# APPLICATION OF MULTIVARIATE STATISTICAL METHODS FOR PFAS SOURCE TRACKING IN THE RIVER VANTAA WATERSHED, SOUTHERN FINLAND

Lappeenranta–Lahti University of Technology LUT

Master's programme in Water Technology

Master's Thesis

2025

Heli Kuru

Examiners: Professor Amit Bhatnagar, PhD

Dr. Marie-Amélie Pétré, Senior Scientist PhD

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Engineering Sciences

Chemical Engineering - Water Technology

Heli Kuru

**Application of multivariate statistical methods for PFAS source tracking in the River Vantaa watershed, Southern Finland**

Master's thesis in Chemical Engineering

2025

86 pages, 12 figures, 8 tables

Examiners: Professor Amit Bhatnagar, PhD and Dr. Marie-Amélie Pétré, PhD, Senior Scientist

Keywords: PFAS source tracking, Contaminant tracing, Environmental forensics, River Vantaa watershed, Per- and polyfluoroalkyl substances, Water contamination, Principal Component Analysis, Hierarchical Clustering on Principal Components

Per- and polyfluoroalkyl substances (PFAS) are persistent environmental contaminants widely detected in water bodies. This thesis investigates PFAS contamination in the River Vantaa watershed, focusing on source tracking using multivariate statistical methods and geospatial analysis. A total of 99 water sample profiles from 58 locations collected between 2014-2024 were used in the statistical analyses. Sample matrices included river, ditch, creek, groundwater, stormwater and influent/effluent from wastewater treatment plants.

PFAS profiles were grouped using Principal Component Analysis (PCA) and Hierarchical Clustering on Principal Components (HCPC), resulting in five clusters that were compared to PFAS source profiles described in the literature. Nine PFAS compounds, common to all datasets, were included in the analysis. Cluster 1 represented diffuse sources and background levels. Clusters 2 and 3 showed profiles typical of wastewater treatment plants, with Cluster 2 also linked to urban runoff. Cluster 4 indicated historical AFFF contamination. Cluster 5 showed the highest levels of all individual PFAS, and the chemical profile reflected contamination from airport firefighting areas, industrial activity and landfill leachate.

The results illustrate challenges in PFAS source tracking due to overlapping contamination signals and limited compound range. Overall, combining multivariate statistical methods and spatial approaches revealed distinct PFAS profiles and spatial trends, demonstrating the potential of integrated methods for future source tracking and monitoring efforts.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUTin insinööritieteiden tiedekunta

Kemiantekniikka

Heli Kuru

**Monimuuttujamenetelmien soveltaminen PFAS-lähteiden jäljitykseen Vantaanjoen valuma-alueella**

Kemiantekniikan diplomityö

2025

86 sivua, 12 kuvaa, 8 taulukkoa

Tarkastajat: Professori Amit Bhatnagar ja Marie-Amélie Pétré, PhD, vanhempi tutkija

Avainsanat: PFAS-lähteiden jäljitys, haitta-aineiden seuranta, Vantaanjoen valuma-alue, per- ja polyfluorialkyyliyhdisteet, lähde-erottelu, vesistöjen pilaantuminen, pääkomponenttianalyysi, hierarkkinen klusterointi pääkomponenttien perusteella, ympäristötutkimus

Per- ja polyfluoratut alkyyliyhdisteet (PFAS) ovat pysyviä ympäristömyrkkyjä, joita esiintyy laajasti Suomen vesistöissä. Tämä tutkielma tarkastelee PFAS-saastumista Vantaanjoen valuma-alueella, keskittyen lähteiden jäljittämiseen monimuuttujamenetelmien ja paikkatietoanalyysin avulla. Tilastollisessa tutkimuksessa käytettiin 99 näyteprofiilia 58 näytepisteeltä vuosilta 2014–2024. Näytematriiseina olivat joki-, oja-, pohjavesi- ja hulevesinäytteet sekä jätevedenpuhdistamon tulo- ja poistovedet. Analysoituja PFAS yhdisteitä oli yhdeksän.

Työssä tilastollisesti samankaltaiset profiilit ryhmiteltiin pääkomponenttianalyysin (PCA) ja siihen perustuvan hierarkkisen klusteroinnin (HCPC) avulla. Tuloksena saatiin viisi erilaista klusteria, joiden profiileja verrattiin kirjallisuudessa kuvattuihin tyypillisiin PFAS lähteisiin. Klusteri 1 edusti hajalähteitä ja taustapitoisuuksia. Klusterit 2 ja 3 osoittivat jätevedenpuhdistamoille tyypillisiä profiileja, mutta klusterin 2 profiilin vaikuttivat myös esim. hulevesivalunta sekä muovi- ja metalliteollisuus. Klusterin 4 korkea PFOS ja PFHxS pitoisuus viittasivat historiallisiin AFFF-saastuneisiin alueisiin. Myös klusterin 5 todennäköinen lähde oli lentokentän paloharjoitusalueet, minkä lisäksi sormenjälki indikoi metalli- ja kemianteollisuuden sekä kaatopaikan suotovesien yhteisvaikutusta.

Tulokset osoittavat, että PFAS-lähteiden jäljittämisessä on haasteita, jotka johtuvat päällekkäisistä saastumissignaaleista ja rajoitetusta yhdistevalikoimasta. Kokonaisuutena tutkimus tukee tilastollisten ja spatiaalisten menetelmien potentiaalia PFAS-lähteiden jäljityksessä ja ympäristötutkimuksessa.

SYMBOLS AND ABBREVIATIONS

Creek characters

| | | |
|---|---|---|
| ρ | Spearman rank-order correlation coefficient | [-] |

Abbreviations

| | |
|---|---|
| AFFF | Aqueous Film-Forming Foam |
| CLR | Centred Logarithmic Ratio |
| CSM | Conceptual Site Model |
| DWD | Drinking Water Directive |
| ECF | Electrochemical Fluorination |
| ECHA | European Chemicals Agency |
| EFSA | European Food Safety Authority |
| EOF | Extractable Organofluorine |
| EQS | Environmental Quality Standard |
| FPP | Forever Pollution Project |
| FT | Fluorotelomer |
| FTOH | Fluorotelomer Alcohol |
| FTS | Fluorotelomer Sulfonate |
| GIS | Geographical Information System |
| GTK | Geologian Tutkimuskeskus (Geological Survey of Finland) |
| GW | Groundwater |
| HCA | Hierarchical Clustering |
| HCPC | Hierarchical Clustering on Principal Components |
| HRMS | High Resolution Mass Spectrometry |
| HWI | Human Waste Indicator |
| LC-MS | Liquid Chromatography with Mass Spectrometry |
| LOE | Lines of Evidence |
| LULC | Land Use and Land Cover |
| MCMC | Markov Chain Monte Carlo |
| MDL | Method Detection Limit |

| | |
|---|---|
| MLR | Multiple Linear Regression |
| OECD | Organisation for Economic Co-operation and Development |
| PC | Principal Component |
| PCA | Principal Component Analysis |
| PFAA | Perfluoroalkyl acids |
| PFAS | Per- and polyfluoroalkyl substances |
| PFBA | Perfluorobutanoic acid |
| PFBS | Perfluorobutanesulfonic acid |
| PFCA | Perfluorocarboxylic acids |
| PFHpA | Perfluoroheptanoic acid |
| PFHxA | Perfluorohexanoic acid |
| PFHxS | Perfluorohexanesulfonic acid |
| PFNA | Perfluorononanoic acid |
| PFOA | Perfluorooctanoic acid |
| PFOS | Perfluorooctanesulfonic acid |
| PFPeA | Perfluoropentanoic acid |
| PFSA | Perfluorosulfonic acids |
| PMF | Positive Matrix Factorization |
| POP | Persistent Organic Pollutants |
| REACH | Registration, Evaluation, Authorisation and Restriction of Chemicals |
| RFE | Recursive Feature Elimination |
| SVHC | Substance of Very High Concern |
| SYKE | Suomen Ympäristökeskus (Finnish Environment Institute) |
| TFA | Trifluoroacetic acid |
| TOP | Total Oxidizable Precursor |
| TWI | Tolerable Weekly Intake |
| UPGMA | Unweighted Pair Group Method with Arithmetic Mean |
| VHVSY | Vantaanjoen ja Helsingin Seudun Vesiensuojeluyhdistys (Water protection association of Vantaa River and Helsinki region) |
| WFD | Water Framework Directive |
| WWTP | Wastewater Treatment Plant |

**Table of contents**

# DECLARATIONS

## Turnitin

The originality of this thesis has been reviewed with the Turnitin similarity checking service.

## AI usage

The author of the thesis, Heli Kuru, used the following AI-tools during the preparation of the thesis:

1. Microsoft 365 Copilot
    a. Purpose of use: Brainstorming, reviewing literature, assistance in drafting the literature review structure
    b. Explanation of the use of the tool: summarizing literature review and organizing key themes during the writing stage
2. ChatGPT
    a. Purpose of use: Help understanding error reports from R (version 4.4.1), understand main points from literature review articles
    b. Explanation of the use of the tool: Discussion about error reports from R, summarizing articles for quicker understanding
3. NotebookLM
    a. Purpose of use: Summarizing and synthesizing research materials to support organizing the literature review and discussion sections
    b. Explanation of the use of the tool: Generate summaries and identify connections between studies

## Responsibility

The author, Heli Kuru, takes full responsibility for the content of this thesis and has reviewed and edited the content generated by the possible use of AI tools.

# 1 Introduction

Per- and polyfluoroalkyl substances (PFAS) are a group of synthetic chemicals widely used in industrial and consumer applications due to their water- and grease-resistant properties (OECD, 2021). However, their persistence and potential toxicity raise concerns in environmental science and public health (OECD, 2018). PFAS contamination is now recognized as a global concern because of widespread detection in surface waters, groundwater, soil and wildlife (Lyu et al., 2022). This thesis investigates PFAS contamination in the River Vantaa watershed. The focus is on source tracking using multivariate statistical methods and geospatial mapping. River Vantaa is recognized as PFAS hotspot in Finland (Junttila et al., 2019). It is essential to understand PFAS sources in this region for effective monitoring, mitigation efforts and risk assessment. Watershed is an important recreational area, and its water is used for municipal water supply as well (Kivimäki et al., 2013). Public awareness is increasing, and EU regulations are tightening regarding PFAS contamination (Council of the EU, 2025).

Previous studies have shown that PFAS contamination can originate from a variety of sources, including industrial discharges, wastewater treatment plants (WWTPs), landfills and the use of aqueous film-forming foams (AFFFs) at airports and military training sites (Zhang et al., 2016; Woodward et al., 2024). PFAS can also enter the environment through biosolids, atmospheric deposition, and septic systems (Silver et al., 2023). To date, most studies have focused on identifying specific point sources or characterizing contamination events rather than cases where multiple inputs complicate attribution to a single origin (Benotti et al., 2020; Charbonnet et al., 2021).

In the River Vantaa watershed research focus has mainly been on quantifying PFAS contamination levels in water (Siimes et al., 2019; Junttila et al., 2021). However advanced source tracking methods have not been utilized; to the best of our knowledge, multivariate statistical tools have not been used to track the origin of the PFAS in this watershed. Source attribution has relied on comparing chemical profiles. While valuable, this approach may not capture the complexity of PFAS transport and transformation in watershed scale environment. Multiple overlapping inputs upstream and downstream from long time period contribute to challenges of source tracking. This gap highlights the need for more

sophisticated forensic approaches capable of differentiating between overlapping PFAS signatures. Multivariate statistical methods, such as Hierarchical Clustering on Principal Components (HCPC), offer promising tools for this purpose, especially when combined with geospatial context and chemical profiling.

This study aims to trace PFAS sources in the River Vantaa watershed by analysing surface water, groundwater, stormwater and WWTP samples collected over a ten-year period (8 July 2014 – 8 March 2024). The dataset combines publicly available monitoring data from the Water Protection Association of Vantaa River and Helsinki region (VHVSY) (Junttila et al., 2021) and Forever Pollution Project (FPP) (FPP, 2023) with internal data from the Geological Survey of Finland (GTK). The study area included 58 sampling locations across the watershed. Ninety-nine individual sample profiles were analysed with a focus on identifying characteristic PFAS profiles and their spatial associations with known potential sources.

The primary research questions this thesis aims to address:

*What are the key patterns and trends in PFAS concentrations across the watershed?*

*What are the main components contributing to the variance in PFAS concentrations in the study area? How do these components correlate with known potential PFAS sources?*

*What are the most significant sources of PFAS contamination in the study area?*

*How do PFAS concentrations vary across different locations within the study area?*

*What are the potential pathways for PFAS transport in the watershed?*

While this study provides important findings, it is limited by the number of PFAS compounds included in the analysis. Only nine substances met the detection frequency threshold across datasets. This largely restricts the depth of source tracking, particularly for emerging and volatile PFAS. Spatial coverage of sampling locations is also uneven, with underrepresentation in the western part of the watershed and limited sampling at major stream junctions. These limitations are discussed further in the discussion chapter.

The study revealed key trends in PFAS concentrations and identified dominant components contributing to variance. Multivariate analysis separated PFAS profiles into five clusters, which were linked to known source categories. The results show influence from potential AFFF use near Helsinki-Vantaa Airport, wastewater treatment plants, metal industry and landfill leachate. Background contamination from diffuse sources was also identified. These findings support the use of unsupervised statistical methods in environmental source tracking and align with previous studies that identified airports, WWTPs, and landfills as major PFAS contributors (Silver et al., 2023; Hu et al., 2016; Morethe et al., 2023). However, overlapping contamination signals and the absence of emerging and sentinel compounds limit the precision of source tracking.

A literature review was conducted to understand current PFAS source tracking methods. It illustrated the role of statistical techniques in identifying contamination patterns and grouping samples. Geospatial and modelling approaches complement statistical tools by linking PFAS to land use and simulating transport (Schaider et al., 2016; Joseph et al., 2023). Together, these methods form a multidisciplinary framework for source attribution. Time-based inclusion criterion is used in the literature review of PFAS source tracking to ensure the relevancy and recency of the data. Source tracking publications from 2015 onward are selected.

In this thesis, chapter 2 provides a literature review of PFAS substances, environmental behaviour, and regulatory framework. Chapter 3 summarizes the main source tracking methods, including statistical, geospatial, and modelling approaches. Chapter 4 outlines the data sources, data processing steps, and analytical methods used. Chapter 5 presents the results of the multivariate analysis and cluster profiles. Chapter 6 discusses the findings in relation to previous research and mentions the limitations of this study. Finally, Chapter 7 concludes with a summary of core findings, limitations, and future research directions.

Funding source

The work is done in collaboration with the Geological Survey of Finland (GTK). This thesis is part of the larger research project EMERGING-FLOW, funded by the Research Council of Finland (decision number 361919). The project investigates per- and polyfluoroalkyl substances, which are widespread in the Finnish aquatic environment.

# 2 Per- and Polyfluoroalkyl Substances

This chapter provides an overview of per- and polyfluoroalkyl substances (PFAS), highlighting their widespread use, environmental persistence and health impacts. It reviews PFAS classification, sources and transport mechanisms that facilitate their movement in the aquatic environment.

## 2.1 Overview of PFAS

PFAS represent a large group of synthetic organofluorine chemicals that have been manufactured and used in industrial and consumer applications since the 1940s (Woodward et al., 2024). The Organization for Economic Co-Operation and Development (OECD) has identified nearly 5,000 PFAS compounds on the global market (OECD, 2018; Valvira, 2024). PFAS are characterized by a fully or partially fluorinated carbon chain, where the carbon–fluorine bond is among the strongest in organic chemistry. This bond gives PFAS remarkable thermal stability, hydrophobic and lipophobic properties, and resistance to biological or chemical degradation. As a result, PFAS are often referred to as “forever chemicals,” because once released, they remain in the environment for decades or longer without significant breakdown (OECD, 2021).

These chemicals are entirely synthetic and enter the environment during manufacturing, use, storage, and disposal of PFAS-containing products (Ehsan et al., 2023). PFAS are valued for their physicochemical properties, including water, grease, and dirt repellency, thermal stability, chemical resistance, and low surface friction (EFSA, 2020). Consequently, they have been widely used in firefighting foams, non-stick cookware, stain-resistant textiles and numerous consumer products (Lyu et al., 2022). In industry, PFAS are applied e.g., in metal plating, paper industry, cleaning solutions and food packaging (OECD, 2021; Dasu et al., 2022). However, these same properties that make PFAS useful cause them to resist chemical, microbial, and photolytic degradation, leading to environmental persistence and bioaccumulation. This raises concern for human and ecosystem health (OECD, 2018; Valvira, 2024).

Because there are various PFAS compounds their identification is aided by classification into several groups based on the structure, functional group, chain length etc. The main classes are perfluoroalkyl and polyfluoroalkyl substances. Among these, perfluoroalkyl acids (PFAAs) comprising perfluoroalkyl carboxylic acids (PFCAs) and perfluoroalkane sulfonic acids (PFSAs), are the most widely detected in the environment (Perkola et al., 2023). PFCA compounds are often categorized as short-chain (C4–C6) and long-chain (C7–C12) compounds while PFSA as short-chain (C4–C5) or long-chain (C6–C12) (OECD, 2021). Ultra-short-chain PFAS have carbon chain length C1-C4. Another subdivision can be made between linear and branched PFAS (Charbonnet et al., 2021). Each compound is assigned an individual CAS number, a distinctive identifier issued by the Chemical Abstracts Service (CAS). The CAS number is an effective distinguisher since several abbreviations might exist for a single compound.

The sources of PFAS can be classified into point and diffuse sources (Nguyen et al., 2017; Woodward et al., 2024). Point sources of PFAS refer to identifiable, localized origins where PFAS are directly discharged into the environment. These include industrial facilities, military bases (Hu et al., 2016), airports (EPA, 2021), landfills (Gallen et al., 2017), and wastewater treatment plants (WWTPs) (Kurwadkar et al., 2022; Morethe et al., 2024) as conventional treatment processes often fail to remove PFAS effectively.

Diffuse sources of PFAS, on the other hand, are non-localized inputs that enter the environment through indirect pathways such as atmospheric deposition, urban runoff, and land application of biosolids. These sources are more challenging to quantify and manage due to their dispersed nature. The review by Lyu et al. (2022) emphasizes atmospheric deposition as a significant diffuse source, especially in remote and rural areas. Volatile PFAS precursors can travel long distances and deposit onto soils and water bodies via wet and dry deposition (Ehsan et al., 2023). Additionally, land-applied biosolids and recycled wastewater used for irrigation introduce PFAS into soils, where they can leach into groundwater (Gallen et al., 2017). Short-chain PFAS are relatively more mobile and can infiltrate deeper into soil profiles, posing long-term risks to aquifers (Li et al., 2020). Ehsan et al. (2023) further studied diffuse contamination in North Carolina, where PFAS were detected in household dust, landfill leachate, and agricultural runoff. These diffuse sources contribute to background PFAS levels in the environment and can accumulate over time, making them an underestimated and persistent threat.

Environmental monitoring has detected PFAS in diverse matrices. Groundwater and surface waters near airports and firefighting training sites show contamination (Benotti et al., 2020; Ruyle, Pickard, et al., 2021; Reinikainen et al., 2022). Sewage sludge, landfill leachate, and wastewater effluents have also been studied as contamination sources (Houtz et al., 2016; Gallen et al., 2017; Morethe et al., 2024). PFAS are present in stormwater, snow (Carlson and Tupper, 2020), fish (Junttila et al., 2019; Lin et al., 2021), soils and sediments (Zhong et al., 2021) and humans (Xiao et al., 2023). Their detection in all environmental matrices and human illustrates their global ubiquity. Some PFAS degrade into stable end products, while others are highly resistant to breakdown both in the environment and in living organisms, leading to long-term accumulation (EFSA, 2018; Fenton et al., 2021; Valvira, 2024).

## 2.2 Implications for Ecosystems and Human Health

PFAS are detected in all environmental matrices including water bodies, soil, air, biota, as well as in the blood samples of humans and wildlife across the globe. Their extreme persistence and mobility have led to serious concerns about ecological safety and public health. PFAS compounds are associated with a range of harmful effects, from reproductive and developmental toxicity to increased disease risks and biodiversity loss (Ehsan et al., 2023).

