeling is the preferred method for obtaining the most accurate results. However, it's important to note that transient modeling is more intricate and time-consuming. Hence, steady-state modeling still holds value as a quicker and simpler method to attain an initial understanding of the situation. Nonetheless, when considering the time scales—50 years compared to infinity - 50 years is more representative of real-life scenarios. Given this perspective, steady-state simulations might prove overly simplistic.

Another interesting observation is that the chosen modeling approach has a significant impact on the obtained result. This highlights the importance of further investigating this matter to gain a better understanding of which method should be used to simulate the transport of PFAS as accurately as possible.

Furthermore, as the results also indicate, it is feasible to create a random forest regression model capable of forecasting the concentration in a specific node by leveraging input parameters. This discovery not only paves the way for additional exploration but also implies that such a model could prove effective in scenarios where a transient model is

necessary without compromising efficiency. Nonetheless, this assumption might be valid due to the model's inherent simplicity. In the event of adopting a more intricate model involving multiple layers and an isotropic aquifer, the random forest regression model's performance could potentially decline as increased complexity might act as a hindrance.

Another notable observation found during the literature review is the presence of significant uncertainties when determining values for various parameters used in groundwater modeling. These uncertainties contribute to substantial variations in the obtained results. The parameter values employed in this study were derived from diverse literature sources, resulting in considerable variations and wide intervals for the input parameters. Although these large intervals introduce uncertainties in the modeling process, they may also have some advantages. The wide range of values provides a large training data for the random forest regression model, allowing for a comprehensive exploration of each parameter's impact on the results. Consequently, this can lead to a greater understanding of the influence of individual parameters on the overall outcome.

6.5 Future work

These results provide an indication of the significant role of sorption in modeling the transport of PFAS in groundwater. However, it is important to note that this conclusion is based on a limited number of simulations, and further investigation is necessary to validate and generalize these findings. This finding underscores the importance of identifying the specific PFAS compounds present, as their varying affinities for sorption can greatly influence their transport behavior in the groundwater system. Therefore, a comprehensive understanding of the composition of PFAS compounds is crucial for accurate modeling and prediction of their fate and transport in the subsurface environment.

To obtain more reliable and robust results, it is necessary to conduct and analyze additional simulations. This includes exploring various scenarios involving different PFAS compounds and aquifer types, beyond just the PFOS in a sandy aquifer scenario. Furthermore, to gain a better understanding of the impact of sorption, it would be valuable to investigate scenarios where multiple PFAS compounds are modeled together and compare them with simulations focusing on a single species.

7 Conclusion

To conclude, MODFLOW/MT3DMS is currently the most commonly used modeling tool for simulating the transport of PFAS in the saturated zone. Sorption is typically the only reactive process considered, and parameter values for sorption are often based on literature findings or experimental data.

In steady-state simulations, hydraulic conductivity appears to have the greatest impact on the results. However, longitudinal dispersivity also plays a significant role when modeling in a sand and gravel aquifer. For transient modeling, sorption becomes more important compared to the steady-state model. Even though the hydraulic conductivity remains an important parameter in this case as well.

Since sorption is the only PFAS-specific parameter and its impact is not significant in steady-state modeling, it can be concluded that the choice of PFAS compound may not matter when using this approach. However, in transient modeling, sorption becomes crucial, indicating that specifying the compounds of PFAS to be modeled is more important. Although, further investigation is needed to explore this aspect in more detail.

Furthermore, the results show that the chosen modeling approach has a significant impact on the results, emphasizing the importance of further investigations in this subject. In general, a conclusion that can be drawn is that using transient modeling instead of steady-state modeling generally provides more accurate results of the situation. Transient modeling takes into account the time-dependent behavior of the system, allowing for a better understanding of how concentrations change over time. This is particularly important for dynamic systems where concentrations can vary significantly. Steady-state modeling, on the other hand, assumes a state of equilibrium where concentrations remain constant, which may not accurately capture the real-world behavior of contaminants in the aquifer.

References

- Adams, E. and L. Gelhar (1992). “Field study of dispersion in a heterogeneous aquifer: 2. Spatial moments analysis”. In: *Water Resources Research* 28.12, pp. 3293–3307. DOI: <https://doi.org/10.1029/92WR01757>.
- Agency, Environment (2019). *Perfluorooctane sulfonate (PFOS) and related substances: sources, pathways and environmental data*. URL: https://consult.environment-agency.gov.uk/environment-and-business/challenges-and-choices/user_uploads/perfluorooctane-sulfonate-and-related-substances-pressure-rbmp-2021.pdf.
- Aguilar, Z. (2013). “Chapter 2 - Types of Nanomaterials and Corresponding Methods of Synthesis”. In: *Nanomaterials for Medical Applications*. Ed. by Zoraida P. Aguilar. Elsevier, pp. 33–82. ISBN: 978-0-12-385089-8. DOI: <https://doi.org/10.1016/B978-0-12-385089-8.00002-9>.
- Ahlstrom, S. W. et al. (1977). *Multicomponent mass transport model: Theory and numerical implementation (discrete-particle-random-walk-version)*. Battelle Pac. Northwest Lab., Richland, Wash.
- Ahrens, L. (2011). “Polyfluoroalkyl compounds in the aquatic environment: a review of their occurrence and fate”. In: *J. Environ. Monit.* 13 (1), pp. 20–31. DOI: 10.1039/COEM00373E.
- ALS Europe (2023). *PFOS och PFOA*. URL: <https://www.alsglobal.se/miljoanalys/pfas/om-pfos>.
- Ambaye, T. et al. (2022). “Recent progress and challenges on the removal of per- and polyfluoroalkyl substances (PFAS) from contaminated soil and water”. In: *Environmental Science and Pollution Research* 29.39, pp. 58405–58428. DOI: 10.1007/s11356-022-21513-2.
- Andersson, M.P., W.W. Woessner, and R.J. Hunt (2015). *Applied Groundwater Modeling - Simulation of Flow and Advective Transport*. 2nd ed. San Diego: Elsevier Scuebce Publishing Co Inc. URL: https://agualabs.edublogs.org/files/2022/02/Applied-Groundwater-Modeling-Second-Edition-Simulation-of-Flow-and-Advective-Transport-by-Mary-P.-Anderson-William-W.-Woessner-Randall-J.-Hunt-z-lib.org_.pdf.
- Arp, H.P., C. Niederer, and K.U. Goss (2006). “Predicting the Partitioning Behavior of Various Highly Fluorinated Compounds”. In: *Environmental Science & Technology* 40.23, pp. 7298–7304. DOI: 10.1021/es060744y.
- ATSDR (2022). *What are the health effects of PFAS?* <https://www.atsdr.cdc.gov/pfas/health-effects/index.html> [Accessed: 15-02-2023].
- Bagheri, R. (Apr. 2022). *Introduction to SHAP Values and their Application in Machine Learning*. [Accessed: 17-05-2023]. URL: <https://towardsdatascience.com/introduction-to-shap-values-and-their-application-in-machine-learning-8003718e6827#:~:text=SHAP%5C%20is%5C%20a%5C%20mathematical%5C%20method,each%5C%20feature%5C%20to%5C%20the%5C%20prediction>.
- Boesel, D. et al. (2000). “Design, performance, evaluation, and modeling of a natural gradient multitracer transport experiment in a contaminated heterogeneous porous aquifer.” In: *Tracers and Modelling in Hydrology*. Burlington: IAHS Publ., pp. 45–52. ISBN: 0144-7815.