Some PFAS compounds are also suspected to be carcinogenic and genotoxic. PFOA is classified as Group 1 “carcinogenic to humans” while PFOS as “possibly carcinogenic to humans” (Group 2B) by the International Agency for Research on Cancer IARC (IARC, 2024). PFAS accumulate in blood and protein-rich organs, particularly the liver, and can be transferred across the placenta and secreted into breast milk (EFSA, 2020; Briassoulis and Ilia, 2025). Fluorinated compounds resist degradation in tissue and blood. In mammals, PFAS half-lives generally increase with chain length, although perfluorohexane sulfonate (PFHxS, C6) has a longer half-life than perfluorooctane sulfonate (PFOS, C8). In humans, the estimated half-lives of PFOS, perfluorooctanoic acid (PFOA), and PFHxS range from 2 to 5 years (and in some cases longer), meaning that even low-level exposures may lead to significant body burden over time (EFSA, 2020).

PFAS exposure can adversely affect fetal development. Epidemiological evidence indicates associations with reduced birth weight and a diminished response to vaccination in children (Fenton et al., 2021). In adults, outcomes include increased cholesterol, immune system disruption, and an increased risk of certain cancers (Fenton et al., 2021). Studies suggest possible or probable links to thyroid disease, testicular cancer, and delayed development of mammary glands (Fenton et al., 2021). Rosen et al. (2022) reported that high concentrations of PFOS and PFNA were associated with increased levels of total and HDL cholesterol.

For humans, diet is the dominant exposure pathway in Europe, accounting for more than 90% of total intake (EFSA, 2018). Estimated dietary exposure ranges from 1.26–20.86 ng/kg/day for PFOS and 1.47–18.27 ng/kg/day for PFOA, with fish and shellfish, meat products, and fruit identified as the most important sources (EFSA, 2018). Drinking water, indoor dust, and air contribute less overall, although groundwater contamination can make drinking water a significant local source. Following regulatory restrictions on PFOS and PFOA in the 2000s, declining blood concentrations of these compounds have been observed in some populations, while levels of shorter-chain PFAS, such as PFHxS and PFBS, have increased (Xiao et al., 2023).

In fish, PFAAs accumulate primarily in the liver and kidneys but are also found in muscle and blood (Junttila et al., 2019) while in soil ecosystems, they can be taken up by invertebrates such as earthworms (Pereira et al., 2018). Although concentrations in surface waters are generally not acutely toxic to aquatic organisms, long-term exposures may disrupt endocrine function and cause developmental and reproductive effects in fish (Ahrens and Bundschuh, 2014). Short-chain PFAAs appear to exert similar effects (Fenton et al., 2021) but are generally considered less harmful than their long-chain counterparts (Smaili and Ng, 2022).

Karamat et al. (2024) reported that terrestrial and aquatic plants are impacted by PFAS from contaminated soil and groundwater. PFAS are less toxic to plants at lower environmental concentrations, but higher PFAS levels impact biomass growth and photosynthetic activity of plants (Karamat et al., 2024). PFAS in plant tissue may cause phytotoxicity that can reduce biomass, alter seed germination, and slow down root elongation (Karamat et al., 2024). PFAS stimulates production of reactive oxygen species, causing oxidative stress and damage in plant cells and cellular respiration (Karamat et al., 2024). These effects threaten food security in contaminated regions.

**PFAS planetary boundary**

Cousins et al. (2022) proposed that PFAS define a distinct planetary boundary due to their exceptional persistence, global distribution, and potential for irreversible harm. This boundary is part of a broader framework that identifies Earth system limits we should not exceed to maintain a stable and safe environment. The study compares environmental concentrations of four major PFAS compounds PFOS, PFOA, PFHxS, and PFNA to health and environmental guideline levels. Their findings show that PFAS are now present in environmental media at levels that exceed safe thresholds globally (Cousins et al., 2022). Therefore, it is suggested that the global spread of these PFAS in the atmosphere has caused the planetary boundary for chemical pollution being exceeded.

## 2.3 Transport Mechanisms

**Surface Water Flow**

The movement of per- and polyfluoroalkyl substances (PFAS) in surface water systems is a critical pathway for environmental dispersion and human exposure. PFAS can enter rivers, streams, and lakes through multiple mechanisms, including surface runoff, sediment transport, lateral flow, and direct discharges from point sources such as wastewater treatment plants and industrial facilities (Weber et al., 2017; Rafiei and Nejadhashemi, 2023). Surface runoff is a dominant transport mechanism, especially in urban and industrial areas where impervious surfaces facilitate rapid water movement. PFAS bound to soil particles or dissolved in water can be mobilized during rainfall events and carried into nearby water bodies. The fate of PFAS in surface waters is influenced by hydrological dynamics such as flow velocity, residence time, and seasonal variations. For example, seasonal fluctuations in groundwater-surface water exchange can alter PFAS concentrations in rivers and lakes (Rafiei and Nejadhashemi, 2023). Once PFAS enters a water body, they can exist both dissolved in water and attached to suspended particles or sediments. These compounds may settle into sediments, be resuspended during turbulent flow, or diffuse between the water column and sediment layers (Lyu et al., 2022). PFAS are extremely persistent and do not degrade easily in surface water. Volatilization is negligible, especially for long-chain PFAS like PFOS and PFOA (Russell et al., 2008).

## Groundwater Flow

Per- and polyfluoroalkyl substances (PFAS) are known to persist in groundwater due to their chemical stability (Weber et al., 2017; Lyu et al., 2022). Contaminated soil areas can remain a source of groundwater pollution long after the original production or disposal activities have ceased (Lyu et al., 2022). This highlights the importance of underground transport pathways even without obvious point sources. PFAS can enter surface waters through groundwater baseflow, especially in areas near known contamination sources like military bases (Reinikainen et al., 2022). Septic systems and application of wastes to agricultural land were found to be a potential source of PFAS in groundwater in Wisconsin (Silver et al., 2023).

Groundwater modelling and field measurements have shown that PFAS-contaminated groundwater discharges into streams and tributaries, contributing significantly to PFAS loads in surface water (Pétré et al., 2022). These contributions are often not linked to visible point sources, indicating subsurface transport. A wide range of PFAS compounds have been detected in groundwater, typically at clearly higher than background levels (Martin et al., 2019; Silver et al., 2023). Detected compounds include legacy substances such as PFOS and PFOA, as well as new, emerging PFAS compounds. Short-chain PFAS are typically more mobile than their long-chain counterparts and reach deeper into the ground (Li et al., 2020). The interaction between groundwater and surface water plays a significant role in the environmental persistence and distribution of PFAS in aquatic systems.

## Atmospheric Deposition

Some PFAS compounds are also volatile and have been observed to be transported in the atmosphere (Johansson et al., 2018). The significance of long-range transport is demonstrated by the fact that PFAS compounds have been detected all over the world, even in polar regions (Johansson et al., 2018). Johansson et al. (2018) showed that atmospheric deposition to the environment does not originate from a single dominant source, but from several overlapping pathways, which are direct manufacturing discharges, transport of sea spray aerosols and degradation of precursors. Factories that produce or use PFAS can release PFAA-containing particles or vapours directly into the atmosphere. These can travel with air currents and eventually be deposited in rainfall or snow. Oceans function as a secondary

source. PFAS that accumulate in seawater can be transferred back into the atmosphere through sea spray. Tiny saltwater droplets carry PFASs and related compounds into the air, which then settle again through atmospheric deposition (Johansson et al., 2018). This process helps explain why PFAS are detected even in remote, oceanic, or polar regions far from known industrial sources. Many PFAS are released into the environment in precursor forms (e.g., fluorotelomer alcohols, FTOH). These precursors are more volatile and can travel long distances in the atmosphere. Over time, they degrade chemically (often via photochemical reactions) into persistent PFAAs, which then return to land and water through deposition (Johansson et al., 2018).

**The role of Wastewater Treatment Plants**

Wastewater treatment plants (WWTPs) play a central role in the environmental transport of PFAS and their transformation. While designed to remove organic pollutants and nutrients, conventional treatment processes are ineffective in eliminating PFAS (Gallen et al., 2018). In some cases, they even help the transformation of precursor compounds into more persistent and mobile short-chain PFAS (Houtz et al., 2016; Morethe et al., 2024). Several studies have shown that PFAS concentrations in effluent often exceed those in influent, particularly for short-chain perfluorocarboxylic acids such as PFHxA and PFPeA, reflecting precursor degradation during biological treatment (Gallen et al., 2018; Smith et al., 2023). This means WWTPs act not only as pathways, discharging PFAS into receiving waters, but also as reactors where precursors are converted into terminal degradation products (Wang et al., 2017). Moreover, PFAS inputs arise from both diffuse domestic sources, such as consumer product use, and localized point sources, including industrial discharges, making WWTP effluent a complex mixture reflecting both everyday and industrial chemical use patterns (Lyu et al., 2022; Morethe et al., 2024). Consequently, WWTPs are recognized as central elements in the environmental PFAS cycle, requiring targeted management strategies to reduce their role as both transport pathways and transformation points.

## 2.4 Regulatory Framework

The European Union (EU) restricts PFAS compounds in various directives and regulations including the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation (European Parliament and Council, 2020). The commercialization and use of PFOS was restricted in 2008 when it was added to the REACH restriction list, and in 2009, it was further listed in the Annex B (restriction of production and use) to the Stockholm Convention on Persistent Organic Pollutants (POPs) (EFSA, 2018). The use of PFOS in firefighting foams has been banned in the EU since 2011, after which foams were replaced with other PFAS-based formulations. The use, production, and import of PFOA and its salts have been restricted in the EU since 2020 (ECHA, 2023). In addition to these, several other PFAS (C8-14-PFCA and PFHxS) have been included on the Candidate List of Substances of Very High Concern (SVHC) by the European Chemicals Agency (ECHA) (ECHA, 2018).

A challenge in regulating PFAS compounds is the ease with which new, structurally similar chemicals can be developed (Reinikainen et al., 2019). Replacement PFAS compounds often contain shorter perfluorinated carbon chains, and they or their degradation products are less bioaccumulative, but equally persistent and highly mobile. In addition, short-chain PFAS compounds (especially PFBA, PFBS) sometimes need to be used in larger quantities, resulting in higher emissions compared to long-chain PFAS compounds (Brendel et al., 2018). Switching to the use of short-chain PFAS compounds does not solve the environmental problem caused by PFAS. In January 2023, Germany, the Netherlands, Norway, Sweden, and Denmark submitted a proposal to restrict the manufacture, commercialization, and use of the entire PFAS group (ECHA, 2023).

EU Revised Drinking Water Directive (DWD) 2020/2184 was published in December 2020. Amending Directive 2006/118/EC was released in 2022 (European Commission, 2022). The revised DWD must be implemented by all EU Member States by 12 January 2026. The directive aims to protect drinking water sources and lower water treatment costs. The combined concentration of the sum of 20 PFAS must not exceed 0.1 µg/L. The term “Sum of PFAS” refers to the total concentration of 20 PFAS compounds identified as relevant to drinking water safety. The limit for parameter “total PFAS” is 0.5 µg/L (European Parliament and Council, 2020; EU Commission, 2022).

Finland is adapting to this directive and will begin testing drinking water for PFAS compounds by 2026. Finnish Drinking Water Decree (1352/2015) states that PFAS20 can be monitored in raw water, water leaving treatment plants, or within the distribution network, since PFAS contamination typically originates from raw water sources (Finnish Ministry of Social Affairs and Health, 2015).

The Environmental Quality Standard (EQS) set by the EU Water Framework Directive 2013/39/EU for PFOS in inland surface waters is 0.65 ng/L and in other surface waters 0.13 ng/L (WFD, 2015). In October 2022, the European Commission proposed environmental quality standards (EQS) for the sum of 24 PFAS compounds, including PFOS, in surface water and groundwater. The proposal was informed by the European Food Safety Authority (EFSA) and supported by scientific opinions from the Scientific Committee on Health, Environmental and Emerging Risks. The suggested EQS value for both surface and groundwater is 4.4 ng/L, expressed as PFOA equivalents (European Commission, 2022). In September 2025, the Council of the EU and the European Parliament made a provisional agreement to update the WFD (Council of the EU, 2025). The revision includes the addition of new priority substances and tighter environmental quality standards. Trifluoroacetic acid (TFA) was added to the initial proposal of the sum of 24 PFAS. The need for the update was recognised as in the EU, 46% of surface waters and 24% of groundwater fail to meet good chemical status.

The European Food Safety Authority (EFSA) has also evaluated the health risks of PFAS exposure via food and drinking water. Based on toxicological assessments, EFSA established a tolerable weekly intake (TWI), which corresponds to a safety limit of 4.4 ng/kg bodyweight for the sum of four PFAS compounds (PFOA, PFOS, PFNA, PFHxS) referred as *PFAS4* (EFSA, 2020).

# 3 PFAS Source Tracking

This section reviews the existing literature on PFAS source tracking methods. It discusses challenges, current methods and future directions. Studies linking specific PFAS to a particular industry or activities are reviewed. This review organizes PFAS source tracking methods into three main categories: statistical techniques, geospatial approaches and modelling techniques. This study considers source tracking publications from 2015 onward.

PFAS source tracking refers to the process of identifying and differentiating the origins of PFAS contamination in environmental media (Charbonnet et al., 2021). It is a critical component of emerging environmental forensics and public health, as it allows researchers and regulators to find contamination sources, understand transport mechanisms, and design effective remediation strategies. Approaches combine analytical measurements, multivariate statistical analyses, and an understanding of PFAS fate and transport within a conceptual site model. Differentiation between sources can be supported by identifying PFAS compounds strongly linked to unique activities (sentinel compounds), examining isomer ratios, grouping statistically similar source profiles (clustering), and considering local hydrogeochemical conditions (Charbonnet et al., 2021).

Although early efforts to associate PFAS in environmental samples with their sources began in the 2010s e.g. Zhang et al. (2013), the systematic study of PFAS source tracking was more clearly defined and structured in the 2020s. One of the central publications by Charbonnet et al. (2021) reported PFAS source tracking within a forensic context. It emphasized using analytical chemistry, statistical tools, and knowledge about sites specifics to trace PFAS origins. It also highlights the need for standardization and advancements in the PFAS forensics field. Terminology in the field varies; terms such as source tracing, tracking, differentiation and attribution are often used interchangeably.

## 3.1 Challenges in PFAS Source Tracking

Source attribution for PFAS remains a considerable challenge due to their widespread use across diverse industrial sectors and environmental pathways. Many PFAS compounds are

common to multiple activities resulting in overlapping chemical profiles (Charbonnet et al., 2021). Environmental transport mechanisms, including water, air, and soil further challenge source differentiation by redistributing and mixing PFAS from several origins (Lyu et al., 2022; Antell et al., 2024). Additionally, several different precursor compounds may degrade into the same terminal PFAS, making it difficult to trace them back to their original sources (Wang et al., 2017; Charbonnet et al., 2021). PFAS forensics has traditionally concentrated on identifying individual sources or isolated contamination events rather than more complex scenarios involving multiple, overlapping sources (Benotti et al., 2020). Determining the type of PFAS releases from chemical fingerprint alone is difficult due to the long history of PFAS use (Woodward et al., 2024), environmental transportation and degradation (Weber et al., 2017), and the incomplete knowledge of all PFAS compounds present (Lin et al., 2021; Lyu et al., 2022).

**Analytical and Sampling Limitations**

PFAS are often present at very low concentrations, typically in the nanogram per litre (ng/L) range. Detecting these trace levels is analytically challenging, particularly for short-chain and emerging PFAS, due to limitations in sensitivity, standard availability, and instrument resolution (Charbonnet et al., 2021; Joseph et al., 2023). The reliability of PFAS source tracking depends heavily on the analytical tools employed. Every method has limitations, and instruments like LC-MS/MS, GC-MS, or high-resolution MS vary in their sensitivity and analyte range. To obtain a more complete PFAS fingerprint for a sample, these methods should be used in combination (Charbonnet et al., 2021).

When doing field sampling and analysing PFAS compounds in the laboratory, the risk of sample contamination must be carefully considered (Carlson and Andersen, 2025). Potential contamination can originate from contact with human skin, laboratory or field materials (e.g., permanent markers, sticky notes). Because PFAS are used in water-repellent clothing and various synthetic materials, trace amounts may accidentally enter the sample during collection, transportation or analysis (Carlson & Andersen, 2025). This risk emphasizes the necessity of rigorous contamination control measures in PFAS monitoring to ensure reliable and accurate measurements and interpretations. The choice of storing containers and gloves is especially important.

**Data Integration and Interpretation Challenges**

Data integration creates additional challenges when source tracking is done on large datasets that consist of measurements from several projects (Gamlin et al., 2024). Different studies may target different PFAS compounds, use varying analytical methods and lack sufficient metadata. This complicates cross-study comparisons and the development of reliable source attribution models. While machine learning has shown promise in distinguishing sources like AFFF from non-AFFF inputs, its effectiveness depends on the availability of large, diverse training datasets to detect statistical similarities and differences (Kibbey et al., 2020).

Differences observed in PFAS fingerprints may not necessarily indicate distinct sources but could instead reflect factors such as the sampled environmental medium, distance from the release point, transport pathways, or site-specific conditions described in the conceptual site model (CSM), including redox environments that influence microbial transformation (Charbonnet et al., 2021). Therefore, PFAS fingerprints must always be interpreted in a site-specific context, taking into account historical land use, known sources, hydrological flow paths, and environmental conditions etc. (Charbonnet et al., 2021).

While standards for legacy compounds such as PFOS and PFOA are commercially available, many precursors and emerging fluorochemicals do not have standardized analytical methods. There is no universally accepted framework for PFAS source tracking (Charbonnet et al., 2021). Commonly quantified PFAS compounds may have limited value for source differentiation due to their widespread use across multiple applications. As a result, contamination from distinct sources may appear chemically similar (Guelfo and Adamson, 2018).

Hydrological variability, such as changes in river volume flow due to rainfall or seasonal snowmelt in surface and groundwater systems, can influence concentrations and further complicate PFAS signatures. This fluctuation can mask contamination signals, making source attribution more difficult when relying solely on concentration data. PFAS loads (mass flux) might be more informative in long term source tracking. For example, Pétré et al. (2022) demonstrated that PFAS loads downstream of a manufacturing plant remained quantifiable even when concentrations varied with flow conditions. Similarly, Carlson and Tupper (2020) observed seasonal changes in PFAS concentrations in snowmelt and runoff, emphasizing the importance of considering hydrological context in PFAS monitoring.

PFAS source tracking is a complex but important field that requires an integrated approach and careful consideration of site-specific factors, analytical limitations, and data integration challenges. No single method can identify, quantify, and trace all PFAS compounds to their origins, and therefore most PFAS source tracking studies employ a combination of tracking techniques. To structure this review, PFAS source tracking approaches are grouped into three main categories: statistical, geospatial, and modelling approaches. The following sections introduce each category, focusing on its principles, strengths, limitations, and applications in recent studies.

## 3.2 Multivariate Statistical Techniques

Multivariate statistical methods are essential tools for interpreting complex PFAS datasets and uncovering patterns that support source attribution. Commonly used statistical tools include Principal Component Analysis (PCA), clustering, correlation analysis and isomer profiling (Johansson et al., 2018; Charbonnet et al., 2021). Statistical methods are primarily used to explore patterns, groupings, and correlations in PFAS data. These techniques are valuable for identifying compositional trends and potential source categories, but they do not quantify source contributions or simulate environmental processes.

Effective application of multivariate techniques requires careful data preprocessing, including normalization, treatment of censored values, and selection of relevant variables. Sample size and data quality significantly influence the robustness of statistical outcomes. Statistical multivariate techniques are powerful because they can manage large datasets, integrate precursor and transformation product data, and work with both targeted and non-targeted PFAS profiles (Charbonnet et al., 2021; Antell et al., 2024). Although various multivariate analysis tools have been adopted for environmental source tracking, the verification and validation of these models have not kept pace and remain a critical need in environmental forensics. If applied incorrectly, these methods can lead to misleading or invalid conclusions (Charbonnet et al., 2021). Multivariate methods require careful preprocessing, accurate large datasets, and supporting contextual information for reliable interpretation (Zhang et al., 2016). Statistical methods are most effective when integrated into a multi-approach framework that includes analytical chemistry, site history, and modelling approaches (Charbonnet et al., 2021).

**Unsupervised pattern recognition**

Unsupervised statistical methods are used to explore PFAS data without predefined categories or training labels. These techniques help reveal natural groupings and underlying structures in the data, which can be linked to different contamination sources (Charbonnet et al., 2021). Common unsupervised methods in PFAS source tracking include PCA, HCA, UPGMA (Unweighted Pair Group Method with Arithmetic Mean) and other clustering algorithms (Ruyle, Pickard, et al., 2021; Joseph et al., 2023; Antell et al., 2024).