- Boonraksasat, W. (2019). “Groundwater Flow and Transport Modelling of PFASs in Åkersberga”. Master’s Thesis. University of Uppsala. URL: <https://www.diva-portal.org/smash/get/diva2:1298129/FULLTEXT01.pdf>.
- Brusseau, M. (2020). “Simulating PFAS transport influenced by rate-limited multi-process retention”. In: *Water Research* 168, p. 115179. ISSN: 0043-1354. DOI: <https://doi.org/10.1016/j.watres.2019.115179>.
- Campos Pereira, H. et al. (2018). “Sorption of perfluoroalkyl substances (PFASs) to an organic soil horizon – Effect of cation composition and pH”. In: *Chemosphere* 207, pp. 183–191. ISSN: 0045-6535. DOI: <https://doi.org/10.1016/j.chemosphere.2018.05.012>.
- Chiang, C. Y. et al. (1989). “Aerobic Biodegradation of Benzene, Toluene, and Xylene in a Sandy Aquifer—Data Analysis and Computer Modeling”. In: *Groundwater* 27.6, pp. 823–834. DOI: <https://doi.org/10.1111/j.1745-6584.1989.tb01046.x>.
- Choudhary, A. (Nov. 2019). *A Unique Method for Machine Learning Interpretability: Game Theory Shapley Values!* [Accessed: 17-05-2023]. URL: <https://www.analyticsvidhya.com/blog/2019/11/shapley-value-machine-learning-interpretability-game-theory/>.
- DHI, MIKE Powered by (2023a). “Description of Mass-Related Material Properties”. In: *FEFLOW 7.2 Documentation*. URL: http://www.feflow.info/html/help72/feflow/09_Parameters/Material_Properties/mass_parameters.html.
- (2023b). “Problem class”. In: *FEFLOW 7.2 Documentation*. URL: http://www.feflow.info/html/help72/feflow/08_ProblemSettings/problem_class.html.
- (2023c). “Transport settings”. In: *FEFLOW 7.2 Documentation*. URL: http://www.feflow.info/html/help72/feflow/08_ProblemSettings/Mass/transport_settings.html?rhsearch=Divergence%5C%20form&rhsyns=%5C%20.
- Du, Z. et al. (2014). “Adsorption behavior and mechanism of perfluorinated compounds on various adsorbents—A review”. In: *Journal of Hazardous Materials* 274, pp. 443–454. ISSN: 0304-3894. DOI: <https://doi.org/10.1016/j.jhazmat.2014.04.038>.
- Earle, S. (2015). “Physical Geology”. In: Victoria, B.C.: BCcampus. URL: <https://opentextbc.ca/geology/>.
- EduPristine (Apr. 2020). *All you want to know about Sensitivity Analysis*. [Accessed: 17-05-2023]. URL: <https://www.edupristine.com/blog/all-about-sensitivity-analysis>.
- Edvinsson, J. (2015). “Grundvattenmodellering och föroreningstransport av PFOS i Bredåkradeltat”. Master’s Thesis. University of Stockholm. URL: <http://diva-portal.org/smash/get/diva2:842948/FULLTEXT01.pdf>.
- Egboka, B.C. E. et al. (1983). “Migration of contaminants in groundwater at a landfill: A case study, 3, Tritium as an indicator of dispersion and recharge”. In: 63, pp. 51–80.
- European Environment Agency (2022). *Emerging chemical risks in Europe - 'PFAS'*. <https://www.eea.europa.eu/publications/emerging-chemical-risks-in-europe> [Accessed: 15-02-2023].
- Fernando, J. (2023). *R-Squared: Definition, Calculation Formula, Uses, and Limitations*. [Accessed 29-05-2023].
- Fitts, C. (2013). “11 - Groundwater Contamination”. In: *Groundwater Science (Second Edition)*. Ed. by Charles R. Fitts. Second Edition. Boston: Academic Press, pp. 499–

585. ISBN: 978-0-12-384705-8. DOI: <https://doi.org/10.1016/B978-0-12-384705-8.00011-X>.
- Freyberg, D. (1986). “A natural gradient experiment on solute transport in a sand aquifer: 2. Spatial moments and the advection and dispersion of nonreactive tracers”. In: *Water Resources Research* 22.13, pp. 2031–2046. DOI: <https://doi.org/10.1029/WR022i013p02031>.
- Garabedian, S. P., L. W. Gelhar, and M. A. Celia (1988). *Large-scale dispersive transport in aquifers: Field experiments and reactive transport theory*. Mass. Inst. of Technol. Cambridge.
- Ge, S. (2003). “HYDROLOGY - Ground and Surface Water”. In: *Encyclopedia of Atmospheric Sciences*. Ed. by James R. Holton. Oxford: Academic Press, pp. 973–979. ISBN: 978-0-12-227090-1. DOI: <https://doi.org/10.1016/B0-12-227090-8/00171-8>.
- Gefell, M. et al. (2022). “Modeling PFAS Fate and Transport in Groundwater, with and Without Precursor Transformation”. In: *Groundwater* 60.1, pp. 6–14. DOI: <https://doi.org/10.1111/gwat.13152>.
- Gelhar, L., C. Welty, and K. Rehfeldt (1992). “A critical review of data on field-scale dispersion in aquifers”. In: *Water Resources Research* 28.7, pp. 1955–1974. DOI: <https://doi.org/10.1029/92WR00607>.
- Gupta, A. (2020). *Feature Selection Techniques in Machine Learning (Updated 2023)*. [Accessed: 29-05-2023]. URL: <https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/?fbclid=IwAR3R4itkn-%5C%5C%20A822saQsBZ4IxTPREfzBV-brKBbxDmiVH7109vTdxsiyEjwbqo>.
- Hartmann, A. (2022). “The Hydrology of Groundwater Systems - From Recharge to Discharge”. In: *Encyclopedia of Inland Waters (Second Edition)*. Ed. by Thomas Mehner and Klement Tockner. Second Edition. Oxford: Elsevier, pp. 324–330. ISBN: 978-0-12-822041-2. DOI: <https://doi.org/10.1016/B978-0-12-819166-8.00097-9>.
- Hassanzadeh, S. Majid (1996). “On the Transient Non-Fickian Dispersion Theory”. In: *Transport in Porous Media* 23, pp. 107–124.
- Haug, L. et al. (2011). “Characterisation of human exposure pathways to perfluorinated compounds — Comparing exposure estimates with biomarkers of exposure”. In: *Environment International* 37.4, pp. 687–693. ISSN: 0160-4120. DOI: <https://doi.org/10.1016/j.envint.2011.01.011>.
- Ho, Y., C. Huang, and H.W. Huang (2002). “Equilibrium sorption isotherm for metal ions on tree fern”. In: *Process Biochemistry* 37.12, pp. 1421–1430. ISSN: 1359-5113. DOI: [https://doi.org/10.1016/S0032-9592\(02\)00036-5](https://doi.org/10.1016/S0032-9592(02)00036-5).
- Hoehn, E (1983). “Advection-dispersion interpretation of tracer observations in a river/-groundwater infiltration system”. In: *International conference on modern approach to groundwater resources management*, pp. 361–371.
- Hoehn, E. and P. Santschi (1987). “Interpretation of tracer displacement during infiltration of river water to groundwater”. In: *Water Resources Research* 23.4, pp. 633–640.
- Interstate Technology Regulatory Council, ITRC (2022). “Environmental Fate and Transport processes”. In: *PFAS Technical and Regulatory Guidance Document and Fact Sheets*. URL: <https://pfas-1.itrcweb.org/5-environmental-fate-and-transport-processes/?print=pdf>.
- Iris, P (1980). *Contribution à l'étude de la valorisation énergétique des aquifères peu profonds*. Ecole des Mines de Paris, Fontainebleau, France.