Principal Component Analysis (PCA) is a widely applied technique in PFAS studies (Charbonnet et al., 2021). It reduces the dimensionality of large datasets by transforming correlated variables into a smaller set of components (Murphy and Morrison, 2014). These components often reflect dominant PFAS compounds or compound groups that may be linked to specific source types (Zhang et al., 2016). PCA has been used to distinguish between PFAS profiles associated with AFFF, wastewater, landfills, and industrial activities (Ruyle, Pickard, et al., 2021; Silver et al., 2023; Joseph et al., 2023; Morethe et al., 2024). In many studies, PCA results show clear clustering of samples based on source type, which supports source attribution.

Hierarchical Clustering (HCA) is another common method that groups samples or variables based on similarity. These cluster profiles can then be compared with known source profiles to find potential origins. It is often used alongside PCA to validate clustering patterns (Morethe et al., 2024). With PCA and HCA, Zhang et al. (2016) were able to group samples into 3 different categories based on likely contamination sources such as airports and textile mills, waste sector and metal plating industries.

Algorithms such as Unweighted Pair Group Method with Arithmetic Mean (UPGMA) have been applied to PFAS datasets to form dendrograms that visualize sample relationships (Nguyen et al., 2017; Ruyle, Pickard, et al., 2021). More recently, Uniform Manifold Approximation and Projection (UMAP) has been introduced as a nonlinear dimensionality reduction method that maintains both the relationships between nearby samples (local structure) and the overall rangement of sample groups in the dataset (global structure). UMAP is especially useful for visualizing complex PFAS datasets, including those with oxidized precursor data from TOP assays (Antell et al., 2024).

Overall, unsupervised pattern recognition methods are valuable for identifying compositional fingerprints and grouping similar samples in PFAS datasets. They do not quantify source contributions or simulate environmental processes, but they give a strong foundation for interpreting compositional variability and linking samples to potential sources. They have enabled the isolation of source-specific PFAS fingerprints and provided insight into precursor degradation and transformation along hydrological flow paths (Ruyle et al., 2021b).

**Correlation and Ratio Analysis**

Correlation analysis is another important statistical approach used in PFAS source tracking as it helps identify co-occurrence relationships between PFAS compounds (Nguyen et al., 2017; Silver et al., 2023). Morethe et al. (2024) reported strong intra-group relationships among PFAS classes, such as PFCA and fluorotelomers (FT), which can indicate shared sources or transformation pathways. In addition to intra-PFAS correlations, the associations between PFAS and other anthropogenic markers, such as human waste indicator (HWI) substances, can be used to support source tracking (Silver et al., 2023). For example, Silver et al. (2023) used correlation test to link short-chain PFAS compounds with HWIs such as artificial sweeteners (acesulfame and sucralose) and pharmaceuticals (e.g. carbamazepine) (Silver et al., 2023). These associations suggest groundwater contamination originating from septic systems or domestic wastewater.

In addition to correlation analysis, PFAS ratio analysis is frequently used to support source identification. Ratios such as PFHpA/PFBS, PFHxA/PFOA, (Guelfo and Adamson, 2018) or ∑PFCA/∑PFSA (Antell et al., 2024) can help differentiate between contamination types. For instance, high PFSA ratios are often associated with AFFF use, while elevated PFCA ratios may indicate landfill leachate or domestic discharge (Antell et al., 2024). These ratios can also be used to infer atmospheric deposition or precursor degradation (Morethe et al., 2024). PFHpA/PFOA ratios were used to distinguish local versus atmospheric sources of PFAS in snow and stream water in the US Quarry Road Trails winter outdoor recreational area (Carlson and Tupper, 2020).

Isomer profiling is another targeted approach that supports source attribution. By analysing the proportion of branched and linear isomers of PFAS compounds like PFOS and PFOA,

researchers can suggest manufacturing origins (Johansson et al., 2018; Langberg et al., 2021). Electrochemical fluorination (ECF) produces a mixture of isomers, while telomerization results in mostly linear compounds. Isomer patterns have been used to distinguish regional differences in PFAS sources and to detect transformation processes in the environment (Ma et al., 2018; Benotti et al., 2020).

Together, correlation and ratio-based analyses enable additional interpretation of PFAS behaviour and source identification. While they do not offer quantitative apportionment, they are important tools for interpreting PFAS profiles and supporting hypotheses about contamination origins.

In summary, multivariate statistical techniques play a central role in PFAS source tracking by revealing compositional patterns, grouping samples, and identifying potential contamination sources. Unsupervised methods such as PCA, HCA, and UMAP are powerful for exploratory analysis, while correlation and ratio analyses help interpret relationships and support source attribution. These methods are most effective when applied to high-quality datasets with sufficient sample size and when combined with contextual information such as site history, land use, and analytical chemistry (Charbonnet et al., 2021). Although statistical techniques alone cannot quantify source contributions or simulate environmental transport, they provide a foundation for understanding PFAS contamination. Their integration with geospatial and modelling techniques, discussed in subsequent chapters, enhances the overall reliability of source attribution and supports decision-making in environmental management.

## 3.3 Geospatial Techniques

Geospatial techniques are essential in PFAS source tracking because they help link contaminant profiles to potential sources and identify environmental pathways (Charbonnet et al., 2021). Two major approaches are described here: spatial sampling design and GIS-based land use and proximity analysis. Spatial sampling design focuses on how and where samples are collected across a study area, using strategies like equal area grids to ensure representative coverage and uniform sampling protocols (Silver et al., 2023). GIS land use and proximity analysis, on the other hand, links PFAS concentrations to surrounding land use types and potential sources by mapping features like buffer zones, capture zones, and land cover categories (Schaider et al., 2016; Lin et al., 2021).

Integrating spatial and hydrogeological data like watershed delineation and flow modelling supports the identification of transport pathways and source zones (Ruyle, Pickard, et al., 2021). Conceptual Site Models (CSMs) enhance spatial attribution by incorporating transport pathways, land use, and exposure scenarios (Charbonnet et al., 2021). Together, these approaches offer a comprehensive framework for understanding and tracking PFAS sources in complex environments.

**Spatial Sampling Design**

A well-designed spatial sampling plan is critical for collecting representative PFAS data across a study area, as it influences the reliability of spatial analysis and the ability to detect contamination gradients (Silver et al., 2023). Approaches such as Equal Area Grid sampling have been used to ensure uniform coverage and minimize spatial bias. For example, Silver et al. (2023) applied an Equal Area Grid design to sample 450 private wells across Wisconsin. Each sampling point was characterized by land use within a 500-meter radius, allowing classification into categories such as developed, agricultural, forested, and grassland. This design enabled spatial comparisons of PFAS detection rates and concentrations across land use types, revealing that higher contamination levels in developed areas were likely influenced by septic systems.

Site-based comparisons are also a common strategy where samples are collected from specific locations to assess spatial variation in PFAS concentrations. Carlson and Tupper (2020) collected snow samples at different locations within a ski area to evaluate PFAS contamination from ski waxes. Concentrations were highest at the race start line and decreased with distance, supporting the hypothesis of local abrasion as the primary source. Also, Morethe et al. (2024), Nguyen et al. (2017) and Joseph et al. (2023) collected samples from representative locations to capture spatial differences in PFAS profiles. Distance-based sampling has also been used to trace contamination sources in non-riverine settings. Langberg et al. (2021) mapped PFAS concentrations across Lake Tyrifjorden, revealing a contamination gradient decreasing with distance from a paper factory. The spatial distribution of PFOS isomers further supported the identification of the factory as the dominant source.

Sampling based on hydrological distances and flow directions gives insight into how PFAS moves through the environment, helping to identify transport pathways and suspected source zones. Pétré et al. (2022) and Geosyntec (2018) used upstream-downstream sampling in the PFAS manufacturing plant at Cape Fear River to quantify the contribution of the plant to the total PFAS load in the watershed. Downstream samples showed elevated concentrations of fluoroethers linked to a manufacturing plant, while upstream variability was attributed to wastewater inputs. Woodward et al. (2024) used time-of-travel-sampling method. They studied PFAS fingerprint from upstream to downstream, combined with data on potential industrial sources to provide context and improve source attribution. Other studies have used flow rate data to estimate PFAS fluxes. Ma et al. (2018) calculated riverine PFAS loads into Taihu Lake, showing spatial distribution consistent with the intensity of nearby industrial activities.

In addition to flow-based sampling, watershed delineation further enhances spatial resolution. (Ruyle, Pickard, et al., 2021) used hydrological data from the U.S. Geological Survey to delineate watersheds and compare PFAS profiles between AFFF-impacted and non-impacted areas. This approach allowed spatial resolution of contamination patterns and linked PFAS signatures to land-use history. Zhang et al. (2016) applied ArcGIS Trace Downstream and USGS StreamStats tools to estimate source impact based on hydrological distance, assuming exponential decay of PFAS concentrations with distance from the source. They also acquired a list of sites subject to environmental regulation and potential PFAS source sites.

**GIS-Based Land Use and Proximity Analysis**

GIS-based techniques are widely used to analyse the relationship between PFAS concentrations and surrounding land use features (Charbonnet et al., 2021). These analyses often involve mapping buffer zones, capture zones, and proximity to point sources such as industrial sites, fire stations, or wastewater treatment plants (Lin et al., 2021).

Proximity and buffer zone analysis is an approach for linking PFAS profiles to nearby point sources. This involves creating buffer zones around sampling locations or distinct PFAS sources like fire stations or industrial sites. For example, in the study by Schaider et al. (2016), GIS was used to define capture zones around domestic wells, representing the area

where groundwater is likely to flow toward the well. Residential density within these zones was analysed to investigate the influence of septic systems on PFAS contamination. Similarly, Silver et al. (2023) conducted proximity analysis to permitted waste application sites, revealing higher PFAS levels near biosolids and septage disposal areas.

Statistical analysis can be used together with geospatial techniques. For instance, Lin et al. (2021) examined spatial relationships between clusters and land use in ArcGis Pro by creating buffer zones with radii of 5, 10, and 20 km around each sampling site, and the area covered by different land-use classes within each buffer was calculated. The land-use categories included forest, shrubland, grassland, cropland, urban, and built-up areas. In addition, datasets of potential point sources, such as airports and industrial facilities, were visually inspected to support source interpretation.

Land use and land cover (LULC) analysis has also been applied to support source identification. Salim et al. (2019) linked changes in urbanization and bare land to pollutant profiles in stormwater runoff, while Morethe et al. (2024) compared PFAS profiles across wastewater treatment plants located in industrial, mining, and residential zones, illustrating spatial differences in contamination sources. Hu et al. (2016) used geospatial analysis by combining PFAS concentrations in U.S. drinking water with four main types of potential point sources, including industrial areas, military fire-training sites, and wastewater treatment plants. A multivariate spatial regression model confirmed that point sources were significant predictors of PFAS concentrations: for example, each additional military site within a zone was associated with 35% increase in PFOS levels.

Overall, geospatial approaches establish essential context for PFAS source tracking by linking contamination patterns to land use, infrastructure, and environmental transport mechanisms. Spatial sampling design ensures representative data collection, GIS analysis identifies potential sources, and hydrological modelling reveals how PFAS moves through the landscape. These methods are most effective when integrated with statistical and modelling analyses, forming a combined methodological framework for reliable source attribution. In the following chapter, modelling techniques are discussed to complement geospatial and statistical approaches by quantifying source contributions.

## 3.4 Modelling Approaches

Modelling techniques in PFAS source tracking are used to simulate environmental transport, estimate source contributions, and assess mass balances of fluorinated compounds (Kibbey et al., 2021; Lyu et al., 2022). These methods often enhance statistical and geospatial analyses by simulating environmental processes and resolving overlapping source profiles. In this section, modelling approaches are grouped into three main categories: receptor models, machine learning classifiers, and process-based and probabilistic models.

**Receptor Models**

Receptor models are commonly used to apportion PFAS sources based on chemical fingerprints observed in environmental samples. Two widely applied methods are Positive Matrix Factorization (PMF) and Principal Component Analysis combined with Multiple Linear Regression (PCA–MLR) (Salim et al., 2019).

PMF is a supervised statistical method that breaks down PFAS concentration data into different source profiles and estimates how much each source contributes to the overall contamination. It has been used to separate complex mixtures of PFAS compounds into identifiable patterns linked to specific sources and to quantify the influence of different sources. For example, Lin et al. (2021) applied PMF to fish tissue data to identify distinct PFAS profiles, supporting source attribution across sampling locations.

PCA–MLR combines dimensionality reduction with regression analysis to identify main components and link them to source categories. Salim et al. (2019) compared PMF and PCA–MLR for modelling PFAS in stormwater runoff and found that PMF performed better in separating overlapping source signals and quantifying their contributions. These applications demonstrate how receptor models can be used to analyse PFAS mixtures but also to trace them back to specific sources. Across studies, PMF has shown better source resolution, while PCA–MLR remains useful for interpretable linear modelling. Also Qi et al. (2016) found that PMF offered better source identification. They used PCA–MLR to identify two major industrial source groups, while PMF provided separation into three emission profiles, including food-packaging, textile and metal industries.

### Machine Learning Classifiers

Machine learning models offer adaptable and versatile data-driven methods for PFAS source tracking. These models can manage large datasets and complex relationships, making them suitable for mixed-source environments and non-linear patterns.

Supervised classification models, such as Random Forests, Support Vector Machines and neural networks, have been used to distinguish PFAS sources based on chemical fingerprints. For example, Kibbey et al. (2021) evaluated twelve supervised classifiers to distinguish between AFFF and non-AFFF sources. The models were trained on over 8000 water samples. The study found that machine learning models outperformed traditional ratio-based methods, especially when PFAS compositions were altered by environmental processes. Similarly, Stults et al. (2023) used supervised machine learning classifiers including k-nearest neighbours, decision trees, and SVM to analyse PFAS chemical fingerprints found in fish tissue. The study classified PFAS profiles into four source types: aqueous film-forming foam (AFFF), paper industry wastewater, diffuse sources, and long-range transport.

Feature selection has also been applied in environmental forensics. Joseph et al. (2023) used three machine learning models: SVC, logistic regression and Random Forest to identify PFAS fingerprints from six source types. The models were trained on both target and suspect PFAS data, which helped improve their ability to recognize patterns. They also used a recursive feature elimination (RFE) method to find the smallest possible set of PFAS compounds that were most useful for identifying sources. For example, 3H-PFENA was found to be a strong indicator of AFFF contamination, while 4:2 FTThA was linked to landfill sources. Including suspect PFASs improved the models' accuracy, especially for samples with low concentrations.

### Process-Based and Probabilistic Models

In addition to receptor and machine learning models, process-based and probabilistic approaches are used to simulate PFAS transport, transformation, and mass balances. These models integrate environmental parameters and chemical properties to estimate source contributions and contaminant dynamics (Rafiei and Nejadhashemi, 2023; Antell et al., 2024).

Recent work illustrates how these methods can offer deeper insights into PFAS origins and behaviour. For example, Ruyle, Pickard, et al. (2021) used Markov Chain Monte Carlo (MCMC)-based Bayesian inference modelling to reconstruct precursor profiles from TOP assay results by combining known oxidation yields of electrochemical fluorination (ECF) and fluorotelomer (FT) precursors. The approach successfully distinguished between ECF and FT AFFF contributions and quantified their proportions. The study also used mass balance modelling with Extractable Organofluorine (EOF) data to quantify the proportion of fluorine not explained by targeted PFAS, suggesting additional sources such as septic systems or consumer products.

Other studies have emphasized the potential of combining precursor oxidation with classification tools. For instance, TOP assays that oxidize precursors to measurable PFCAs can be paired with PCA-based classification models to distinguish between sources such as AFFF, WWTP effluent, and landfill leachate with high accuracy (Antell et al., 2024). Bayesian inference was also applied to estimate precursor concentrations and synthesis pathways.

At larger spatial scales, integrated hydrogeochemical models have been developed to track PFAS fate and transport across watersheds. Rafiei and Nejadhashemi (2023) developed a distributed hydrogeochemical model integrating SWAT-MODFLOW-RT3D to simulate PFOS transport. The model included advection-dispersion equations (how PFAS moves and spreads through soil and water), adsorption isotherms (how PFAS attaches to soil), and mass balance equations for PFOS in various environmental compartments. The model quantified PFOS transport via runoff, sediment, lateral flow, and leaching, identifying urban areas and known potential PFAS contamination sites as major contamination contributors.

Similarly, load estimation tools like LOADEST have been applied to quantify daily PFAS loads from industrial facilities based on river discharge and concentration data (Pétré et al., 2022). By linking flow dynamics with contaminant levels, these models not only enable assessment of emission rates and temporal trends but also contribute to source tracking by identifying when and where major PFAS releases occur. Together, these studies show that process-based and probabilistic models can go beyond simple concentration measurements to offer a system-based understanding of PFAS transport and source apportionment. By combining chemical profiling, precursor transformation, and environmental modelling,

these approaches allow researchers to explain complex contamination scenarios and estimate contributions from multiple, overlapping sources.

In conclusion, modelling approaches offer essential quantitative tools for PFAS source tracking. Receptor models such as PMF and PCA–MLR help resolve overlapping sources, while machine learning classifiers offer flexible and accurate source identification. Process-based and probabilistic models simulate environmental transport and transformation, enabling mass balance assessments and precursor reconstruction. These methods are most effective when integrated with statistical and geospatial analyses, forming a multidisciplinary framework for rigorous and detailed source attribution.

## 3.5 Summary of Central Methods Used for Tracking PFAS Sources

Table 1 below provides an overview of central methods used for tracking PFAS sources, summarizing their principles and limitations.

Table 1. Summary of PFAS source tracking methods.

| Category | Method | Description | Limitations | References |
|---|---|---|---|---|
| **Statistical Approaches** | Unsupervised Pattern Recognition (e.g. PCA, HCA, UPGMA, UMAP, clustering) | Explores data structure and compositional patterns without prior source labels; reduces data dimensionality; groups samples by similarity and reveals dominant PFAS fingerprints | Does not quantify source contributions; sensitive to data preprocessing and outliers. | (Zhang et al., 2016; Nguyen et al., 2017; Guelfo and Adamson, 2018; Ruyle, Pickard, et al., 2021; Langberg et al., 2021; Joseph et al., 2023; Silver et al., 2023; Morethe et al., 2024; Antell et al., 2024) |
| | Correlation and Ratio Analysis (e.g. Spearman/Pearson, PFAS ratios, isomer profiling) | Examines relationships between PFAS and indicators or among PFAS; ratios and isomer patterns to suggest sources and transformation; can distinguish manufacturing origins | Correlations might suggest a shared source, but does not prove it; ratios can be influenced by environmental processes (degradation) which alter the original PFAS composition; isomer data limited | (Zhang et al., 2016; Nguyen et al., 2017; Johansson et al., 2018; Ma et al., 2018; Guelfo and Adamson, 2018; Benotti et al., 2020; Langberg et al., 2021; Silver et al., 2023; Morethe et al., 2024) |
| **Geospatial Approaches** | Spatial Sampling and Hydrological Design (e.g. Equal Area Grid, upstream/downstream, watershed delineation) | Ensures spatially representative data collection; reduces sampling bias and supports spatial trend analysis; identify source zones by integrating spatial and hydrogeological information | Requires detailed knowledge of hydrological and site data, resource intensive | (Weber et al., 2017; Nguyen et al., 2017; Geosyntec, 2018; Carlson and Tupper, 2020; Ruyle, Pickard, et al., 2021; Langberg et al., 2021; Pétré et al., 2022; Silver et al., 2023; Morethe et al., 2024) |
| | GIS-Based Source Proximity and Land Use Analysis (e.g. LULC, buffer zones, capture zones) | Links PFAS occurrence to land use types, source proximity, and potential impact zones using GIS tools | Spatial data may be outdated or low resolution; does not account for subsurface transport | (Schaider et al., 2016; Zhang et al., 2016; Hu et al., 2016; Salim et al., 2019; Lin et al., 2021; Silver et al., 2023) |

| | | | | |
|---|---|---|---|---|
| **Modelling Approaches** | Receptor Models (e.g. PMF, PCA–MLR, UNMIX, CMB) | Break down PFAS data into source profiles and estimate contributions from each source using statistical models | Requires large high-quality datasets; sensitive to input uncertainties; may not resolve all sources in complex overlapping environments | (Qi et al., 2016; Salim et al., 2019; Lin et al., 2021) |
| | Bayesian & Probabilistic Models (e.g. Bayesian inference, MCMC) | Estimates source contributions, precursor profiles, or transformation pathways with uncertainty quantification; integrate prior knowledge with observed data | Requires statistical expertise; computationally intensive; results depend on prior assumptions | (Ruyle, Thackray, et al., 2021; Ruyle, Pickard, et al., 2021; Antell et al., 2024) |
| | Fate & Transport Modelling (e.g. SWAT, MODFLOW, RT3D, hydrogeochemical models) | Simulates PFAS movement, transformation, and distribution in environmental compartments using flow dynamics, chemical reactions and site-specific parameters | Requires detailed site and chemical data; model calibration can be challenging; environmental variability may affect model accuracy | (Weber et al., 2017; Langberg et al., 2021; Rafiei and Nejadhashemi, 2023; Beisner et al., 2024) |
| | Mass Balance & Load Estimation (e.g. LOADEST) | Provides quantitative estimates of PFAS fluxes and stocks in environmental systems using measured concentrations and flow data; can give hint on source inputs | Often assumes steady-state or simplified flows; sensitive to sampling frequency and data quality; variability in flow can lead to uncertainty | (Nguyen et al., 2017; Ma et al., 2018; Pétré et al., 2022; Woodward et al., 2024) |
| | Supervised Machine Learning (e.g. Random Forest, SVM, Neural Networks, RFE). | Trains algorithms to classify or predict PFAS sources based on compositional data; can identify characteristic features; handle complex patterns | Requires large, labelled datasets; interpretability can be challenging; performance depends on data representativeness | (Kibbey et al., 2020; Kibbey et al., 2021; Joseph et al., 2023; Stults et al., 2023; Dong et al., 2024) |

The literature review presented different PFAS source tracking methods, including statistical, geospatial, and modelling approaches. Each method has its own strengths, and their use depends on the available data and the goals of the study. In this thesis, the selected methods are Principal Component Analysis (PCA), hierarchical clustering on principal components (HCPC), Spearman correlation test and compositional data analysis. These methods were chosen to help answer the research question: What are the main components contributing to the variation in PFAS concentrations in the study area, and how do they relate to known potential PFAS sources? The selected methods are explained in more detail in the following chapters.

# 4 Materials and Methods

This chapter introduces the study area, outlines the data collection, integration and statistical analysis methods. It describes selection of parameters in the database and further explains the software and tools that are used for source tracking application.

## 4.1 Study Area - River Vantaa Watershed

The River Vantaa Watershed is located in southern Finland. It is in the Uusimaa and Kanta-Häme regions and it spans over 14 municipalities (Kivimäki et al., 2013). The total area of the watershed is 1680 $km^2$. The main stream, River Vantaa is over 100 km long. It originates from Hausjärvi (southern Kanta-Häme), and it runs southward into the sea at Vanhankaupunginlahti Bay at the Gulf of Finland. The largest tributaries are Keravanjoki, Luhtajoki, Palojoki, Lepsämänjoki, Tuusulanjoki, and Kytäjoki in the catchment area. Several tributaries branch into smaller, stream-like reaches in the upper catchment area (Kivimäki et al., 2013).

River Vantaa constitutes a backup water source for the Helsinki metropolitan area. Area's groundwater sources are also used for municipal drinking water in the upstream municipalities (Kivimäki et al., 2013). The watershed area is also an important recreational area for over 1 million citizens (Junttila et al., 2021).

The watershed contains only a few lakes, which are generally small and shallow. The largest lake in the region is Lake Tuusula, with a surface area of 6.0 $km^2$ and an average depth of 3.1 meters (Kivimäki et al., 2013). Other major lakes include Lake Hirvi, Ridas, and Kytäjärvi, all of which are located in the upstream. Low lake density reduces the watershed's capacity to buffer flow variations (Junttila et al., 2021). Rivers, lakes and the catchment area are shown in Figure 1.

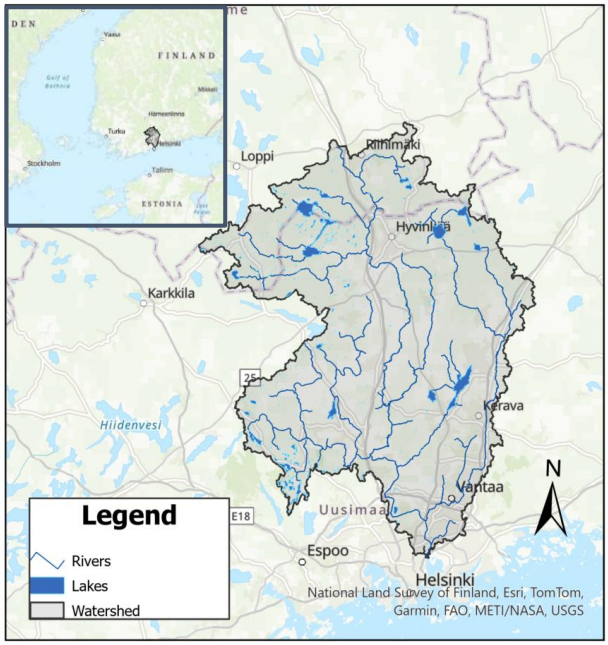


Figure 1. River Vantaa Watershed.

Within the watershed, there are a total of 29 groundwater areas, many of which are hydraulically connected to river systems (Kivimäki et al., 2013). Of these, 21 areas are used for municipal water supply. Kivimäki et al. (2013) have studied the interactions of groundwater and surface water in the watershed. Some rivers in the area are known to be strongly influenced by groundwater, especially in its upper reaches. During low-flow periods, groundwater contributes considerably to the baseflow and therefore contaminated groundwater can further contaminate the surface water. Over 250 groundwater discharge points were identified along the riverbanks, demonstrating the major role of groundwater in river flows. Groundwater also discharges into the Rekolanoja stream, and especially during low-flow periods, the proportion of groundwater in the stream is significant (Turtiainen, 2023). These hydraulic connections can influence the movement of contaminants like PFAS between aquatic matrices in the watershed.

River water can also infiltrate into groundwater and reach water intake wells (Kivimäki et al., 2013). The presence of chemical compounds could pose significant risks to groundwater quality in areas where hydraulic connections exist. The entry of dissolved compounds into the intake wells may be possible.

**PFAS Contamination in the Watershed**

PFAS concentrations in the River Vantaa watershed are exceptionally high in the national context (Turtiainen et al., 2025). Junttila et al. (2019) reported that River Vantaa had the highest PFAS concentrations among 13 Finnish rivers studied. The chemical status of the water bodies has deteriorated, and both River Vantaa and Lake Tuusula are classified as failing to meet good status (Junttila et al., 2021). PFAS contamination in the River Vantaa watershed has been the subject of multiple investigations over the past decade. Studies by Junttila et al. (2019, 2021), Siimes et al. (2019), Turtiainen (2023) and Turtiainen et al. (2025) have documented elevated concentrations of PFAS compounds, particularly PFOS, in surface waters, wastewater, stormwater, fish tissue, and groundwater.

In VHVSY report, Junttila et al. (2021) PFAS compounds were present in waste and surface waters throughout the year. PFAS loading increased during periods of high runoff and flow. In the groundwater samples studied, PFAS concentrations did not exceed the limits set by the EU DWD. Wastewater treatment plants were identified as major contributors, particularly during low-flow periods when treated effluent forms a significant portion of river water (Junttila et al., 2021).

Seasonal variation in river flow is high in the River Vantaa watershed, and it influences PFAS transport and concentrations (Junttila et al., 2021). During low-flow periods, especially in summer and early autumn, the relative share of municipal wastewater in river water increases due to reduced surface runoff (Junttila et al., 2021). This can increase PFAS concentrations downstream of wastewater discharge points. Conversely, during spring snowmelt and heavy rainfall events, surface runoff becomes the dominant contributor to river flow, leading to increased diffuse PFAS loading from urban areas, landfills, and contaminated soils (Junttila et al., 2019). Elevated PFAS concentrations and a wider range of compounds were observed in stormwater samples collected after such events, particularly in June 2020 following a 14 mm rainfall event (Junttila et al., 2021).

PFAS loading has been especially high in the lower reaches of the river and Keravanjoki. PFOS, a long-chain PFAS compound, accumulates in fish tissue at levels that exceed or approach the environmental quality standard for biota in the lower reaches of River Vantaa and Keravanjoki (Junttila et al., 2021). In the lower Vantaa River, the standard is exceeded multiple times.

In groundwater taken from Riihimäki and Hyvinkää, PFAS compounds were detected near old landfill sites and industrial areas (Junttila et al., 2021). The 2025 VHVSY study found PFAS in all groundwater samples, with the highest PFAS concentration observed in groundwater monitoring well HPS8 (2749 ng/l), located near the northern boundary of the closed Kerava landfill (Turtiainen et al., 2025). In contrast, PFAS concentrations south of the waste management area were only about one percent of the concentrations observed in northern wells, even in locations very close to the landfill's leachate discharge point. This was likely due to the bedrock surface, which prevents groundwater from flowing southward. The study showed that PFAS contamination in groundwater is spatially variable and often linked to historical land use and hydrological connectivity with surface waters.

**Potential sources**

The PFAS load to the watershed originates from multiple sources. Identified emission sources include industrial sites, landfills, and the Helsinki-Vantaa Airport accident site, which has been recognized as a major contributor to PFAS load in the lower reaches of the river (Junttila et al., 2021).

A list and geospatial data for potential PFAS sources in the watershed are available in the national soil contamination database MATTI, maintained by SYKE. The register contains information on approximately 25,000 sites where current or past activities may have released harmful substances into the soil. The data has been collected since the early 1990s and includes both active and closed sites, with operational histories reaching back up to 150 years. Some sites have confirmed contamination, while others are included due to known risks associated with similar activities (Pyy et al., 2013).

This list included 644 possible point source polluters (contaminated soil) in River Vantaa watershed. Eleven industrial sectors (233 contamination sites) were identified as highly potential PFAS emission sources: airports, chemical and plastics industry (including chemical manufacturing), metal industry including metal production, industrial landfills, municipal landfills, hazardous waste treatment, other waste management (e.g. composting, sludge handling, incineration, fill areas), wastewater treatment plants, machine workshops, surface treatment industry and drum washing.

Oil and chemical spill sites were excluded due to their unclear nature and widespread distribution across the watershed, which could not bring meaningful separation between clusters. Fuel storage and distribution stations were also excluded as they were not considered highly potential sites and were similarly dispersed throughout the area, making it impossible to associate them with any specific cluster. The locations of 233 potential PFAS pollution sites were mapped in ArcGis Pro (Figure 2) to help trace industrial activities linked to the presence of PFAS contamination.

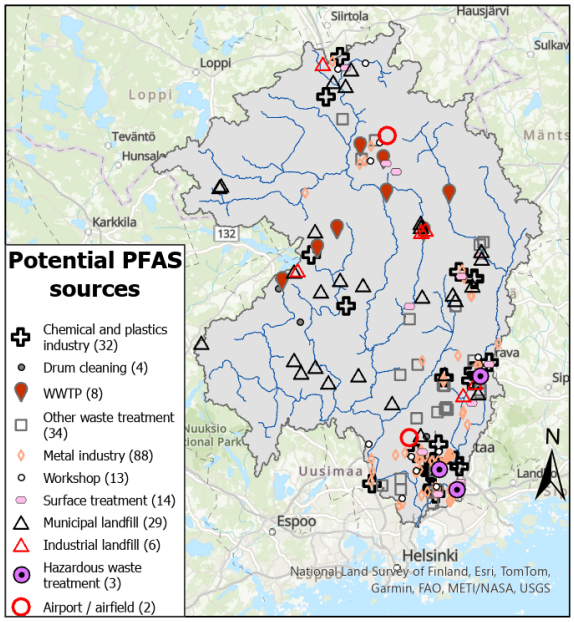


Figure 2. MATTI-register potential PFAS sources (SYKE, 2018).

It is important to point out that a limitation of this analysis is that not all facilities listed in the MATTI-register database necessarily emit PFAS, and the database may not completely capture all potential sources.

## 4.2 Data Collection

Data for PFAS source tracking was gathered from multiple sources, including scientific publications, and environmental monitoring reports. These datasets are combined to create a detailed database of PFAS concentrations and potential sources. The data utilized in this study were not originally collected for source tracking purposes but rather compiled by

various stakeholders for diverse monitoring and regulatory objectives. Therefore, the dataset reflects a range of sampling strategies, frequency and timing of data collection, and analytical frameworks (number of targeted compounds, method detection limits, laboratory protocols etc.) This variability can affect the representativeness and comparability of the data. The uncertainty arising from the integration of data from multiple sources is acknowledged; however, given its comprehensiveness and spatial-temporal coverage, it remains highly valuable for source tracking application.

**GTK (Geological Survey of Finland) database**

GTK and the University of Helsinki have performed PFAS monitoring between March 2023 and March 2024, resulting in 57 river water samples (Ulmanen, 2024). Fifty PFAS substances were targeted and 14 were found in concentrations higher than their detection limit at least once. In addition, 3 groundwater samples were collected in 2022 in the north of the watershed.

**VHVSY (Water protection association of Vantaa River and Helsinki region)**

Another source of PFAS data was gathered from the VHVSY open file report "Julkaisu 89/2021, Vantaanjoen PFAS-hanke Loppuraportti" (Junttila et al., 2021). The PFAS data from the project were collected during 5.5.2020-5.5.2021. The report studied occurrence of eighteen target PFAS compounds from various aquatic matrices: 12 river sites (n=60), 3 stormwater sites (n=18), 4 groundwater wells (n=4), and 6 wastewater streams from 2 WWTPs (n=40).

**FPP (Forever Pollution -project)**

The third source of PFAS data in this study is the Forever Pollution Project (FPP) open database. FPP is a cross-border journalism initiative launched in 2021 to investigate the widespread PFAS contamination across Europe. The database contains data from 40 sampling locations within the Vantaa River watershed, with samples collected between 2014 and 2022. Of these, 15 overlap with VHVSY data and 25 added new data for this study. The

retrieved dataset includes PFAS concentration profiles from 6 creeks (n=38), 16 ditches (n=121) and 3 rivers (n=24).

**Data Integration**

PFAS data from the abovementioned three sources were compiled into a single dataset of 382 water samples from 58 sites in the River Vantaa Watershed. Locations based on aquatic matrices are shown in Figure 3. These include:

- 60 samples from 10 locations collected through GTK monitoring, a snapshot campaign, and three groundwater wells
- 139 samples from 24 locations gathered during VHVSY campaigns targeting groundwater, surface water, stormwater, and wastewater treatment plants (WWTPs)
- 183 data records from the Forever Pollution Project (FPP), covering 25 locations

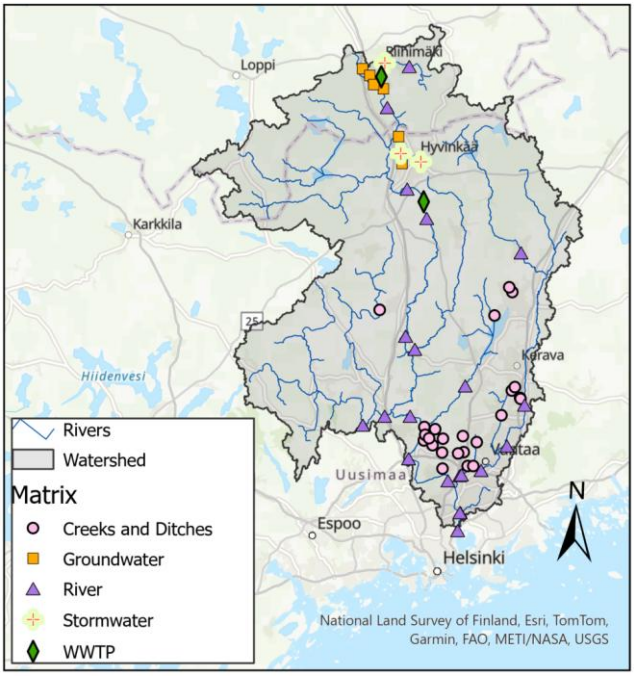


Figure 3. Spatial distribution of the sampling locations selected for the statistical analysis. Shape and colours of the markers indicate the matrix type (creeks and ditches, groundwater (GW), river, stormwater and WWTP).

## 4.3 Data Processing

The PFAS dataset compiled for this study integrates measurements from multiple sources across the River Vantaa Watershed. To enable good statistical analysis, the data were pre-processed to ensure consistency, completeness, and comparability across sampling locations and matrix types. The processing included compound selection based on detection frequency, exclusion of incomplete or outlier profiles, substitution of censored values, and normalization and standardization of the dataset. Only complete sample profiles containing results for all target PFAS compounds were retained.

Each dataset included 18-50 targeted PFAS compounds; 18 compounds were shared across all datasets. Detection percentages were determined, and only 9 PFAS that had detection rates exceeding 80% were retained for integration (Cloutier et al., 2008). If detection frequency is low, the compound does not adequately represent the variance in the dataset (Cloutier et al., 2008). Compounds included in the analysis are: PFBA, PFPeA, PFHxA, PFHpA, PFOA, PFNA, PFBS, PFHxS and PFOS. These represent both perfluoroalkyl carboxylic acids (PFCAs) and sulfonic acids (PFSAs) and include a mix of short-chain (PFBA, PFPeA, PFHxA, PFBS) and long-chain (PFHpA, PFOA, PFNA, PFHxS, PFOS) substances. Detection frequencies ranged from 87% (PFNA) to 99% (PFOA). A summary of compounds, their PFAA group, and chain-length classification is presented in Table 2.

Table 2. Nine selected PFAS compounds classified into PFCA/PFSA groups and short-chain/long-chain categories. Long-chain and PFSA compounds are shown in bold for clarity.

| | PFBA | PFPeA | PFHxA | PFHpA | PFOA | PFNA | PFBS | PFHxS | PFOS |
|---|---|---|---|---|---|---|---|---|---|
| PFCA or PFSA | PFCA | PFCA | PFCA | PFCA | PFCA | PFCA | **PFSA** | **PFSA** | **PFSA** |
| Short- or long-chain | Short-chain | Short-chain | Short-chain | **Long-chain** | **Long-chain** | **Long-chain** | Short-chain | **Long-chain** | **Long-chain** |

Sampling events with missing values for any of the selected nine compounds were excluded. Additionally, extreme outliers were eliminated. Repeated sampling was performed in most sites (e.g., 51 repeated monitoring throughout the year). Duplicates were removed. To facilitate more focused analysis and avoid bias in the analysis due to multiple samplings at

a given location, a smaller, representative sub-dataset was created. This subset was selected to preserve the primary characteristics and variability of the full dataset while allowing for more efficient processing and interpretation. The selection criteria and methodology used to define the sub-dataset are described in detail below.

To represent each of the 58 sampling locations, one to three complete sample profiles were selected depending on the observed variation among samples in each site. Sampling locations which could be represented by 1 individual measured consistent concentration over the measurement period, while sites with greater variability or longer monitoring periods required up to three samples to capture their chemical fingerprint.

A final representative subset of 99 environmental samples from 58 sites was compiled to capture spatial and compositional variation across the watershed. The data included creeks and ditches (n = 47, 25 sites), groundwater (n = 7, 6 sites), river (n = 31, 22 sites), stormwater (n = 6, 3 sites), and wastewater treatment plant effluent (n = 8, 2 sites (influent and effluent from both WWTPs)).

Several records in the raw dataset fall below the MDL. These measurements were substituted with MDL/2 (Beaudry et al., 2018; Zhong et al., 2021). Method detection limits from each data source are reported below in Table 3.

Table 3. Method detection limits in ng/L.

| Method Detection Limits (ng/L) | PFBA | PFPeA | PFHxA | PFHpA | PFOA | PFNA | PFBS | PFHxS | PFOS |
|---|---|---|---|---|---|---|---|---|---|
| **GTK surface waters** | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| **GTK groundwater** | 0.5 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **VHVSY surface water and groundwater** | 0.5 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **VHVSY stormwater and WWTP** | 0.8 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 |
| **Forever Pollution Project** | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.1 |

Data analysis was conducted with the integrated representative dataset of 99 analyses (individuals). Data must be normally distributed and standardized for successful multivariate statistical analysis. R 4.4.1. was used to check the normality of data; Shapiro-Wilk normality test, Quantile-comparison (QQ) plot and histograms were done (Field et al., 2012). Original data were not normally distributed, so a Box-Cox transformation was applied (Box and Cox,

1982). After normalization, variable standardization was done in R 4.4.1. Standardization (also called Z-score normalization) transforms data to have a mean of 0 and a standard deviation of 1. The relative distribution (skewness) remains the same (Field et al., 2012).