- ITRC, Interstate Technology Regulatory Council (2020a). “Environmental Fate and Transport for Per- and Polyfluoroalkyl Substances”. In: URL: https://pfas-1.itrcweb.org/fact_sheets_page/PFASFact_Sheet_Fate_and_Transport_April2020.pdf.
- (2020b). “Naming Conventions and Physical and Chemical Properties of Per- and Polyfluoroalkyl Substances (PFAS)”. In: URL: https://pfas-1.itrcweb.org/fact_sheets_page/PFAS_Fact_Sheet_Naming_Conventions_April2020.pdf.
- Jensen, K., K. Bitsch, and P. L. Bjerg (1993). “Large-scale dispersion experiments in a sandy aquifer in Denmark: Observed tracer movements and numerical analyses”. In: *Water Resources Research* 29.3, pp. 673–696. DOI: <https://doi.org/10.1029/92WR02468>.
- Jonasson, S. et al. (2007). *Modeller för transport och spridning av föroreningar fas 2*. 5692. Naturvårdsverket. URL: <https://www.diva-portal.org/smash/get/diva2:1634023/FULLTEXT01.pdf>.
- Kies, B (1981). “Solute transport in unsaturated field soil and in ground- water, Ph.D. dissertation”. Ph.D. Dep. of Agron., N.M., State Univ., Las Cruces.
- Kim, M. et al. (2015). “Selecting reliable physicochemical properties of perfluoroalkyl and polyfluoroalkyl substances (PFASs) based on molecular descriptors”. In: *Environmental Pollution* 196, pp. 462–472. ISSN: 0269-7491. DOI: <https://doi.org/10.1016/j.envpol.2014.11.008>.
- Konikow, L. and J. Bredehoeft (1974). “Modeling flow and chemical quality changes in an irrigated stream-aquifer system”. In: *Water Resources Research* 10.3, pp. 546–562.
- Kreft, A. et al. (1974). “Determination of effective porosities by the two-well pulse method”. In: *International Atomic Energy Agency (IAEA): IAEA*.
- Kutsuna, S. and H. Hori (2008). “Experimental determination of Henry’s law constant of perfluorooctanoic acid (PFOA) at 298K by means of an inert-gas stripping method with a helical plate”. In: *Atmospheric Environment* 42.39, pp. 8883–8892. ISSN: 1352-2310. DOI: <https://doi.org/10.1016/j.atmosenv.2008.09.008>.
- Larsson, M. (2003). *Hydrogeologiska transportprocesser och modeller för föroreningsspridning*. Total försvarets forskningsinstitut. URL: <https://www.foi.se/rest-api/report/FOI-R--0830--SE>.
- Lau, L. K. and W. J. Kaufman and D. K. Todd (1957). *Studies of dispersion in a radial flow system, Canal Seepage Research: Dispersion Phenomena in Flow Through Porous Media*. Sanit. Eng. Res. Lab., Dep. of Eng. and School of Public Health, Univ. of Calif., Berkeley.
- Lee, D.R., J. Cherry, and J. Pickens (1980). “Groundwater transport of a salt tracer through a sandy lakebed”. In: *Limnology and Oceanography* 25.1, pp. 45–61. DOI: <https://doi.org/10.4319/lo.1980.25.1.0045>.
- Li, F., X. Fang, et al. (2019). “Adsorption of perfluorinated acids onto soils: Kinetics, isotherms, and influences of soil properties”. In: *Science of The Total Environment* 649, pp. 504–514. ISSN: 0048-9697. DOI: <https://doi.org/10.1016/j.scitotenv.2018.08.209>.
- Li, H., D. Ellis, and D. Mackay (2007). “Measurement of Low AirWater Partition Coefficients of Organic Acids by Evaporation from a Water Surface”. In: *Journal of Chemical & Engineering Data* 52.5, pp. 1580–1584. DOI: [10.1021/jc600556d](https://doi.org/10.1021/jc600556d).

- Livsmedelsverket (2022). *Pfas I Dricksvatten och livsmedel - kontroll*. <https://www.livsmedelsverket.se/foretagande-regler-kontroll/dricksvattenproduktion/riskhantering-pfas-i-dricksvatten-egenfangad-fisk> [Accessed: 13-02-2023)].
- Livsmedelsverket et al. (2021). “Kartläggning av per- och polyfluorerade alkylsubstanser (PFAS) i Sveriges kommunala rå- och dricksvatten.” In: *Livsmedelsverketsrapportserie* 21. URL: <https://www.livsmedelsverket.se/globalassets/publikationsdatabas/rapporter/2021/1-2021-nr-21-kartlaggning-av-per-och-polyfluorerade-alkylsubstanser.pdf>.
- Lu, J. (2015). “Chapter 6 - Identification of Forensic Information from Existing Conventional Site-Investigation Data”. In: *Introduction to Environmental Forensics (Third Edition)*. Ed. by Brian L. Murphy and Robert D. Morrison. Third Edition. San Diego: Academic Press, pp. 149–164. ISBN: 978-0-12-404696-2. DOI: <https://doi.org/10.1016/B978-0-12-404696-2.00006-0>.
- Lyu, X., F. Xiao, and D. Wang (2021). “Critical Review on the Fate and Transport of Per- and Polyfluoroalkyl Substances (PFAS) in Subsurface Environments”. In: *Authorea*. DOI: <https://doi.org/10.1002/essoar.10508278.1> [Accessed: 16-02-2023)].
- Mas-Pla, J. et al. (1992). “A Forced Gradient Tracer Experiment in a Coastal Sandy Aquifer, Georgetown Site, South Carolina”. In: *Groundwater* 30.6, pp. 958–964. DOI: <https://doi.org/10.1111/j.1745-6584.1992.tb01579.x>.
- Milinic, J. et al. (2015). “Sorption behaviour of perfluoroalkyl substances in soils”. In: *Science of The Total Environment* 511, pp. 63–71. ISSN: 0048-9697. DOI: <https://doi.org/10.1016/j.scitotenv.2014.12.017>.
- Molinari, J. and P. Peaudeceff (1977). “Peaudeceff, Essais conjoints en laboratoire et sur le terrain en vue d’une approche simplifiée de la prévision des propagations de substances miscibles dans les aquifères réels”. In: *Symposium on Hydrodynamic Diffusion and Dispersion in Porous Media*.
- Moltyaner, G. L. et al. (1993). “Numerical simulations of Twin Lake Natural-Gradient Tracer Tests: A comparison of methods”. In: *Water Resources Research* 29.10, pp. 3433–3452. DOI: <https://doi.org/10.1029/93WR01276>.
- Naymik, T. and M. Barcelona (1981). “Characterization of a contaminant plume in ground water, Meredosia, Illinois”. In: *Groundwater* 19.5, pp. 517–526.
- Nguyen, T. et al. (2020). “Influences of Chemical Properties, Soil Properties, and Solution pH on Soil–Water Partitioning Coefficients of Per- and Polyfluoroalkyl Substances (PFASs)”. In: *Environmental Science & Technology* 54.24. PMID: 33249833, pp. 15883–15892. DOI: 10.1021/acs.est.0c05705.
- OECD, Organisation for Economic Co-operation and Development (2015). “Working towards a global emission inventory of PFASs: Focus on PFCAs - Status quo and the way forward”. In: URL: <https://www.oecd.org/chemicalsafety/Working%20Towards%20a%20Global%20Emission%20Inventory%20of%20PFAS.pdf>.
- Palmer, C. and R. Nadon (1986). “A Radial Injection Tracer Experiment in a Confined Aquifer, Scarborough, Ontario, Canada”. In: *Groundwater* 24.3, pp. 322–331. DOI: <https://doi.org/10.1111/j.1745-6584.1986.tb01008.x>.
- Persson, J. and N. Andersson (2016). “Modeling groundwater flow and PFOS transport”. Master’s Thesis. KTH Royal institute of technology. URL: <https://www.diva-portal.org/smash/get/diva2:946450/FULLTEXT02.pdf>.

- Pettersson, K. (2020). “Groundwater Movement and PFAS Transportation in the Vreta-Bålsta Esker”. Master’s Thesis. University of Uppsala. URL: <http://diva-portal.org/smash/get/diva2:1484720/FULLTEXT01.pdf>.
- Phillips, F.M. and M.C. Castro (2014). “7.11 - Groundwater Dating and Residence-Time Measurements”. In: *Treatise on Geochemistry (Second Edition)*. Ed. by Heinrich D. Holland and Karl K. Turekian. Second Edition. Oxford: Elsevier, pp. 361–400. ISBN: 978-0-08-098300-4. DOI: <https://doi.org/10.1016/B978-0-08-095975-7.00513-1>.
- Pickens, J. and G. Grisak (1981). “Scale-dependent dispersion in a stratified granular aquifer”. In: *Water Resources Research* 17.4, pp. 1191–1211. DOI: <https://doi.org/10.1029/WR017i004p01191>.
- Pinder, G. (1973). “A Galerkin-finite element simulation of groundwater contamination on Long Island, New York”. In: *Water Resources Research* 9.6, pp. 1657–1669.
- Postigo, C. et al. (2018). “Groundwater Pollution: Sources, Mechanisms, and Prevention”. In: *Encyclopedia of the Anthropocene*. Ed. by Dominick A. Dellasala and Michael I. Goldstein. Oxford: Elsevier, pp. 87–96. ISBN: 978-0-12-813576-1. DOI: <https://doi.org/10.1016/B978-0-12-809665-9.09880-3>.
- Prommer, H. (2023). *Email to Clara Eklund*.
- Ptak, T. and G. Teutsch (1994). “Forced and natural gradient tracer tests in a highly heterogeneous porous aquifer: instrumentation and measurements”. In: *Journal of Hydrology* 159.1, pp. 79–104. ISSN: 0022-1694. DOI: [https://doi.org/10.1016/0022-1694\(94\)90250-X](https://doi.org/10.1016/0022-1694(94)90250-X).
- Rajaram, H and L. Gelhar (1991). “Three-dimensional spatial moments analysis of the Borden Tracer Test”. In: *Water Resources Research* 27.6, pp. 1239–1251. DOI: <https://doi.org/10.1029/91WR00326>.
- Rapp, B. (2017a). “Chapter 11 - Conservation of Momentum: The Navier-Stokes Equation”. In: *Microfluidics: Modelling, Mechanics and Mathematics*. Ed. by Bastian E. Rapp. Micro and Nano Technologies. Oxford: Elsevier, pp. 273–289. ISBN: 978-1-4557-3141-1. DOI: <https://doi.org/10.1016/B978-1-4557-3141-1.50011-3>.
- (2017b). “Chapter 9 - Fluids”. In: *Microfluidics: Modelling, Mechanics and Mathematics*. Ed. by Bastian E. Rapp. Micro and Nano Technologies. Oxford: Elsevier, pp. 243–263. ISBN: 978-1-4557-3141-1. DOI: <https://doi.org/10.1016/B978-1-4557-3141-1.50009-5>.
- Raschke, A., A. Pouyan Nejadhashemi, and V. Rafiei (2022). “Overview of Modeling, Applications, and Knowledge Gaps for Integrated Large-Scale PFAS Modeling”. In: *Journal of Environmental Engineering* 148.9, p. 03122003. DOI: 10.1061/(ASCE)EE.1943-7870.0002033.
- Rayne, S. and K. Forest (2009). “Perfluoroalkyl sulfonic and carboxylic acids: A critical review of physicochemical properties, levels and patterns in waters and wastewaters, and treatment methods”. In: *Journal of Environmental Science and Health, Part A* 44.12, pp. 1145–1199. DOI: 10.1080/10934520903139811.
- Rayne, S., K. Forest, and K. Friesen (2009). “Estimated congener specific gas-phase atmospheric behavior and fractionation of perfluoroalkyl compounds: Rates of reaction with atmospheric oxidants, air-water partitioning, and wet/dry deposition lifetimes”. In: *Journal of Environmental Science and Health, Part A* 44.10, pp. 936–954. DOI: 10.1080/10934520902996815.