## 4.4 Multivariate Statistical Analysis

Principal Component Analysis (PCA) was done to reduce the dimensionality of the dataset and to identify the variance in PFAS concentrations in different parts of the watershed. Principal component approaches can be used as a pre-processing step prior to hierarchical clustering and partitional clustering (Husson et al., 2010). Then hierarchical clustering (HCA) was performed to group statistically similar samples (Cloutier et al., 2008). It has been applied in earth sciences and hydrochemical studies to classify water samples and identify patterns in complex datasets (Güler et al., 2002). In this study, clustering was performed using the HCPC (Hierarchical Clustering on Principal Components) method, which combines principal component analysis (PCA) with hierarchical clustering. The clustering was performed on the principal component coordinates of the standardized PFAS concentration dataset.

Both PCA and HCA were performed using the package “FactoMineR” (Lê et al., 2008) and its HCPC function (Husson et al., 2010). By default, the analysis used Euclidean distance as the dissimilarity measure and Ward’s method as the linkage criterion to construct the dendrogram (Husson et al., 2010). The Euclidean distance and Ward’s method are commonly used in environmental and hydrogeochemical studies due to their ability to produce compact and interpretable clusters (Güler et al., 2002; Cloutier et al., 2008). FactoMineR performed clustering by default on the first 5 dimensions of PCA.

A supplementary qualitative variable named "matrix" (river, creeks and ditches, groundwater, WWTP, stormwater) was added to differentiate sample types. A supplemental variable does not contribute to the construction of the PCA axes but can be used to help in their interpretation.

The Spearman correlation test was done in R 4.4.1. Raw data in ng/L were used for the Spearman test. For the Spearman rank-order test, pairwise p-values and adjusted p-values (Holm's method) are obtained. If $p<0.05$, correlation coefficients are statistically significant (Field et al., 2012).

## 4.5 Geospatial Analysis

To investigate potential spatial associations between potential PFAS sources and the studied sites, a geospatial analysis was performed using ArcGIS Pro. This analysis builds on the results of multivariate statistical methods, which group samples based on similarities in their PFAS chemical profiles. Clusters potentially reflect different contamination sources. The spatial distribution of these groups together with the map of potential PFAS sources could provide an understanding of the source attribution. The dataset comprised two primary point layers: (1) PFAS measurement points, grouped into five clusters based on chemical profiles, with approximately 9-43 points per cluster; and (2) MATTI-registered potential PFAS source locations, totalling 223 points, each assigned to one of 11 distinct source categories (e.g., metal industry, landfills, wastewater treatment). Both datasets were georeferenced using the ETRS-TM35FIN coordinate system.

The objective of the analysis was to determine whether specific PFAS source categories exhibit spatial proximity or concentration around particular clusters, thereby suggesting potential source-cluster relationships.

# 5 Results

This chapter presents the results of multivariate statistical and spatial analyses conducted to explore the relationships between PFAS sources and cluster sample profiles. Hierarchical Clustering on Principal Components (HCPC), Spearman correlation analysis, and spatial mapping in ArcGIS Pro were used to identify groupings, patterns and potential source associations. Both absolute concentration data (ng/L) and relative compositional profiles (%) were examined to characterize PFAS patterns.

## 5.1 PFAS Composition by Matrix Type

To better understand the distribution of PFAS across different environmental media, average concentrations (ng/L) of nine commonly detected PFAS compounds were calculated for each water matrix: creeks and ditches, groundwater, rivers, stormwater, and WWTP samples. Table 4 summarizes the matrix-specific median concentrations, providing a basis for interpreting the multivariate results presented in subsequent sections.

Table 4. Matrix-specific median concentration (ng/L) profiles for 9 PFAS common to all datasets. n = number of samples in the matrix.

| | **Creeks and Ditches** | **Groundwater** | **River** | **Stormwater** | **WWTP** |
|---|---|---|---|---|---|
| **n** | 47 | 7 | 31 | 6 | 8 |
| **PFBA** | 5.00 | 0.25 | 2.10 | 1.94 | 3.34 |
| **PFBS** | 4.00 | 0.44 | 0.60 | 0.20 | 1.56 |
| **PFHpA** | 8.00 | 0.48 | 1.46 | 1.15 | 2.85 |
| **PFHxA** | 14.70 | 0.48 | 3.00 | 0.97 | 8.70 |
| **PFHxS** | 21.00 | 1.31 | 0.70 | 0.20 | 2.41 |
| **PFNA** | 4.00 | 0.05 | 0.41 | 0.20 | 0.48 |
| **PFOA** | 16.00 | 0.98 | 2.00 | 1.24 | 3.90 |
| **PFOS** | 51.00 | 0.53 | 3.00 | 0.85 | 6.10 |
| **PFPeA** | 15.00 | 0.28 | 2.80 | 0.52 | 8.60 |
| **$\sum_9$PFAS** | 238.25 | 7.03 | 16.28 | 7.81 | 38.87 |

Table 4 data reveal significant variation in PFAS levels between matrices. Creeks and ditches exhibit the highest median total PFAS concentrations (median 238.25 ng/L), dominated by PFOS (51 ng/L), PFHxS (21 ng/L), and PFOA (16 ng/L). PFBS and PFNA show the lowest median (4 ng/L) in this water type. In contrast, groundwater samples show the lowest concentrations overall (median 7.03 ng/L), with individual compounds median exceeding 1 ng/L only for PFHxS.

WWTP samples display increased levels of short-chain PFAS, particularly PFHxA (8.7 ng/L), PFPeA (8.6 ng/L) and PFBA (3.34 ng/L), consistent with the influence of wastewaters (Morethe et al., 2024). Rivers and stormwater show intermediate concentrations ($\sum_9$PFAS medians 16.28 ng/L and 7.81 ng/L, respectively), with PFHxA, PFBA, and PFOA contributing notably. These concentration profiles might reflect the hydrological connectivity and point or diffuse sources of PFAS distribution.

To supplement the median concentration data, Figure 4 presents the average PFAS composition of each matrix type, highlighting the relative contributions of individual compounds.

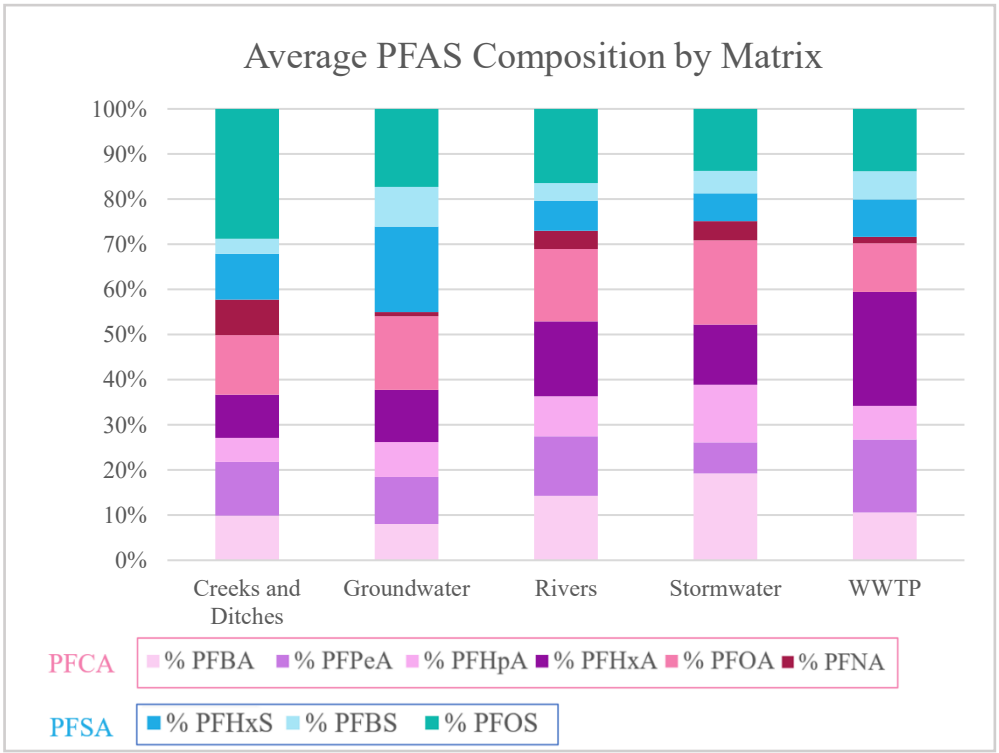


Figure 4. The average PFAS composition of each matrix type considering the 9 PFAS common to all datasets.

The mean percentage composition of PFAS compounds varies notably across different water matrices. In creeks and ditches, PFOS is the dominant compound, accounting for 28.8% of the total PFAS load, followed by PFOA (13.2%) and PFPeA (11.9%). Groundwater samples show a different profile, with PFHxS (18.9%) and PFOS (17.3%) being the most prominent, while PFNA has a low contribution (0.94%). River samples are characterized by a high proportion of PFHxA (16.6%), alongside elevated levels of PFOS (16.4%) and PFOA (16.1%). Stormwater exhibits the highest relative contribution of PFBA (19.3%) and PFHpA (12.8%), with PFOA also elevated (18.7%). WWTP samples are dominated by PFHxA (25.2%) and PFPeA (16.2%), indicating a strong presence of short-chain PFAS. These differences might reflect the influence of land use and origins of PFAS across the watershed.

Overall, creeks and ditches show the highest total PFAS concentrations, dominated by long-chain PFCA and PFSA, while WWTP effluent exhibits the highest relative contribution of short-chain PFAS. Rivers and stormwater display PFAS profiles dominated by PFCA. Groundwater samples have the lowest concentrations and a greater proportion of long-chain compounds. These compositional differences offer insight into PFAS sources and transformation processes. For example, the dominance of PFOS and PFHxS in creeks and ditches may reflect legacy contamination from firefighting foams, while the prevalence of short-chain PFAS in WWTP matrix suggests modern industrial or consumer product inputs.

## 5.2 Principal Component Analysis

Principal Component Analysis was applied on the standardized PFAS concentration dataset to reduce dimensionality and identify dominant patterns in the data (Davis, 1986). The first two principal components (PC) together explained 87.42% of the total variance in the dataset (76.22% for Component 1 and 11.20% for Component 2). PC3 explained 3.3% of the variance, bringing the cumulative explained variance to over 90%. This indicates that the first two dimensions are sufficient to describe most of the variation in the data. The number of components retained for interpretation was based on the Kaiser criterion, which recommends keeping only components with eigenvalues greater than one (Field et al., 2012). Accordingly, only the first two PC are displayed here, as they had eigenvalue >1 and together explained 87.4% of the total variance in the dataset. Table 5 presents the loadings for the first two components along with their respective variances.

Table 5. Principal component loadings and explained variance for the first two principal components.

| Parameters | Component 1 | Component 2 |
|---|---|---|
| **PFBA** | 0.504 | 0.817 |
| **PFBS** | 0.913 | -0.140 |
| **PFHpA** | 0.961 | 0.091 |
| **PFHxA** | 0.938 | 0.168 |
| **PFHxS** | 0.908 | -0.255 |
| **PFNA** | 0.878 | -0.218 |
| **PFOA** | 0.870 | -0.095 |
| **PFOS** | 0.910 | -0.295 |
| **PFPeA** | 0.889 | 0.277 |
| | | |
| **Explained variance %** | 76.22 | 11.20 |
| **Cumulative % of variance** | 76.22 | 87.42 |

The first component (PC1) is strongly positively correlated with e.g. PFOS, PFHxS, PFOA, suggesting a potential link to industrial sources. The second component is positively associated with PFBA, PFPeA, PFHxA and negatively with PFOS, PFHxS (Figure 5), indicating a contrast between short- and long-chain PFAS. These components reflect distinct contamination patterns, possibly related to differences in PFAS usage, mobility, or degradation pathways. Figure 5 presents a plot of loadings in PCA space.

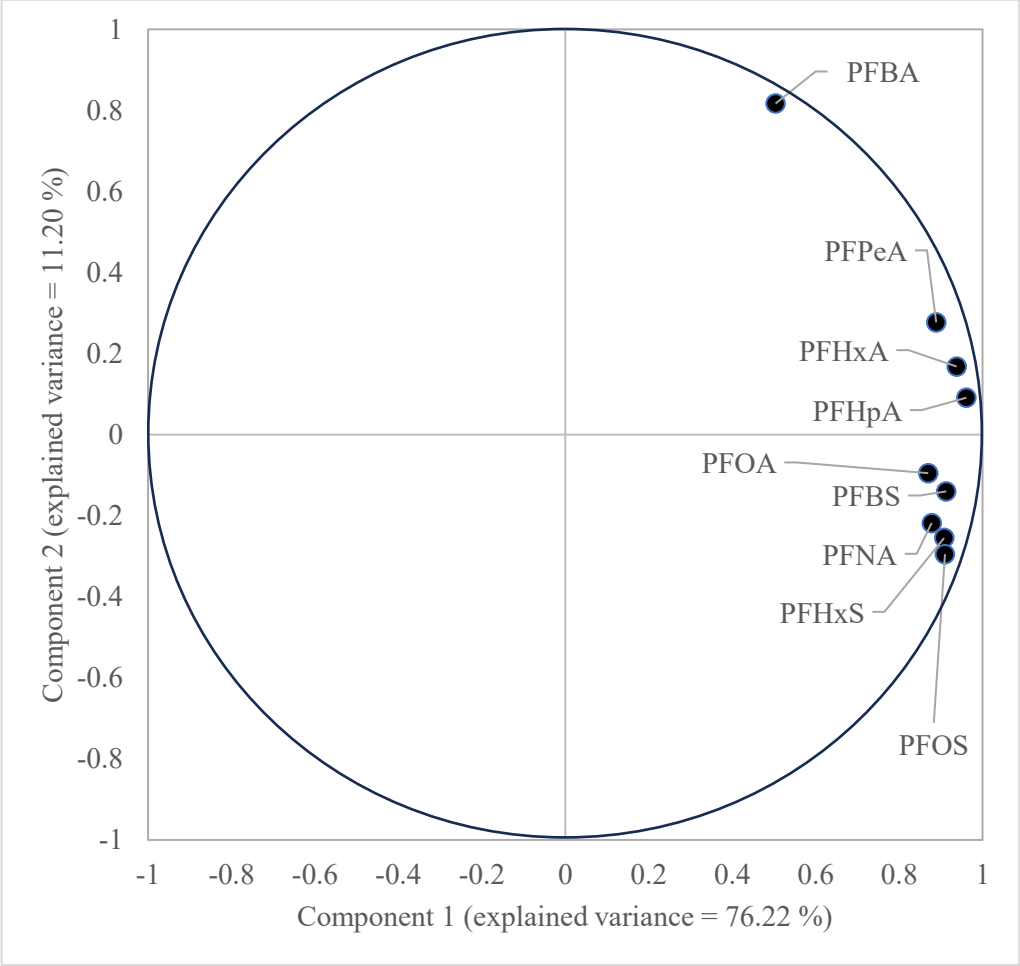


Figure 5. PCA plot of variable loadings for the first two principal components.

Figure 5 shows that all PFAS substances are strongly loaded on the positive side of PC1 except PFBA. This suggests that PC1 represents a general PFAS concentration gradient, capturing overall variation in total PFAS levels. In contrast, PC2 shows greater separation among individual compound loadings.

Supplementary categories representing different water source types were projected onto the PCA space to aid interpretation of PFAS distribution patterns. These categories were not used to compute the principal components but to understand their relationship with the PFAS profiles. Each individual and associated supplementary category is represented below in a multivariate scatter plot (Figure 6).

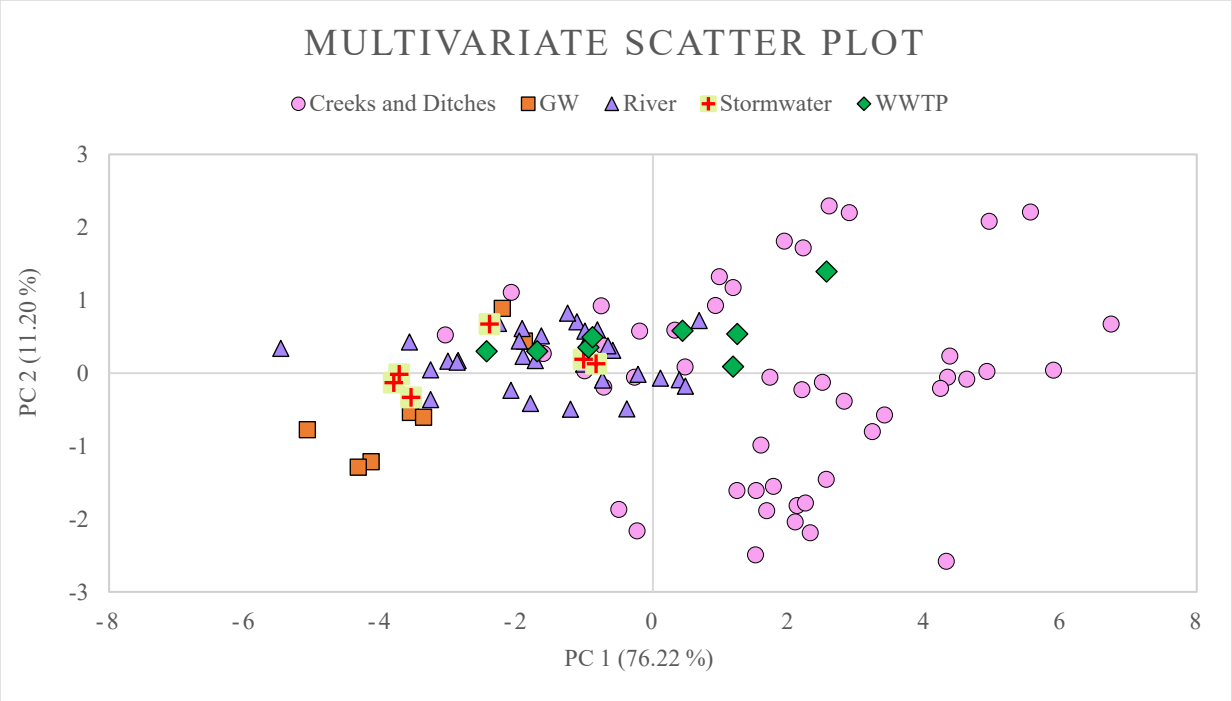


Figure 6. Multivariate scatter plot with supplementary factors.

Creeks and ditches were positioned positively along PC1, indicating elevated concentrations of multiple PFAS compounds. In contrast, groundwater, river, and stormwater samples were located negatively along this axis, suggesting generally lower PFAS levels in these natural water bodies. WWTP samples showed distinct positioning along the components, reflecting unique PFAS profiles. However, it is important to note that the WWTP category includes both influent and effluent samples. Therefore, the influence of treatment processes is not uniform across all WWTP data points. Some samples reflect untreated wastewater, while others represent post-treatment effluent. This mixed composition likely contributes to the distinct positioning of WWTP in the PCA space, especially along PC2, which may capture chemical transformations or selective removal of certain PFAS compounds.

These findings support the interpretation that multiple processes and source types contribute to PFAS distribution across the watershed.

## 5.3 HCPC Clustering Analysis

The primary output of the HCPC is the dendrogram (Figure 7) that illustrates the hierarchical clustering structure between the individual samples based on their PFAS profile similarity. The optimal number of clusters was automatically determined by the HCPC function in FactoMineR, which identifies the partition that results in the highest relative loss of within-cluster inertia. Five clusters were identified with sample sizes ranging from 9 to 43. Cluster 2 was the largest group (n = 43), while Cluster 5 was the smallest (n = 9). The dendrogram was cut at a linkage distance of approximately 0.4, and samples below that threshold were grouped into the same cluster. The number of clusters can be changed by adjusting the phenon line up or down (Güler et al., 2002).