- Rodhe, A. et al. (2006). *Grundvattenbildning i svenska typjordar*. University of Uppsala. URL: https://www.sgu.se/globalassets/grundvatten/grundvattennivaer-old/grundvattenbildning/rodhe-et-al_2006.pdf.
- Sahu, O. and N. Singh (2019). “13 - Significance of bioadsorption process on textile industry wastewater”. In: *The Impact and Prospects of Green Chemistry for Textile Technology*. Ed. by Shahid-ul-Islam and B.S. Butola. The Textile Institute Book Series. Woodhead Publishing, pp. 367–416. ISBN: 978-0-08-102491-1. DOI: <https://doi.org/10.1016/B978-0-08-102491-1.00013-7>.
- Saini, A. (Apr. 2023). *Decision tree algorithm - A complete guide*. URL: <https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/>.
- Schulze-Makuch, D. (2005). “Longitudinal dispersivity data and implications for scaling behavior”. In: *Groundwater* 43.3, pp. 443–456. DOI: <https://doi.org/10.1111/j.1745-6584.2005.0051.x>.
- SGU, Geological survey of Sweden (2021). *PFAS i grundvatten*. <https://www.sgu.se/om-sgu/verksamhet/miljoovervakning/pfas-i-grundvatten/> [Accessed: 13-02-2023)].
- Silliman, S. E. and E. S. Simpson (1987). “Laboratory evidence of the scale effect in dispersion of solutes in porous media”. In: *Water Resources Research* 23.8, pp. 1667–1673. DOI: <https://doi.org/10.1029/WR023i008p01667>.
- Silva, J., J. Guelfo, et al. (2022). “Simulated leaching of PFAS from land-applied municipal biosolids at agricultural sites”. In: *Journal of Contaminant Hydrology* 251, p. 104089. ISSN: 0169-7722. DOI: <https://doi.org/10.1016/j.jconhyd.2022.104089>.
- Silva, J., J. Šimůnek, and J. McCray (2020). “A Modified HYDRUS Model for Simulating PFAS Transport in the Vadose Zone”. In: *Water* 12.10, pp. 27–58. ISSN: 2073-4441. DOI: [10.3390/w12102758](https://doi.org/10.3390/w12102758).
- Sima, M. and P. Jaffé (2021). “A critical review of modeling Poly- and Perfluoroalkyl Substances (PFAS) in the soil-water environment”. In: *Science of The Total Environment* 757, p. 143793. ISSN: 0048-9697. DOI: <https://doi.org/10.1016/j.scitotenv.2020.143793>.
- Sparks, D. (2003). “5 - Sorption Phenomena on Soils”. In: *Environmental Soil Chemistry (Second Edition)*. Ed. by Donald L. Sparks. Second Edition. Burlington: Academic Press, pp. 133–186. ISBN: 978-0-12-656446-4. DOI: <https://doi.org/10.1016/B978-012656446-4/50005-0>.
- Sruthi, E. R. (Apr. 2023). *Understand random forest algorithms with examples (updated 2023)*. URL: <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>.
- Sudicky, E.A., J.A. Cherry, and E.O. Frind (1983). “Migration of contaminants in groundwater at a landfill: A case study: 4. A natural-gradient dispersion test”. In: *Journal of Hydrology* 63.1. Migration of Contaminants in Groundwater at a Landfill: A Case Study, pp. 81–108. ISSN: 0022-1694. DOI: [https://doi.org/10.1016/0022-1694\(83\)90224-X](https://doi.org/10.1016/0022-1694(83)90224-X).
- Sunderland, E. et al. (2019). “A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects”. In: *Journal of Exposure Science Environmental Epidemiology* 29.2, pp. 131–147. ISSN: 1559-064X. DOI: <https://doi.org/10.1038/s41370-018-0094-1>.

- Swedish Chemical agency (2021). *Kunskapssammanställning om PFAS*. <https://www.kemi.se/download/18.3f6f225517c0af779871bc0/1632907246253/PM-1-21-Kunskapssammanst%C3%A4llning-om-PFAS.pdf> [Accessed: 13-02-2023)].
- Swedish EPA (2023). *Högfluorerade ämnen i miljön, PFAS*. <https://www.naturvardsverket.se/amnesomraden/miljofororeningar/organiska-miljogifter/hogfluorerade-amnen-i-miljon-pfas/> [Accessed: 13-02-2023)].
- Sykes, JF et al. (1983). *The Validation of SWENT, a Geosphere Transport Model; W: Scientific Computing, red. R. Stapleman i in; IMAES*.
- Trefry, M. and C. Muffels (2007). “FEFLOW: A Finite-Element Ground Water Flow and Transport Modeling Tool”. In: *Groundwater* 45.5, pp. 525–528. DOI: <https://doi.org/10.1111/j.1745-6584.2007.00358.x>.
- USEPA (2014). “Emerging Contaminants – Perfluorooctane Sulfonate (PFOS) and Perfluorooctanoic Acid (PFOA), Emerging Contaminants Fact Sheet – PFOS and PFOA”. In: U.S. Environmental Protection Agency (EPA). URL: <https://www.epa.gov/work/HQ/100002767.pdf> [Accessed: 2016-02-2023)].
- (2023). *CompTox Chemicals Dashboard*. <https://www.epa.gov/chemical-research/comptox-chemicals-dashboard> [Last Modified 3-02-2023].
- Vaccaro, JJ and EL Bolke (1983). *Evaluation of water-quality characteristics of part of the Spokane aquifer, Washington and Idaho, using a solute-transport digital model*. Tech. rep. US Geological Survey,
- Valocchi, A. et al. (1981). “Simulation of the transport of ion-exchanging solutes using laboratory-determined chemical parameter values”. In: *Groundwater* 19.6, pp. 600–607.
- Wallis, I. et al. (2022). “Model-based identification of vadose zone controls on PFAS mobility under semi-arid climate conditions”. In: *Water Research* 225, p. 119096. ISSN: 0043-1354. DOI: <https://doi.org/10.1016/j.watres.2022.119096>.
- Wang, Z. et al. (2011). “Using COSMOtherm to predict physicochemical properties of poly- and perfluorinated alkyl substances (PFASs)”. In: *Environmental Chemistry* 8, pp. 389–398. DOI: <https://doi.org/10.1071/EN10143>.
- Wendland, E. and G. Schmid (2000). “A Symmetrical Streamline Stabilization scheme for high advective transport”. In: *International Journal for Numerical and Analytical Methods in Geomechanics* 24.1, pp. 29–45. DOI: [https://doi.org/10.1002/\(SICI\)1096-9853\(200001\)24:1<29::AID-NAG51>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1096-9853(200001)24:1<29::AID-NAG51>3.0.CO;2-F).
- Wiebenga, WA et al. (1967). “Radioisotopes as groundwater tracers”. In: *Journal of Geophysical Research* 72.16, pp. 4081–4091.
- Wikipedia (2022). *Mean squared error* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 29-May-2023]. URL: https://en.wikipedia.org/w/index.php?title=Mean_squared_error&oldid=1127519968.
- (2023a). *Root-mean-square deviation* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 29-May-2023]. URL: https://en.wikipedia.org/w/index.php?title=Root-mean-square_deviation&oldid=1145735789.
- (2023b). *Upwind scheme* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 5-August-2023]. URL: https://en.wikipedia.org/w/index.php?title=Upwind_scheme&oldid=1153053273.

- Xiao, F. (2017). “Emerging poly- and perfluoroalkyl substances in the aquatic environment: A review of current literature”. In: *Water Research* 124, pp. 482–495. ISSN: 0043-1354. DOI: <https://doi.org/10.1016/j.watres.2017.07.024>.
- Xiao, F. et al. (2022). “Per- and Polyfluoroalkyl Substances (PFAS) in Subsurface Environments: Occurrence, Fate, Transport, and Research Prospect”. In: *Reviews of Geophysics* 60.3. DOI: <https://doi.org/10.1029/2021RG000765>.
- Xu, J., Z. Liu, et al. (2020). “Enhanced adsorption of perfluorooctanoic acid (PFOA) from water by granular activated carbon supported magnetite nanoparticles”. In: *Science of The Total Environment* 723, p. 137757. ISSN: 0048-9697. DOI: <https://doi.org/10.1016/j.scitotenv.2020.137757>.
- Xu, M. and Y. Eckstein (1995). “Use of Weighted Least-Squares Method in Evaluation of the Relationship Between Dispersivity and Field Scale”. In: *Groundwater* 33.6, pp. 905–908. DOI: <https://doi.org/10.1111/j.1745-6584.1995.tb00035.x>.
- Yu, C., J.J. Cheng, et al. (2023). “Data collection handbook to support modeling the impacts of radioactive material in soil”. In: DOI: 10.2172/10162250.
- Yu, C., C. Loureiro, et al. (1993). *Data collection handbook to support modeling impacts of radioactive materials in soil*. Argonne National Laboratory, Argonne, Illinois.
- Zhang, D., Y. Liang, and W. Zhang (2019). “Adsorption of perfluoroalkyl and polyfluoroalkyl substances (PFASs) from aqueous solution - A review”. In: *Science of The Total Environment* 694, p. 133606. ISSN: 0048-9697. DOI: <https://doi.org/10.1016/j.scitotenv.2019.133606>.
- Zhang, X., T. Brown, et al. (2010). “Assessment of chemical screening outcomes based on different partitioning property estimation methods”. In: *Environment International* 36.6, pp. 514–520. ISSN: 0160-4120. DOI: <https://doi.org/10.1016/j.envint.2010.03.010>.
- Zhang, Z., Y. Xiao, et al. (June 2017). “Sustainable Application of Unbound Quarry Byproducts in Pavement Construction through Aggregate Gradation Optimization”. In: DOI: 10.12783/dtmse/ictim2017/10187.

Appendix A - Henry's sorption coefficient

Henry's sorption coefficient values found in the literature for PFOS can be seen in Table A. These values were the base for the values used in the simulation, where the 10th and 90th were used as the input interval.

Table A: Henry's sorption coefficient for PFOS collected from several different sources. Density used for calculating K_h from K_d is obtained from ¹ Yu, Loureiro, et al., 1993 and ² Zhang, Xiao, et al., 2017, see table 6.

Sorption coefficient for PFOS			
K_d (L/kg)	K_H (-)	Comment	References
4.40E-01	6.78E-01	<i>Sand</i> ¹	Edvinsson, 2015
	9.24E-01	<i>Sand and Gravel</i> ²	
1.50E+01	2.31E+01	<i>Sand</i> ¹	Boonraksasat, 2019; Edvinsson, 2015
	3.15E+01	<i>Sand and Gravel</i> ²	
1.70E+01	2.62E+01	<i>Sand</i> ¹	Persson and Andersson, 2016
	3.57E+01	<i>Sand and Gravel</i> ²	
1.56E+01	2.25E+01	<i>Loam</i>	Silva, Šimunek, et al., 2020
2.62E+00	3.77E+00	<i>Loamy sand</i>	
	1.90E+03		Kim et al., 2015
	4.00E-03		Arp et al., 2006
	9.00E-02		
	4.50E-01		
	5.00E-02		Zhang, Brown, et al., 2010
	2.00E-03		
	4.00E-02		
	7.60E-10		
	2.00E-02		USEPA, 2023
	1.70E-01		Wang et al., 2011
	4.00E-02		Xiao, 2017
	4.50E-01		Rayne and Forest, 2009
	6.10E-06		Rayne, Forest, and Friesen, 2009

Henry's sorption coefficient values found in the literature for PFOA can be seen in Table B. These values were the base for the values used in the simulation, where the 10th and 90th were used as the input interval.

Table B: *Henry's sorption coefficient for PFOA collected from several different sources. Density used for calculating K_h from K_d is obtained from* ¹ *Yu, Loureiro, et al., 1993 and* ² *Zhang, Xiao, et al., 2017, see table 6.*

Sorption coefficient for PFOA			
K_d (L/kg)	K_H (-)	Comment	References
1.10E+00	1.69E+00	<i>Sand</i> ¹	Boonraksasat, 2019
	2.31E+00	<i>Sand and Gravel</i> ²	
1.08E-02	1.66E-02	<i>Sand</i> ¹	Gefell et al., 2022
	2.27E-02	<i>Sand and Gravel</i> ²	
1.08E-01	1.66E-01	<i>Sand</i> ¹	Gefell et al., 2022
	2.27E-01	<i>Sand and Gravel</i> ²	
1.99E+00	2.87E+00	Loam	Silva, Šimůnek, et al., 2020
5.70E-01	8.21E-01	Loamy sand	
	3.70E+02		Kim et al., 2015
	1.00E-03		Li, Ellis, et al., 2007
	4.30E-03		Arp et al., 2006
	2.00E-02		
	3.70E+00		Zhang, Brown, et al., 2010
	4.00E-02		
	3.00E-02		
	3.60E-01		
	8.20E-03		Kutsuna and Hori, 2008
	8.30E-09		USEPA, 2023
	1.00E-02		Wang et al., 2011
	1.00E-01		Xiao, 2017

Appendix B - Longitudinal dispersivity

Table C shows the value for the longitudinal dispersivity in sand found in the literature. Table C also shows the belonging scale and the calculated values of the dispersivity based on the scale and Equation 9.