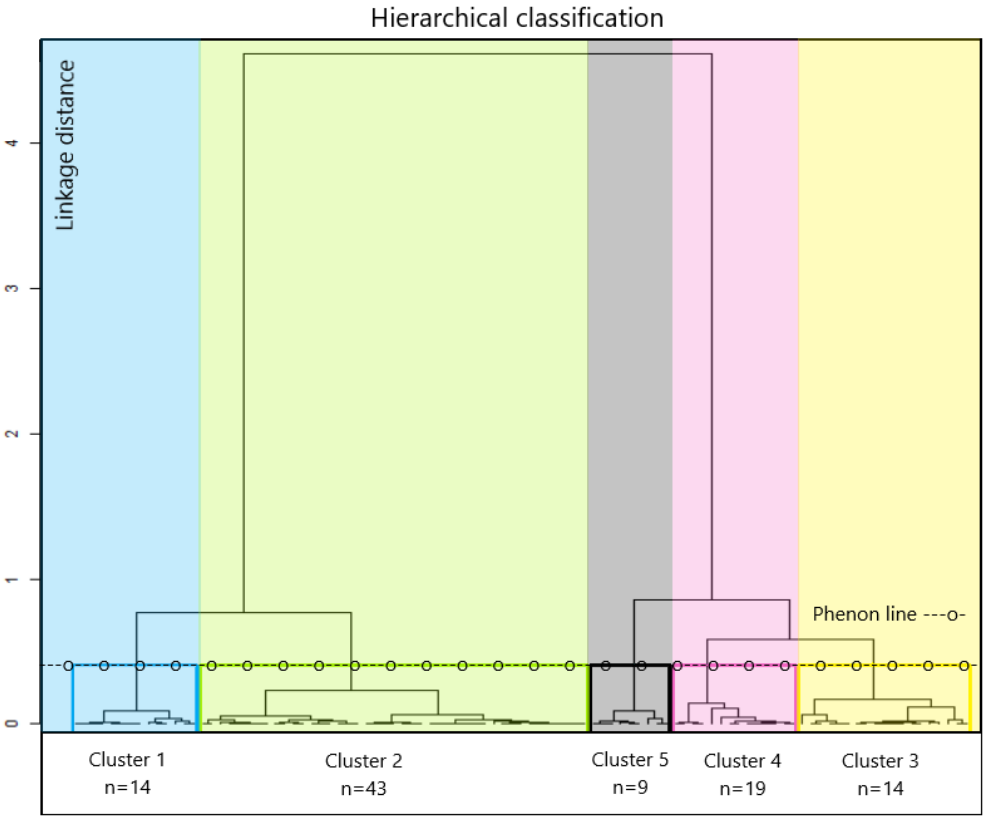


Figure 7. Hierarchical tree from HCPC for the 99 environmental water samples, showing the grouping into 5 clusters.

The vertical height at which branches merge reflects the similarity between clusters: clusters joined at lower heights are more similar (Cloutier et al., 2008). For example, Clusters 1 and 2 appear to be statistically closer to each other than Cluster 5 to Clusters 3 or 4, which are joined at higher linkage distances. Clusters 3 and 4 are most similar among the clusters while clusters 1 and 2 significantly differ from clusters 3,4 and 5. This pattern suggests varying degrees of similarity in PFAS profiles across the dataset.

The 99 samples are represented on the factor map produced by the first two principal components and coloured according to their cluster in Figure 8. The plot displays the first two principal components (PC 1 and PC 2). Each point represents an individual sample, and colours indicate the five clusters identified by HCPC. The spatial separation of clusters in the PCA space reflects differences in PFAS profiles among the groups.

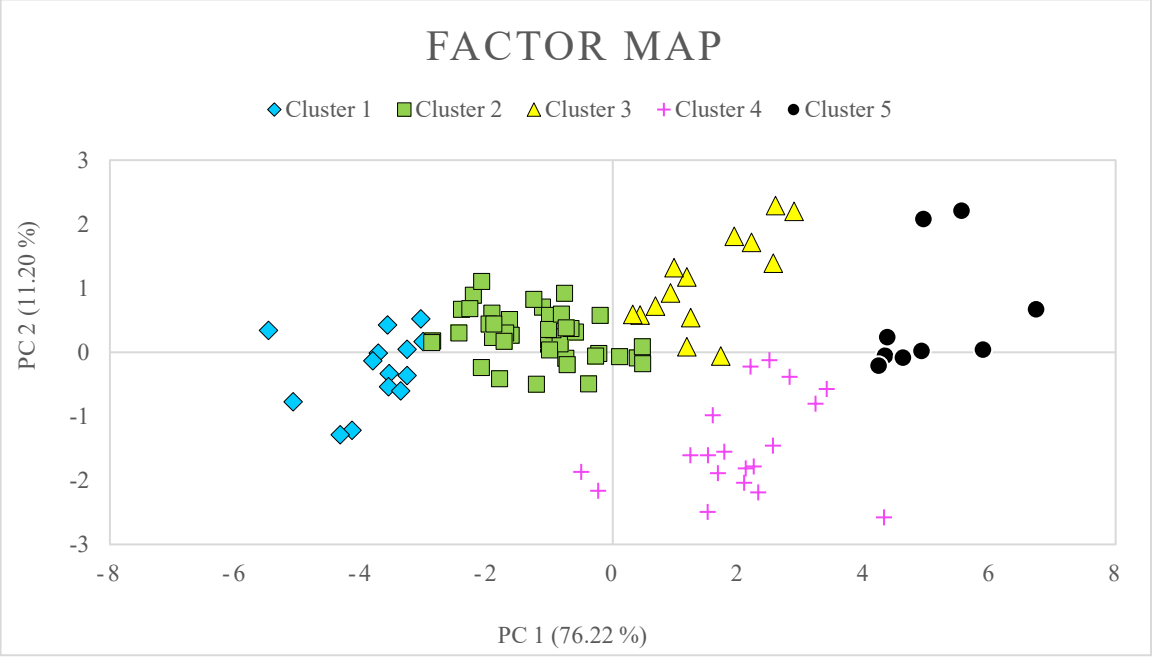


Figure 8. Multivariate factor map showing the distribution of samples in the PCA space coloured by cluster grouping.

Cluster 1 has low PFAS concentration values, Cluster 2 moderate but lower than average, Cluster 3 higher levels of PFBA and short-chain PFAS, Cluster 4 higher PFOS and PFHxS, and Cluster 5 presents the highest concentrations of the dataset across most PFAS. Each cluster is associated with specific categories: Cluster 1 mostly with groundwater, Cluster 2

with rivers, Cluster 3 with WWTP, and Clusters 4 and 5 with creeks and ditches. To further characterize the clusters, Table 6 presents the median concentrations (ng/L) of PFAS compounds for each group.

Table 6. Median PFAS concentration for each cluster in ng/L. The highest median across clusters is bolded and the lowest median is underlined. n = number of samples in the cluster.

| | **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** | **Cluster 5** |
|---|---|---|---|---|---|
| **n** | 14 | 43 | 14 | 19 | 9 |
| **PFBA** | 0.8 | 2.62 | 9 | <u>0.25</u> | **22** |
| **PFBS** | <u>0.32</u> | 0.81 | 2.23 | 5 | **20** |
| **PFHpA** | <u>0.49</u> | 1.76 | 7.5 | 9 | **61** |
| **PFHxA** | <u>0.62</u> | 3 | 16 | 15 | **90** |
| **PFHxS** | <u>0.28</u> | 1 | 8.5 | 26 | **130** |
| **PFNA** | <u>0.16</u> | 0.51 | 1.11 | 6 | **26** |
| **PFOA** | <u>0.79</u> | 3.3 | 5.71 | 22 | **160** |
| **PFOS** | <u>0.59</u> | 3 | 12.05 | 90 | **200** |
| **PFPeA** | <u>0.31</u> | 3 | 20 | 14 | **64** |
| **$\sum_9$PFAS** | <u>4.64</u> | 20.5 | 94.74 | 239.25 | **1571** |
| **∑PFCA/∑PFSA** | 3.05 | 2.54 | **4.51** | <u>0.69</u> | 1.32 |

Table 6 shows that Cluster 1 has the lowest median of all compounds except PFBA, which is lowest in Cluster 4. Cluster 5, on the other hand, showed the highest median of all compounds and generally the medians are significantly higher than in other clusters. There is a noticeable pattern of median concentration increasing from cluster 1 to 5.

Figure 9 displays the median concentrations (ng/L) of selected PFAS compounds across the five clusters identified by HCPC. Compounds are grouped by class: PFCAs are shown in shades of purple, while PFSAs are shown in shades of blue. The chart illustrates differences in PFAS composition between clusters, with Cluster 5 showing the highest overall concentrations, particularly of PFOA and other PFCAs, while Clusters 1 and 2 exhibit the lowest PFAS levels.

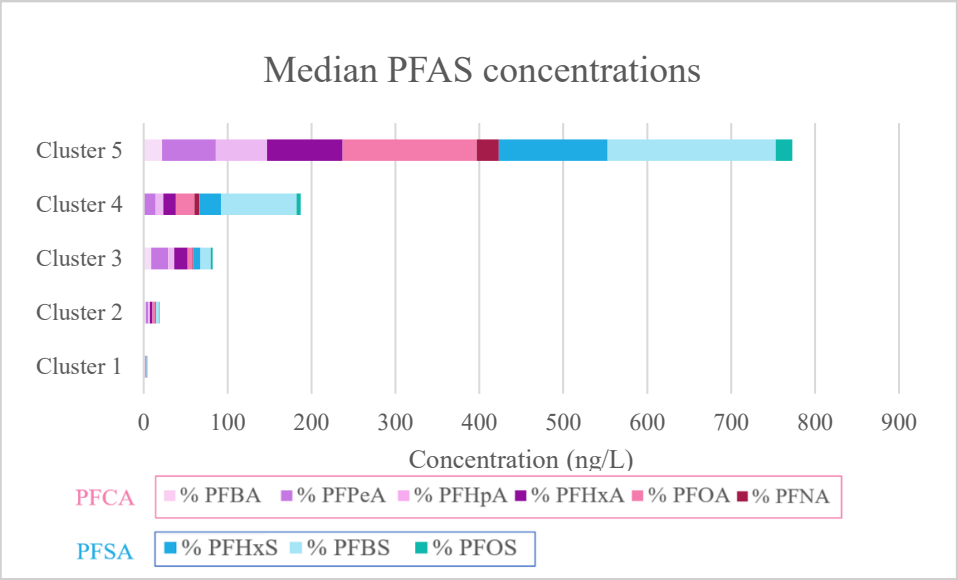


Figure 9. Median PFAS concentration across each cluster in ng/L.

Figure 10 displays the relative contribution (%) of individual PFAS compounds to the total PFAS concentration within each cluster. PFCAs are shown in shades of purple, while PFSAs are shown in shades of blue. Clusters 1, 2 and 3 compose over 66%, 72% and 75% of PFCA, respectively, while cluster 4 has the lowest share of PFCA (41%). Cluster 5 showed equivalent contributions of PFSA and PFCA.

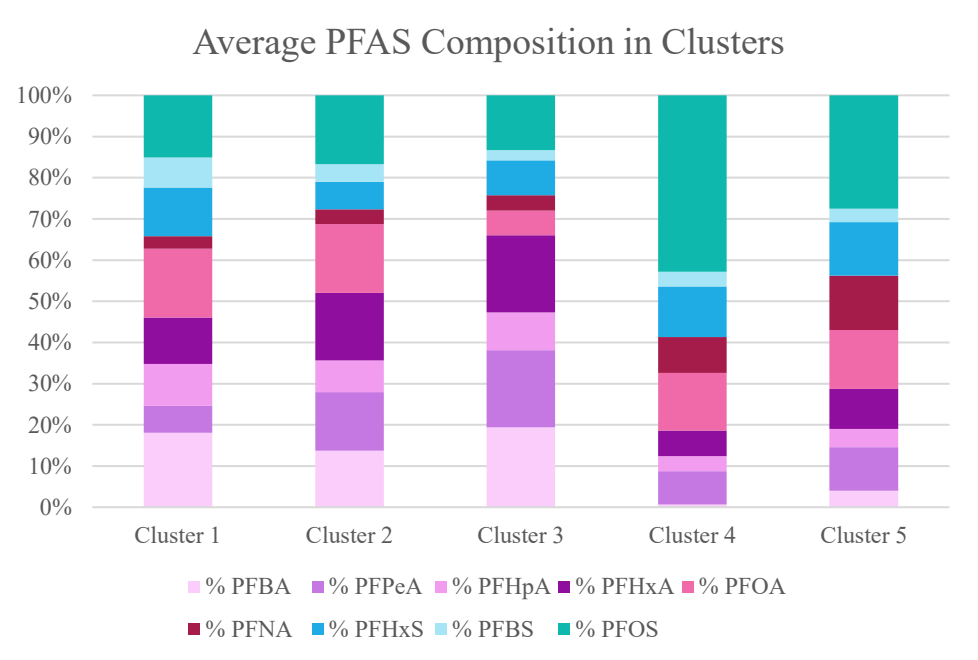


Figure 10. Average PFAS compositions across five clusters.

Cluster 1 had the lowest PFAS concentrations overall, with PFBA, PFOA and PFOS contributing the highest relative percentages. The cluster is primarily associated with groundwater samples. This cluster likely represents background levels or diffuse or background contamination possibly from long-range transport or legacy sources (Dasu et al., 2022; Silver et al., 2023). Cluster 2 presented moderate PFAS concentrations. PFOS, PFOA and PFHxA were among the most abundant compounds, suggesting mixed sources such as urban runoff or wastewater effluent (Zhang et al., 2016; Nguyen et al., 2017).

Cluster 3 showed a distinct profile characterized by short-chain PFAS compounds. PFBA, PFPeA, and PFHxA were dominant in both concentration and composition, with PFPeA reaching a median of 20 ng/L and contributing on average 18.7% of the total PFAS load. These patterns may reflect influence from consumer products or landfill leachate, where short-chain PFAS are more prevalent due to their mobility and widespread use. Dominance of short-chain PFAS has also been linked to wastewater treatment plants, which usually do not effectively remove these compounds (Ruyle, Pickard, et al., 2021; Antell et al., 2024). Samples from WWTPs generally display lower concentrations of PFAS compared to AFFF impacted sites (Joseph et al., 2023).

Cluster 4 was dominated by PFOS both in terms of absolute concentration and relative contribution to the total PFAS profile. PFOS made up on average 42.8% of the PFAS profile, with a median concentration of 90 ng/L in cluster 4. This strong PFOS signature may be related to contamination from aqueous film-forming foams (AFFF), commonly used in firefighting training areas and airports (Zhang et al., 2016; Hu et al., 2016; Ruyle, Pickard, et al., 2021; Reinikainen et al., 2022). Hu et al. (2016) determined that median PFOA and PFOS concentrations in groundwater impacted by AFFF near the contaminated areas are significantly higher compared to the WWTP effluent sources.

Cluster 5 exhibited considerably higher median concentrations across all PFAS compounds, particularly PFOS (200 ng/L), PFHxS (130 ng/L), and PFOA (160 ng/L). These elevated levels suggest a strong industrial influence, possibly linked to metal plating or manufacturing activities or nearby airport (Ehsan et al., 2023; Gamlin et al., 2024). The percentage composition data supports this interpretation, with PFOS accounting for 27.5% (mean) of the total PFAS load in Cluster 5, and PFHxS contributing 12.9%.

The alignment between absolute concentrations and relative compositions strengthens the interpretation of PFAS sources. Clusters with high PFOS and PFHxS concentrations also showed high relative contributions of these compounds, supporting hypotheses of industrial or AFFF-related contamination. Clusters dominated by short-chain PFAS in both metrics suggest consumer or landfill sources. The use of both concentration and composition data allows a more consistent insight into PFAS profiles and enhances confidence in source attribution.

## 5.4 Spatial Distribution

The spatial distribution of the clusters was mapped in ArcGis Pro (Figure 11) to explore potential geographic patterns and associations with known PFAS sources or environmental pathways.

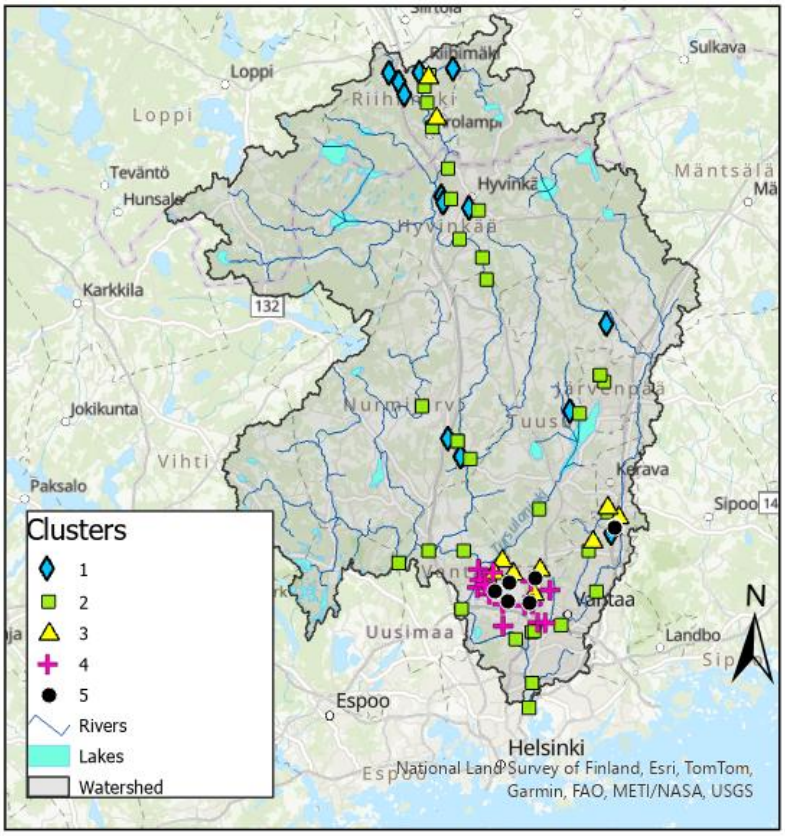


Figure 11. Regional distribution of five clusters in River Vantaa Watershed.

Cluster 1 samples are concentrated in the upstream part of the watershed with three samples also located in middle watershed rivers. Clusters 2 on the other hand is the largest cluster with 43 samples and is widely distributed across the northern, middle and southern parts of the watershed. Generally, WWTP, stormwater and GW are in the north part, river samples mainly in the south and middle while ditch and creek samples in the middle and southeastern areas. The spatial distribution supports the chemical profiles which suggested moderate PFAS levels from mixed sources like urban runoff. Samples grouped to cluster 3 are mainly located at the southern watershed, although several samples in this cluster originate from Riihimäki WWTP and Arolamminkoski River which are situated upstream.

Samples from cluster 4 and 5 are centred around the Helsinki-Vantaa airport, Southern part of the watershed. This spatial pattern supports the earlier interpretation based on PFAS profiles, where cluster 4 showed elevated levels of PFOS and PFHxS, consistent with AFFF-related PFAS contamination. Cluster 5 samples are also situated near the airport but are more tightly clustered in its immediate surroundings. However, two samples from the Karhuntassunoja ditch, located northeast of the airport in Kerava, also belong to cluster 5. Although these samples are not spatially close to the main airport cluster, they share a statistically similar PFAS profile, suggesting a common or similar source influence.

Cluster 4 and 5 showed highest median concentrations of the 9 PFAS. The highest PFAS concentrations were observed in samples collected from creeks and ditches compared to other matrices. This could be due to less water flow or point sources nearby. The ditches west of the airport flow towards the Vantaa River and the streams converge by Vantaankoski river sampling point where PFAS profile is more moderate than in the upstream ditches and creeks. This could suggest dilution or mixing with cleaner water sources. Figure 12 presents the spatial distribution of clusters alongside potential PFAS contamination sources. Some of the potential contamination sites are not active anymore e.g. some of the WWTPs are not in use these days and some landfills do not receive waste anymore but may still contribute to PFAS contamination due to legacy pollution and ongoing leaching.

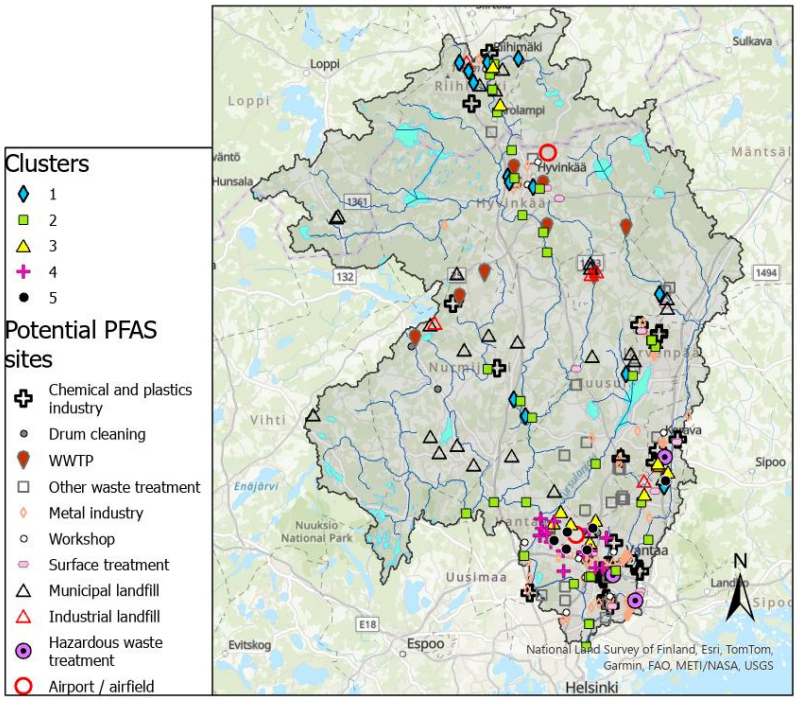


Figure 12. Clusters and MATTI-registered potential PFAS contamination sites.