Table C: Values for the longitudinal dispersivity in sand found in the literature, together with the belonging scale. $\alpha(eq)$ are the values obtained by inserting the scale together into equation (9).

Longitudinal dispersivity - Sand			
Scale (m)	Alpha (m)	Alpha (eq.)	Source
0.150	0.0580	-	
0.460	0.0470	-	
0.910	0.0760	-	
1.37	0.0940	0.00681	
1.83	0.0930	0.0329	Silliman and Simpson, 1987
0.150	0.0540	-	
0.460	0.0710	-	
0.910	0.0730	-	
1.37	0.0750	0.00681	
1.83	0.127	0.0329	
125	0.450	4.95874	Jensen et al., 1993
5.00	0.100	0.3496	
5.00	0.250	0.34962	Mas-Pla et al., 1992
5.00	0.050	0.3496	
5.11	0.240	0.36114	
25.8	0.190	1.9076	
71.5	0.190	3.68526	
80.0	0.340	3.9237	
94.8	0.340	4.30067	Moltyaner et al., 1993
93.4	0.210	4.2668	
102	0.230	4.46038	
116	0.510	4.7797	
266	0.550	7.04241	
6.50	0.0700	0.5034	Palmer and Nadon, 1986

Table D shows the value for the longitudinal dispersivity in sand and gravel found in the literature. Table D also shows the belonging scale and the calculated values of the dispersivity based on the scale and Equation 9.

Table D: *Values for the longitudinal dispersivity in sand and gravel found in the literature, together with the belonging scale. Alpha(eq) are the values obtained by inserting the scale together into equation (9).*

Longitudinal dispersivity - Sand and Gravel			
Scale (m)	Alpha (m)	Alpha (eq.)	Source
12.8	1.50	1.06	
8.90	0.710	0.732	
9.90	1.33	0.821	
12.4	6.61	1.03	
12.2	2.00	1.01	
8.93	1.38	0.735	
17.1	2.50	1.37	
9.90	1.02	0.821	
17.1	0.600	1.37	
120	0.600	4.87	
52.2	11.0	3.06	Ptak and Teutsch, 1994
32.7	4.35	2.26	
34.1	3.99	2.33	
41.0	3.83	2.63	
56.4	5.52	3.21	
44.5	10.6	2.77	
32.7	5.00	2.26	
34.1	4.00	2.33	
41.0	5.92	2.63	
56.4	6.00	3.21	
44.5	7.09	2.77	
350	10.0	7.91	Chiang et al., 1989
26.6	1.50	1.95	
28.6	1.50	2.06	
44.2	2.20	2.76	
30.0	4.00	2.13	
58.0	4.00	3.26	
26.0	1.00	1.92	Boesel et al., 2000
63.1	6.00	3.43	
43.3	5.00	2.73	
80.0	5.00	3.92	
91.5	3.50	4.22	
234	30.0	6.66	
223	30.0	6.52	
200	7.50	6.21	Adams and Gelhar, 1992

Appendix C - Transversal dispersivity and hydraulic conductivity

Table E shows the values for transverse dispersivity and hydraulic conductivity in sand obtained from the literature. Table E also presents the ratio between transverse and longitudinal dispersivity.

Table E: Values for the transversal dispersivity, the ration between D_T and D_L as well as the conductivity in sand. All the values were obtained during the literature review.

Transversal dispersivity and Conductivity Sand			
D_T (m)	D_T/D_L	Conductivity (m/s)	Source
		1.00E-04	Egboka et al., 1983
		1.00E-06	
0.0390	0.0907	7.20E-05	Freyberg, 1986
0.760	0.475	9.55E-05	Kies, 1981
		3.10E-05	Kreft et al., 1974
		1.50E-04	
		3.20E-05	Lee et al., 1980
0.110	0.0567		Molinari and Peaudeceff, 1977
0.110	0.0403		
		2.00E-05	Pickens and Grisak, 1981
		2.00E-04	
0.0500	0.100	7.20E-05	Rajaram and Gelhar, 1991
0.00500	0.500	4.80E-05	Sudicky et al., 1983
0.0300	0.375	7.60E-05	
		5.80E-05	Sykes et al., 1983
		7.20E-05	

Table F shows the values for transverse dispersivity and hydraulic conductivity in sand obtained from the literature. Table F also presents the ratio between transverse and longitudinal dispersivity.

Table F: *Values for the transversal dispersivity, the ration between D_T and D_L as well as the conductivity in sand. All the values were obtained during the literature review.*

Transversal dispersivity and Conductivity Sand and Gravel			
D_T (m)	D_T / D_L	Conductivity (m/s)	Source
		1.00E-04	Adams and Gelhar, 1992
		1.00E-02	
18.3	0.600	5.70E-04	Ahlstrom et al., 1977
		3.00E-02	
0.0180	0.0188	1.30E-03	Garabedian et al., 1988
		9.20E-04	Hoehn, 1983
		6.60E-03	
		8.10E-05	Hoehn and Santschi, 1987
		6.60E-03	
1.50	0.500		Iris, 1980
		2.40E-04	Konikow and Bredehoeft, 1974
9.10	0.298	1.00E-02	
		9.00E-04	Lau and Todd, 1957
0.610	0.286		Naymik and Barcelona, 1981
0.915	0.273		
4.20	0.197	7.50E-04	Pinder, 1973
27.4	0.300		Vaccaro and Bolke, 1983
0.100	0.100		Valocchi et al., 1981
		5.50E-03	Wiebenga et al., 1967

Appendix D - Shapley values

Figure A presents the Shapley values obtained for the six different parameters and at the four different scenarios in observation point 2.