Near cluster 1 samples, there are notably less potential PFAS sources than in the vicinity of other clusters supporting earlier findings that the PFAS profile doesn't suggest any distinct point sources. Based on the potential PFAS site's geographical location, PFAS contamination in cluster 1 samples could rise from a mixture of minor releases from landfills, Hyvinkää airfield, urban runoff and several waste treatment facilities.

Samples in cluster 2 are distributed largely in the watershed and therefore there are multiple MATTI-registered potential sources nearby. However, several river sampling locations don't have any plausible registered sources in the immediate vicinity. In the Northern watershed municipal landfills, metal industries, old WWTPs and chemical and plastics industry sites are recognized nearby sampling locations linked to cluster 2. The findings support the moderate mixed PFAS source profile of cluster 2. Also 4 WWTP individuals grouped to cluster 2 which is in line with WWTPs found in the vicinity.

Interestingly, cluster 3 is associated with WWTP water matrix, but on the map, cluster 3 individuals are not specifically located nearby old or existing WWTPs. However, Riihimäki WWTP is not in the MATTI-register, and several cluster 3 samples were from that WWTP. Other potential sites nearby cluster 3 locations are waste treatment, metal industry, drum cleaning and the airport.

One interesting finding is between creeks and ditches in clusters 4 and 5. While both clusters are linked to this category, the PFAS levels differ remarkably. Cluster 4 and 5 are notably near the airport which is the main factor contributing to the PFAS profile. Firefighting practices and AFFF-products have been practised around the airport (Reinikainen et al., 2022). By visually inspecting the potential sites, the difference between clusters 4 and 5 is that several ceased and existing metal industry facilities and a few chemical and plastics industry sites are located near cluster 4 individuals. On the other hand, a few waste treatment sites and municipal landfills are in the vicinity of some individuals in cluster 5.

## 5.5 Correlation Matrix

The Spearman rank-order correlation test was performed in R 4.4.1. Concentration data in ng/L were used and adjusted p-values (Holm's method) were obtained (Field et al., 2012). All p-values were below the 0.05 threshold, indicating statistically significant correlations. Spearman correlation coefficients (ρ) are presented in Table 8.

Table 7. Spearman correlation coefficients (ρ) for the 9 PFAS found in River Vantaa Watershed. ***p<0.001, **p<0.005, *p<0.05

| | **PFBA** | **PFBS** | **PFHpA** | **PFHxA** | **PFHxS** | **PFNA** | **PFOA** | **PFOS** | **PFPeA** |
|---|---|---|---|---|---|---|---|---|---|
| **PFBA** | 1.00 | 0.35** | 0.52*** | 0.51*** | 0.28** | 0.32** | 0.39** | 0.30** | 0.58*** |
| **PFBS** | | 1.00 | 0.85*** | 0.84*** | 0.86*** | 0.78*** | 0.84*** | 0.83*** | 0.77*** |
| **PFHpA** | | | 1.00 | 0.93*** | 0.84*** | 0.84*** | 0.84*** | 0.86*** | 0.89*** |
| **PFHxA** | | | | 1.00 | 0.81*** | 0.77*** | 0.77*** | 0.80*** | 0.93*** |
| **PFHxS** | | | | | 1.00 | 0.84*** | 0.78*** | 0.91*** | 0.76*** |
| **PFNA** | | | | | | 1.00 | 0.79*** | 0.88*** | 0.71*** |
| **PFOA** | | | | | | | 1.00 | 0.81*** | 0.74*** |
| **PFOS** | | | | | | | | 1.00 | 0.75*** |
| **PFPeA** | | | | | | | | | 1.00 |

One of the interesting features of the matrix is the consistently strong positive correlations among many compounds, particularly within the PFSA (PFBS, PFHxS, PFOS) and short-chain carboxylates (PFPeA, PFHxA, PFHpA) excluding PFBA. For example, PFBS and PFHxS show a correlation of 0.86, and PFHxA and PFHpA 0.93. High correlations suggest

that these compounds are frequently co-detected, likely due to shared sources such as WWTP effluent or AFFF applications, or due to similar environmental transport mechanisms. The strongest correlations were observed between PFHxA and PFHpA (ρ =0.93), PFPeA and PFHxA (ρ =0.93), and PFOS and PFHxS (ρ =0.91).

PFOS was strongly correlated with all other substances except PFBA. It indicates its widespread presence as legacy PFAS in aquatic matrices. Interestingly, PFOS shows strong correlations with PFHxS and PFNA, which suggests that they co-occur from a historical source that might still be affecting current PFAS contamination profiles. This could reflect persistent contamination from older industrial or firefighting sources, where both sulfonates and long-chain carboxylates were used.

PFBA showed weak and moderate correlations with other compounds (e.g., ρ ranged from 0.28 with PFHxS to 0.58 with PFPeA), suggesting somewhat independent behaviour. However, PFBA has high mobility and a short chain length (Silver et al., 2023), which means it might travel faster and further than other compounds. As a result, its weaker correlations may not necessarily imply a different source (Kibbey et al., 2020). PFBA may also be influenced by diffuse or distant inputs, such as atmospheric deposition or precursor degradation.

High correlation coefficients between PFHxA, PFHpA, and PFPeA (all >0.89) are notable. These compounds are often associated with modern consumer products or replacement chemicals, and their co-occurrence may point to recent or ongoing inputs, possibly from urban runoff or treated wastewater (Morethe et al., 2024). From a source tracking perspective, these correlation patterns support the idea that multiple overlapping sources are influencing the PFAS profiles observed. Correlation matrix highlights the complexity of environmental PFAS mixtures, where both legacy and emerging sources are often found simultaneously.

## 5.6 Summary of Results

Cluster summary table (Table 7) provides an overview of the composition of each cluster portraying the number of samples from each studied matrix, total number of samples, cluster location description, and relatively PFAS abundances (high/low) across clusters.

Table 8. Cluster summary table.

| CLUSTER | Composition (number of samples) | Total number of samples | Location | PFAS – Higher Relative Abundance | PFAS – Lower Relative Abundance | Source interpretation |
|---|---|---|---|---|---|---|
| **Cluster 1** | River (5)<br>Creeks/Ditches (1)<br>GW (5)<br>Stormwater (3)<br>WWTP (0) | 14 | Concentrated in the upper watershed, less potential PFAS sources nearby. All Stormwater/GW are in the upstream, River samples and the creek individual in the middle watershed. | | PFBA, PFHxS, PFBS, PFNA, PFOA, PFPeA, PFOS, PFHpA, PFHxA | Background or diffuse contamination from landfills, WWTPS and other industries |
| **Cluster 2** | River (25)<br>Creeks/Ditches (9)<br>GW (2)<br>Stormwater (3)<br>WWTP (4) | 43 | Widely distributed in watershed, north, middle and south. WWTP/Stormwater/ GW are in the north part, River locations mainly in the south and middle, ditch and creek in the middle/east part. | | PFHxS, PFBS, PFOS, PFHpA, PFNA, PFHxA, PFOA, PFPeA | Urban runoff or wastewaters, mixture from several sources, metal industry, upstream point sources |
| **Cluster 3** | River (1)<br>Creeks/Ditches (9)<br>GW (0)<br>Stormwater (0)<br>WWTP (4) | 14 | Mostly at the downstream part of the watershed, but WWTP and 1 river site located upstream near WWTP | PFBA, PFPeA, PFHxA, PFHpA | PFBS, PFHxS, PFNA, PFOA | Consumer products, WWTPs, municipal landfill leachate |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Cluster 4** | River (0)<br>Creeks/Ditches (19)<br>GW (0)<br>Stormwater (0)<br>WWTP (0) | 19 | Located north-west and south-east of the Helsinki-Vantaa Airport | PFOS, PFHxS, PFBS, PFOA, PFNA | PFBA | AFFF-related contamination |
| **Cluster 5** | River (0)<br>Creeks/Ditches (9)<br>GW (0)<br>Stormwater (0)<br>WWTP (0) | 9 | Mostly located near the Helsinki-Vantaa Airport and industrial areas south of the watershed. 3 Ditch individuals in Kerava (more upstream, to the east, no airport) | PFBA, PFPeA, PFHxA, PFHpA, PFNA, PFOA, PFBS, PFHxS, PFOS | | Industrial or airport-related contamination, landfill leachate |

The clustering of environmental samples reveals distinct PFAS profiles that align with land use and potential anthropogenic sources across the watershed. Cluster 1, composed entirely of non-WWTP samples from upstream riverine, groundwater, and stormwater environments, exhibits the lowest relative PFAS concentrations, suggesting minimal contamination and serving as a reference for background levels. In contrast, cluster 2, the largest and most spatially dispersed, includes a mix of WWTP, river, and ditch samples and shows an elevated proportion of long-chain PFAS such as PFOS and PFOA, indicative of diffuse urban and point-source inputs. Cluster 3, dominated by WWTP and ditch samples, is characterized by high proportions of short-chain PFAS (PFBA, PFPeA, PFHxA), likely reflecting modern substitution trends in treatment plant effluents. Clusters 4 and 5, located near the Helsinki-Vantaa Airport and industrial zones, show distinct PFAS signatures: cluster 4 is dominated by legacy compounds (PFOS, PFHxS), while cluster 5 exhibits the highest levels of all PFAS, suggesting intense and possibly mixed-source contamination.

# 6 Discussion

This discussion section interprets the results of multivariate statistical analysis and geospatial patterns to understand PFAS contamination in the River Vantaa watershed. Cluster profiles and their respective interpretations are compared to findings from existing source tracking literature. The analysis addresses key research questions, including the main patterns and trends in PFAS concentrations, the components contributing most to variance and their relation to known source fingerprints, and the most significant sources of contamination in the watershed. The chapter also considers the study's limitations and outlines directions for future research.

## 6.1 Cluster-Specific PFAS Signatures and Source Attribution

**Cluster 1: Background and Diffuse Inputs**

Cluster 1 represents a low-concentration, mixed PFAS signature that is typical for surface waters in urban or lightly impacted areas, or background sites. Concentration levels are clearly lower than those observed near identified point sources, supporting the interpretation of moderately clean or diffusely impacted environment. The mixed profile dominated by PFBA, PFOA, PFHxA, and PFOS resembles low-level WWTP effluent or diffuse urban inputs (Morethe et al., 2024). These low yet diverse concentrations are characteristic of a mixed profile or background noise in areas without a dominant point source, likely resulting from diffuse inputs such as urban runoff and diluted wastewater influence.

Cluster 1 did not show a clear PFAS profile linked to any specific potential source. It is common that locations without evident point sources nearby show a mixture of different signals (Benotti et al., 2020). Background PFAS concentrations often reflect overlapping source signals that are not statistically distinguishable (Zhang et al., 2016). The unclear profile of cluster 1 supports diffuse contamination as a likely contributor.

**Cluster 2: Urban and Mixed Industrial Influence**

Cluster 2, on the other hand, is characterized by widespread distribution across the watershed and sampling locations surrounded by multiple MATTI-registered potential sources, including municipal landfills, metal industry, WWTPs, chemical and plastics manufacturing. This pattern is typical of complex contamination scenarios, where surface waters are influenced by both point and nonpoint sources (Woodward et al., 2024). PFAS emissions from plastic production have been shown to contaminate soil and groundwater via airborne transport (Ehsan et al., 2023). In such environments with multiple overlapping sources, it is difficult to identify a single dominant contributor, and contamination may also result from long-range transport or diffuse urban inputs (Lin et al., 2021).

Chemically cluster 2 presents a moderate-concentration, PFCA-dominant (median ΣPFCA/ΣPFSA ≈ 2.5) fingerprint. This profile is consistent with contamination from multiple urban sources, particularly wastewater treatment plant (WWTP) effluent and diffuse urban runoff. The median total PFAS concentration Σ9PFAS in cluster 2 is 20.5 ng/L, which is significantly higher than in cluster 1 (4.64 ng/L), but lower than concentrations typically observed near strong point sources such as heavily polluted industrial waters (Ehsan et al., 2023). The concentration level is comparable to PFAS levels reported in contaminated Swedish rivers, where $\Sigma_{26}$PFAS averaged around 30 ng/L (Nguyen et al., 2017). PFCA-dominance excludes a typical PFOS-based AFFF source (Kibbey et al., 2021) and instead points toward urban WWTP effluent and overlapping diffuse sources. The composition of cluster 2 closely resembles WWTP signatures reported by Morethe et al. (2023), where PFBA, PFOA, and PFOS were consistently detected. Legacy compounds such as PFOS (median 3.00 ng/L) and PFOA (median 3.30 ng/L) are present at relatively elevated levels, suggesting a mixture of historical contamination combined with ongoing WWTP discharges (Rosen et al., 2022). Overall, cluster 2 reflects a WWTP-influenced profile shaped by multiple overlapping urban sources and legacy contamination, which is also supported by Figure 12 of geospatial mapping of potential sources and clusters.

**Cluster 3: WWTP and Landfill-Influenced Cluster**

Compared to clusters 1 and 2, cluster 3 shows a shift toward higher concentrations of short-chain PFAS, particularly PFBA, PFPeA and PFHxA, which together account for 56.8% on

average of the total $\Sigma_9$PFAS. The cluster's median ΣPFCA/ΣPFSA ratio (4.51) confirms a PFCA-dominant profile. This pattern is typical of wastewater treatment plant (WWTP) effluent, where short-chain PFCAs are present due to their resistance to conventional treatment processes (Weber et al., 2017; Antell et al., 2024). Cluster 3 is strongly associated with WWTP matrix, and it contains 4 samples from Riihimäki WWTP. Although not all samples are located directly adjacent to registered WWTPs, WWTP influence extends downstream or mixes with other urban inputs in downstream so WWTP related PFAS fingerprints can be observed downstream of the WWTP.

The median ΣPFAS concentration in cluster 3 is 94.74 ng/L, which is high but still within the lower range of WWTP effluent concentrations reported internationally e.g., $\Sigma_7$PFCA values of 140–459 ng/L in European municipal effluent (Dasu et al., 2021) and $\Sigma_9$PFAS ranging from 21 to 560 ng/L in Australian WWTPs (Gallen et al., 2018). Importantly, median PFOS level in cluster 3 (12.05 ng/L) exceeds the EU Water Framework Directive Environmental Quality Standard (EQS) (0.65 ng/L) for PFOS in inland surface waters, indicating potential ecological concern.

In addition to WWTPs, landfills may also contribute to cluster 3's PFAS profile. Short-chain compounds such as PFBA and PFBS are frequently dominant in untreated landfill leachate due to their high water solubility and mobility (Gallen et al., 2017). Reported leachate concentrations range from 33.3 to 780 ng/L (Ehsan et al., 2023), supporting the possibility of landfill influence in some samples. Geospatial analysis also confirms that Savio, Korttionmäki and Ruskeasanta landfill areas are situated upstream of a few cluster 3 sampling sites. Overall, cluster 3 reflects a WWTP-type fingerprint dominated by short-chain PFCAs, likely originating from a combination of consumer product emissions, WWTP discharges, and landfill leachate.

Riihimäki WWTP effluent samples and influent line 1 samples were grouped into the same cluster 3, while Kalteva influent, effluent and Riihimäki influent lines 2 and 3 were associated with cluster 2. The majority of influent coming to Riihimäki WWTP flows through influent line 1. Riihimäki WWTP receives both municipal and industrial wastewaters (Junttila et al., 2021). Influent line 1 measures significantly higher PFAS concentrations than lines 2 and 3. Since WWTP data includes both influent and effluent, the clustering suggests that treatment processes have not uniformly reduced PFAS levels, or that effluent still retains a similar chemical signature to influent.

**Clusters 4 and 5: High-Concentration airport- and industry-adjacent sites**

Clusters 4 and 5 contain the highest PFAS concentrations in the dataset with samples from creeks and ditches near Helsinki-Vantaa Airport. Cluster 4 has a median $\Sigma_9$PFAS concentration of 239.25 ng/L, while cluster 5 shows 1571 ng/L, clearly representing the most heavily contaminated sites in the study. These small water bodies possibly measure increased PFAS levels due to limited dilution and proximity to point sources. Airports and military sites are globally recognized PFAS hotspots due to the historical and ongoing use of AFFFs in firefighting practices (Silver et al., 2023). These activities have led to elevated PFAS concentrations in soil, groundwater, and surface water surrounding such facilities (Hu et al., 2016). For example, PFOS concentrations at Gingin Satellite Airfield in Australia reached 36,400 ng/L in a well near AFFF contamination (Gamlin et al., 2024).

Cluster 4 is PFSA-dominant (median ΣPFCA/ΣPFSA = 0.69), with PFOS (median: 90 ng/L) and PFHxS (26 ng/L) as the main compounds. This profile is typical of legacy AFFF contamination, particularly formulations based on electrochemical fluorination (ECF) processes (Gamlin et al., 2024; Kibbey et al., 2021). PFOS and PFHxS are widely used additives in AFFFs and are consistently elevated in water bodies impacted by firefighting activities (Hu et al., 2016; Ruyle, Pickard, et al., 2021; Lin et al., 2021; Woodward et al., 2024). In addition to AFFF, nearby metal plating and chemical industry facilities may contribute to the PFSA-heavy profile of cluster 4. PFOS has been used in anti-corrosion surface treatments in metal coating (Dasu et al., 2022). PFOS, PFPeA, and 6:2 FTS have been detected in waters impacted by metal plating (Zhang et al., 2016). Sometimes non-AFFF industrial samples (i.e. from metal finishing) can chemically resemble legacy AFFF profiles due to PFSA dominance (Kibbey et al., 2021).

Cluster 5, with the highest median Σ9PFAS concentration (1571 ng/L), likely reflects sampling sites located immediately adjacent to strong point sources. High ΣPFAS concentrations in surface waters are frequently associated with water bodies adjacent to airports and military facilities (Dasu et al., 2021; Ruyle et al., 2021a), and cluster 5 spatial concentration near the airport supports this. Although PFOS (200 ng/L) is the dominant compound, the cluster shows a higher PFCA contribution (median ΣPFCA/ΣPFSA = 1.32) than cluster 4. This may be explained by precursor degradation: AFFF formulations often contain fluorotelomer-based precursors that degrade into stable PFCAs in aquatic environments (Benotti et al., 2020). However, non-AFFF and AFFF classification based

solely on PFCA/PFSA ratios has been shown to be imprecise (accuracy 79.6%) in supervised machine learning models (Kibbey et al., 2021). Therefore, cluster 5 PFCA-rich profile may reflect intense precursor degradation or additional inputs from waste management sources.

Additional contributions may come from nearby waste treatment facilities and landfills. Landfill leachates often contain high levels of PFBA, PFBS, PFOS, and PFOA due to the high solubility and mobility of these compounds (Silver et al., 2023), and landfills are also known sources of airborne fluorotelomer alcohols (Zhang et al., 2016). This mixture may explain the high ΣPFAS concentrations in cluster 5, where AFFF signatures are mixed with waste-related inputs. Notably, samples from Karhuntassunoja ditch in Kerava, northeast of the airport, are grouped into cluster 5 despite geographic separation. This suggests a similar source influence, possibly another AFFF application site. The upper part of Karhuntassunoja receives runoff and leachate from landfills (Turtiainen et al., 2025), and PFOS concentrations (median 200 ng/L) and the presence of known landfill leachate indicator 5:3 FTCA in Turtiainen et al. (2025) study, support this interpretation.

Overall, clusters 4 and 5 show typical AFFF-related PFAS contamination characteristics, combined with industrial and waste management influences. Their distinct chemical profiles, cluster 4 with PFOS/PFHxS dominance and cluster 5 with higher PFCA contribution, illustrate the complexity of PFAS source tracking. The separation of clusters based on compound dominance demonstrates the value of multivariate analysis in distinguishing between potential sources.