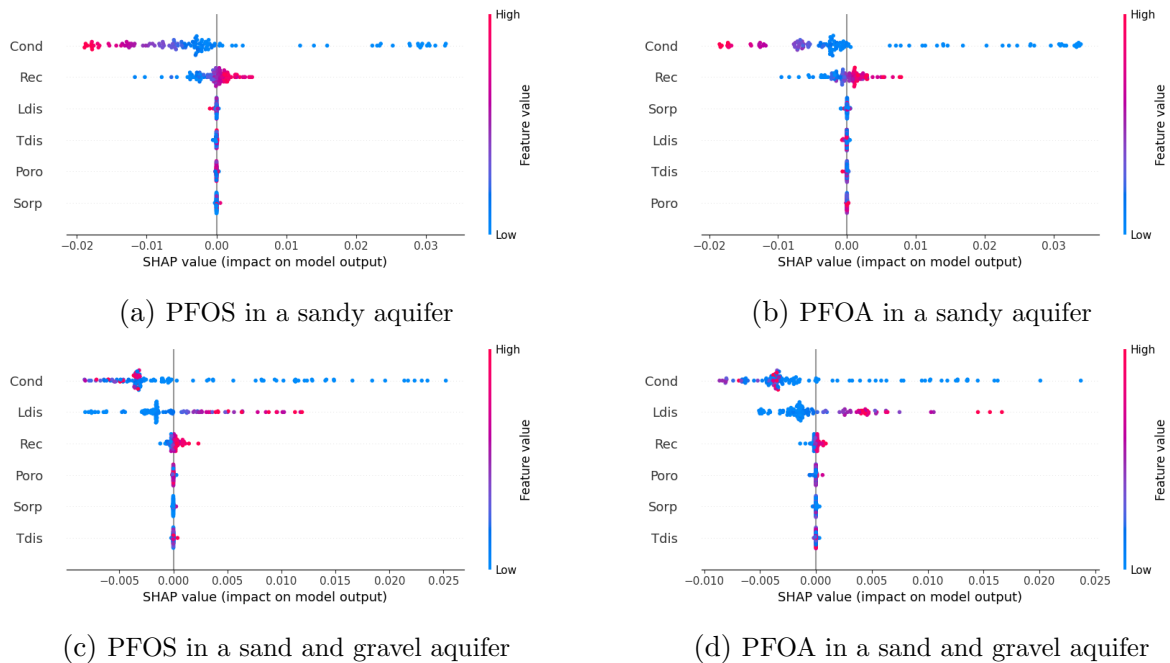


Figure A: *The Shapley values for the four different scenarios in Observation point 2. Blue color indicates low values of the parameters, while the pink color indicates high values (as the scale to the right present).*

Figure B presents the Shapley values obtained for the six different parameters and at the four different scenarios in observation point 3.

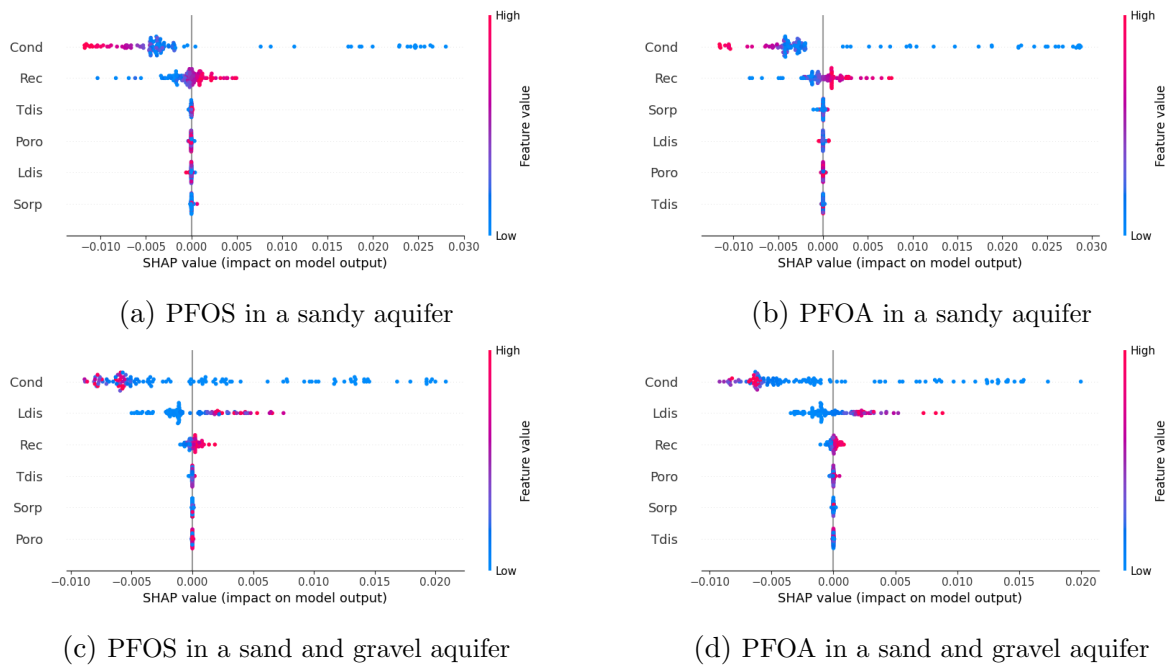


Figure B: *The Shapley values for the four different scenarios in Observation point 3. Blue color indicates low values of the parameters, while the pink color indicates high values (as the scale to the right present).*

Appendix E - Random forest importance

Figure C shows the random forest importance values obtained for the five different models created in observation point 2.

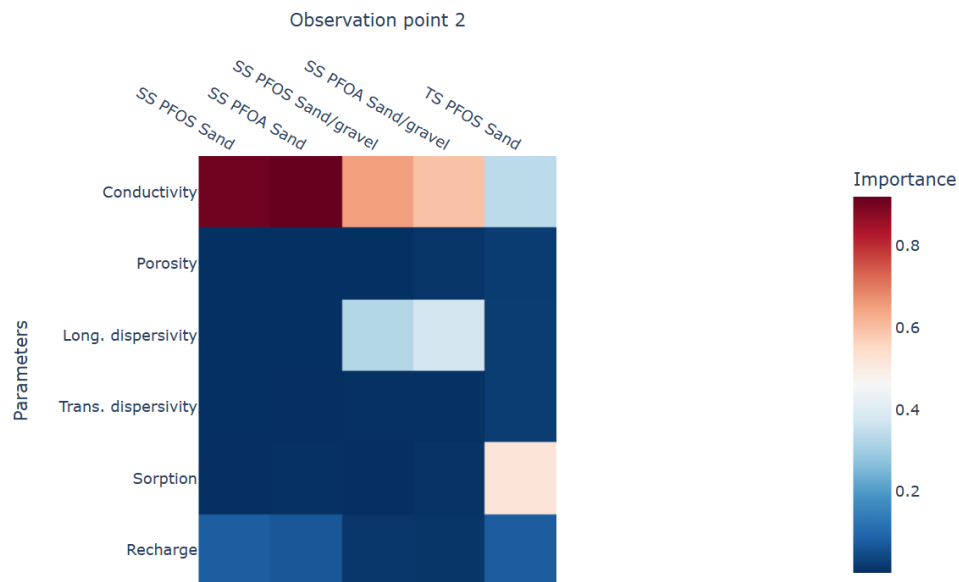


Figure C: *The heat-map for the random forest importance values obtained for the five different models in observation point 2.*