## 6.2 Spatial Distribution and Source Influence

This study shows that the highest PFAS contamination in water occurs in the downstream sections of the watershed, particularly near the airport adjacent to the main streams, rather than at the very terminus. This pattern might be due to cumulative PFAS loading in downstream, higher population density, or simply the presence of more potential PFAS sources such as industrial facilities nearby. Several studies support the observation that the highest PFAS concentrations and loads are often found in downstream sections of watersheds or in geographically southern areas, where population density and the number of potential sources is typically higher (e.g. Lin et al., 2021; Nguyen et al., 2017; Woodward et al., 2024).

This pattern is largely attributed to the persistence of PFAS compounds and their tendency to accumulate from multiple overlapping sources. For example, Nguyen et al. (2017) found that PFAS concentrations in Swedish rivers were generally lower in the north compared to rivers along the southeastern and southwestern coasts. Similarly, in Finland, the highest PFAS concentrations in both surface waters and fish were observed in southern water bodies, which are more densely populated (Junttila et al., 2019). The increase could be attributed to additional inputs from other catchment areas. In theory, PFAS concentrations should decline with distance and time unless new sources or precursor transformations are present (Gamlin et al., 2024).

In this study, the five southern most (downstream) sampling locations (river sites) showed moderate PFAS profile that were grouped into cluster 2. These profiles differ significantly from those in clusters 3, 4, and 5, which are located adjacent to the main streams. Interestingly, typical cluster 2 profiles are also measured upstream of the heavily industrialized area and the airport, where most of cluster 4 and 5 samples are situated. This suggests that the PFAS loads from smaller water bodies in the industrial area may not be high enough to alter the overall chemical profile at the downstream end of the main river. Tracking PFAS loads in addition to concentrations could offer more understanding into this.

## 6.3 Correlation Patterns Among PFAS Compounds

The Spearman correlation matrix revealed consistently strong positive associations among several PFAS compounds, particularly within the PFSA group (PFBS, PFHxS, PFOS) and short-chain carboxylates (PFPeA, PFHxA), excluding PFBA. These patterns are in line with the findings of Silver et al. (2023), who reported positive correlations among shorter-chain PFAS, with PFHxA showing correlations ($\rho > 0.49$) with PFBA, PFBS, PFPeA, PFHxS, PFHpA, PFOA, PFOS, and PFNA. In our study, PFHxA had ($\rho > 0.51$) with all compounds and PFHxA–PFHpA showed the highest correlation ($\rho = 0.93$). This suggests co-occurrence due to shared sources or similar environmental behaviour.

PFOS also correlated strongly with PFHxS, PFBS and PFNA in our dataset, which corresponds with Silver et al. (2023) interpretation of legacy contamination from historical sources such as AFFFs and industrial discharges. PFOS and PFHxS have also been linked with southern latitude and densely populated areas (Nguyen et al., 2017).

Morethe et al. (2023) observed similar correlations among short-chain PFCAs and PFSAs in South African WWTP samples. These correlations were interpreted as indicative of shared sources such as domestic discharge, industrial inputs, and precursor degradation. Correlations for those compounds were strongly positive in our study as well except for PFBA which showed only moderate association (e.g. $\rho = 0.58$ with PFPeA), suggesting distinct behaviour or diffuse inputs. PFBA also stood out in Silver et al. (2023) and Nguyen et al. (2017) studies due to its weaker correlations with other PFAS. Silver et al. attributed PFBA's prevalence to its high mobility and potential for long-range transport via precipitation and atmospheric deposition. They also suggested PFBA's linked to non-agricultural sources.

Although our study did not include fluorotelomers, Morethe et al. (2023) found strong correlations between 6:2 FTS and short-chain PFAS, suggesting transformation within WWTP processes. Interestingly, their seasonal analysis also showed variation in PFBA–PFOA correlations in dry and wet seasons, highlighting the importance of temporal factors in source tracking. While our study did not include seasonal correlation analysis, this could be a valuable direction for future research.

Across all three studies, a recurring theme is the strong internal correlation among most PFAS compounds except PFBA. Our findings, like those of Morethe et al. (2024), Nguyen et al. (2017) and Silver et al. (2023) support the interpretation that multiple overlapping sources such as WWTP effluent, AFFF applications, and waste management activities contribute to the observed PFAS profiles. The correlation patterns across studies underscore the complexity of PFAS mixtures in aquatic environments and emphasize the value of multivariate approaches in source tracking.

## 6.4 Limitations

While the principal component analysis and clustering provided valuable interpretation into PFAS source profiles, several limitations must be considered when interpreting these results. First spatial coverage of sampling locations is uneven (Figure 3), with underrepresentation in the western part of the watershed and limited sampling at major stream junctions. More comprehensive monitoring at these junctions would improve the ability to compare geospatial differences and strengthen source attribution. Secondly, a major limitation of this

study is the limited compound selection. Only nine PFAS were consistently detected across all datasets with a detection frequency above 80% and thus were selected for further analysis. This narrow compound range enables only a simple PFAS profile, which limits source attribution depth. Several widely used and commonly detected PFAS compounds with relevance for source tracking, like PFDA (Lin et al., 2021), PFHpS (Carlson and Andersen, 2025) and emerging substances like TFA (Liu et al., 2019) and GenX (Pétré et al., 2022) were missing from the selected sample analyses.

Additionally, our study did not cover fluorotelomer carboxylic acids, which are frequently dominant in landfill leachate, known degradation products of FTS, and precursors to short-chain PFCAs, some of which are detectable post-treatment but not in raw water (Joseph et al., 2023). Similarly, important precursor compounds such as fluorotelomer sulfonates (FTS) e.g. 6:2 FTS (Benotti et al., 2020) and sulfonamidoacetic acids (Joseph et al., 2023), which are relevant for precursor tracking, were missing from our analysis. The absence of 6:2 FTS in analysis complicates the assessment of AFFF-related sources and the influence of wastewater treatment plants (Woodward et al., 2024).

For example, Turtiainen et al. (2025) reported high concentrations of 5:3 FTCA (a compound typically associated with landfill leachate) in Karhuntassunoja ditch, yet it was not included in our compound list. Same location is included in our study, but the essential FTCA signature is missing. However, due to other fingerprints and geospatial mapping of potential sources revealed landfill was revealed as a potential source for Karhuntassunoja sample that belonged to cluster 5.

While this study explored PFAS source differentiation using multivariate and geospatial methods, several promising approaches could not be fully implemented due to time and resource constraints. Compositional Data Analysis (CoDA) through Centred-logarithmic ratio (CLR) -transformed PCA is a way to address the limitations of concentration-based analyses, which tend to be dominated by magnitude differences. CoDA could provide another framework for interpreting PFAS profiles based on relative proportions, potentially revealing source-related patterns that absolute concentrations may hide (Aitchison, 1982; Filzmoser et al., 2019; Ruyle, Pickard, et al., 2021). Also, additional geospatial methods, including buffer analysis and watershed delineation using ArcGIS tools could support more precise identification of PFAS transport mechanisms and source zones across the watershed.

Lastly, knowing the composition of an original formulation or site profile is not sufficient to confirm a source, as the differing mobilities of PFAS compounds can lead to highly variable profiles in environmental samples within a contaminated area (O'Carroll et al., 2020). This was visible in the results of this study as well: no cluster could be uniquely linked to a single source with these methods. Several clusters had profiles that could be explained by multiple industries. The compounds in the analysis (PFBA, PFPeA, PFHxA, PFHpA, PFOA, PFNA, PFBS, PFHxS and PFOS) are widely used and commonly detected in the environment, which limits their specificity for source identification. Compounds such as legacy PFOS, PFOA, and short-chain PFCA have been released from multiple sources, and their ratios can shift under environmental processes (Kibbey et al., 2020).

## 6.5 Recommendations and Future Directions

This thesis has reviewed current methods and challenges in PFAS source tracking, highlighting the complexity of tracking contamination in diverse environmental settings. While progress has been made in environmental forensics and PFAS source attribution techniques, several research gaps remain that limit the reliability and comparability of source tracking efforts across the field.

Recent studies emphasize the need to expand detection capabilities beyond conventional targeted analysis. High-resolution mass spectrometry (HRMS) allows for the identification of a broader range of PFAS compounds, including unknowns and transformation products (Benotti et al., 2020). The Total Oxidizable Precursor (TOP) assay can detect precursor compounds missed by targeted analysis and, when combined with PCA, enhance source apportionment (Antell et al., 2024). Standardizing PFAS analytical methods and terminology would help researchers compare results across studies and regions (Charbonnet et al., 2021). Improved transparency such as open-access databases and industry disclosure of PFAS product composition would further support effective source tracking and data sharing (Gamlin et al., 2024).

Environmental transformation of PFAS compounds further complicates source tracking. Some compounds persist for decades, move at different speeds, or transform into other PFAS types (Ehsan et al., 2023; Kibbey et al., 2021). Integrating chemical data with environmental transport information is another important direction. Conceptual site models (CSMs) can

help interpret these changes by combining PFAS measurements with hydrogeology, land use, and discharge data (Charbonnet et al., 2021). Improving these models requires considering long-term PFAS behaviour, including groundwater–surface water exchange and interactions with soil that may slow down movement (Lyu et al., 2022; Rafiei & Nejadhashemi, 2023). Machine learning offers promising tools for analysing complex PFAS datasets and improving source attribution, especially when combined with large and diverse training data (Kibbey et al., 2021; Joseph et al., 2023; Stults et al., 2023).

Uncertainty remains a major challenge, especially when combining chemical, spatial, and statistical data. Quantifying and communicating this uncertainty are essential for improving the reliability of source attribution (Liu et al., 2019; Gamlin et al., 2024). Building multiple lines of evidence (LOE) is one way to address this. Combining indicators like concentration trends, geospatial patterns, and site-specific conditions can strengthen interpretation. When LOEs conflict, they should be examined carefully (Gamlin et al., 2024).

Calculating PFAS mass loads, rather than relying only on concentrations, can offer a more meaningful understanding into contaminant movement. Approaches like the Time-of-Travel Sampling Method (Woodward et al., 2024) can help estimate PFAS loads from specific sources, such as wastewater treatment plants or military sites. Temporal monitoring is also needed to capture seasonal variations, episodic releases, and long-term trends (Pétré et al., 2022; Carlson and Andersen, 2025), which are crucial for distinguishing persistent background levels from active sources. Including emerging fluorinated alternatives like GenX and identifying sentinel compounds could also support more targeted source tracking (Charbonnet et al., 2021).

Finally, future research should actively involve local communities and stakeholders in the development of PFAS monitoring and mitigation strategies. This approach helps ensure that source tracking efforts reflect local concerns and contribute to promoting public health and environmental justice in affected areas.

# 7 Conclusions

This thesis represents the first PFAS source tracking effort in the River Vantaa watershed, applying multivariate statistical methods to characterize contamination patterns and identify potential sources. The research reached its core objectives by revealing key trends in PFAS concentrations, identifying dominant components contributing to variance, and linking chemical profiles to known source categories.

The results demonstrate that PFAS source tracking in this watershed is complex due to overlapping contamination signals, compound specific environmental behaviour, and limited compound range. However, the use of Hierarchical Clustering on Principal Components (HCPC) enabled the separation of distinct PFAS profiles into five clusters. Cluster 1 reflects background levels from diffuse sources. Clusters 2 and 3 show influence from wastewater treatment plants (WWTPs), urban runoff, landfill leachate and potentially metal industry. Clusters 4 and 5 clearly reflect AFFF-related contamination near Helsinki-Vantaa Airport and a combination of heavy industry and industrial landfill leachate. The successful separation of clusters based on compound dominance (e.g., PFHxA, PFHpA, PFOA, PFHxS, PFOS) illustrates the value of unsupervised pattern recognition in environmental source tracking. The first two principal components explained 87.4% of the variance, confirming strong compositional patterns in the dataset.

PFAS concentrations varied widely across the watershed, with Cluster 5 showing the highest median $\Sigma_9$PFAS concentration of 1571 ng/L, while Cluster 1 exhibited background levels at 4.64 ng/L. Clusters 4 and 5 were concentrated near Helsinki-Vantaa Airport and adjacent industrial zones, while Cluster 1 was located upstream in areas with fewer potential PFAS sources. The spatial distribution and comparison with MATTI-registered potential PFAS sources, support the interpretation that airport firefighting activities and industrial operations are major contributors to PFAS contamination in the southern watershed.

These findings are consistent with previous studies that have identified airports, WWTPs, and landfills as major PFAS contributors. However, this study also illustrates the challenge of distinguishing between overlapping sources, especially when common compounds like PFOS and PFOA are present across multiple clusters. The uneven spatial coverage and restricted compound range were major limitations, highlighting the need for expanded

monitoring and inclusion of emerging PFAS for more precise source attribution. In particular, the absence of central precursors and emerging compounds such as 6:2 FTS, FTCA, and GenX limited the ability to differentiate between overlapping source signals.

While geospatial analysis in this study was limited, the observed spatial patterns particularly the concentration of PFAS in southern, industrialized areas are consistent with broader regional trends. The correlation analysis further supports the interpretation that multiple overlapping sources, including WWTP effluent, AFFF applications, and consumer products, contribute to the observed PFAS profiles. Strong positive correlations among most PFAS compounds indicate co-occurrence from shared sources. PFBA showed weaker correlations, suggesting distinct environmental behaviour or diffuse inputs.

From a practical perspective, the findings support the prioritization of airport-adjacent creeks, WWTPs, and landfill areas for further monitoring and PFAS removal efforts. The study contributes to the growing PFAS research area by demonstrating how multivariate analysis can be used to resolve complex contamination profiles in urban watersheds. It also emphasizes the need for improved compound coverage, temporal monitoring, and load-based assessments to enhance source identification. Several challenges and research gaps continue to limit the reliability and comparability of current approaches. Integrating chemical data with environmental transport models and applying machine learning techniques can further improve the interpretation of complex datasets and strengthen source attribution. Moreover, ongoing monitoring that captures temporal variations and includes emerging and sentinel PFAS will support more comprehensive assessments.

Overall, the application of multivariate statistical methods combined with geospatial analysis represents a valuable first step in using existing data to assess PFAS source and patterns across the River Vantaa watershed. This integrated approach, although limited by the number of individual PFAS, revealed spatial trends in PFAS concentrations and compositional profiles and highlighted its potential for guiding future source tracking and monitoring efforts.

# References


Ahrens, L. and Bundschuh, M. 2014. Fate and effects of poly- and perfluoroalkyl substances in the aquatic environment: A review. *Environmental Toxicology and Chemistry*. **33**(9), pp.1921–1929.

Aitchison, J. 1982. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)*. **44**(2), pp.139–177.

Antell, E.H., Yi, S., Olivares, C.I., Ruyle, B.J., Kim, J.T., Tsou, K., Dixit, F., Alvarez-Cohen, L. and Sedlak, D.L. 2024. The Total Oxidizable Precursor (TOP) Assay as a Forensic Tool for Per- and Polyfluoroalkyl Substances (PFAS) Source Apportionment. *ACS ES and T Water*. **4**(3), pp.948–957.

Beaudry, C., Lefebvre, R., Rivard, C. and Cloutier, V. 2018. Conceptual model of regional groundwater flow based on hydrogeochemistry (Montérégie Est, Québec, Canada). *Canadian Water Resources Journal*. **43**(2), pp.152–172.

Beisner, K.R., Travis, R.E., Alvarez, D.A., Barber, L.B., Fleck, J.A. and Jasmann, J.R. 2024. Temporal variability and sources of PFAS in the Rio Grande, New Mexico through an arid urban area using multiple tracers and high-frequency sampling. *Emerging Contaminants*. **10**(3).

Benotti, M.J., Fernandez, L.A., Peaslee, G.F., Douglas, G.S., Uhler, A.D. and Emsbo-Mattingly, S. 2020. A forensic approach for distinguishing PFAS materials. *Environmental Forensics*. **21**(3–4), pp.319–333.

Box, G.E. and Cox, D.R. 1982. An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association*. **77**(377), pp.209–210.

Brendel, S., Fetter, É., Staude, C., Vierke, L. and Biegel-Engler, A. 2018. Short-chain perfluoroalkyl acids: environmental concerns and a regulatory strategy under REACH. *Environmental Sciences Europe*. **30**(1).

Briassoulis, G. and Ilia, S. 2025. Exposure to Per- and Polyfluoroalkyl Substances (PFASs) in Healthcare: Environmental and Clinical Insights. , pp.1–26.

Carlson, G.L. and Andersen, M. 2025. Tracking environmental contamination from multiple sources of per- and polyfluoroalkyl substances. *Environmental Research*. **276**(March), Article 121470.

Carlson, G.L. and Tupper, S. 2020. Ski wax use contributes to environmental contamination by per- and polyfluoroalkyl substances. *Chemosphere*. **261**.

Charbonnet, J.A., Rodowa, A.E., Joseph, N.T., Guelfo, J.L., Field, J.A., Jones, G.D., Higgins, C.P., Helbling, D.E. and Houtz, E.F. 2021. Environmental Source Tracking of Per- And Polyfluoroalkyl Substances within a Forensic Context: Current and Future Techniques. *Environmental Science and Technology*. **55**(11), pp.7237–7245.

Cloutier, V., Lefebvre, R., Therrien, R. and Savard, M.M. 2008. Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *Journal of Hydrology*. **353**(3–4), pp.294–313.

Council of the EU 2025. *Water pollution: Council and Parliament reach provisional deal to update priority substances in surface and ground waters* [Online]. [Accessed 19 August 2025]. Available from: www.consilium.europa.eu/press.

Cousins, I.T., Johansson, J.H., Salter, M.E., Sha, B. and Scheringer, M. 2022. Outside the Safe Operating Space of a New Planetary Boundary for Per- and Polyfluoroalkyl Substances (PFAS). *Environmental Science and Technology*. **56**(16), pp.11172–11179.

Dasu, K., Xia, X., Siriwardena, D., Klupinski, T.P. and Seay, B. 2022. Concentration profiles of per- and polyfluoroalkyl substances in major sources to the environment. *Journal of Environmental Management*. **301**(May 2021), Article 113879.

Davis, J.C.. 1986. *Statistics and data analysis in geology*. New York: John Wiley & Sons Inc.

Dong, J., Tsai, G. and Olivares, C.I. 2024. Prediction of 35 Target Per- and Polyfluoroalkyl Substances (PFASs) in California Groundwater Using Multilabel Semisupervised Machine Learning. *ACS ES and T Water*. **4**(3), pp.969–981.

ECHA 2018. Committee for Risk Assessment (RAC) Committee for Socio-economic Analysis (SEAC) Background document. . **1**(June), p.385.

ECHA 2023. PFAS restriction proposal submitted by five national authorities. *ECHA/NR/25/24*. [Online]. [Accessed 7 September 2025]. Available from: https://echa.europa.eu/-/echa-publishes-updated-pfas-restriction-proposal.

EFSA 2020. Risk to human health related to the presence of perfluoroalkyl substances in food. *EFSA Journal*. **18**(9).

EFSA 2018. Risk to human health related to the presence of perfluorooctane sulfonic acid and perfluorooctanoic acid in food. *EFSA Journal*. **16**(12).

Ehsan, M.N., Riza, M., Pervez, M.N., Khyum, M.M.O., Liang, Y. and Naddeo, V. 2023. Environmental and health impacts of PFAS: Sources, distribution and sustainable management in North Carolina (USA). *Science of the Total Environment*. **878**(February), Article 163123.

EU Commission 2022. Amending Directive 2000/60/EC establishing a framework for Community action in the field of water policy, Directive 2006/118/EC on the protection of groundwater against pollution and deterioration and Directive 2008/105/EC on environmental quality standard. *540 Final - 543 Final*. **0344**, pp.1–81.

European Commission 2022. *Proposal for a directive of the European Parliament and of the Council amending Directive 2000/60/EC* [Online]. Brussels. [Accessed 26 September 2025]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0540.

European Parliament and Council 2020. DIRECTIVE (EU) 2020/2184 OF THE COUNCIL of 16 December 2020 on the quality of water intended for human consumption (recast). *Official Journal of the European Union*. **L435**(December), pp.1–62.

Fenton, S.E., Ducatman, A., Boobis, A., DeWitt, J.C., Lau, C., Ng, C., Smith, J.S. and Roberts, S.M. 2021. Per- and Polyfluoroalkyl Substance Toxicity and Human Health Review: Current State of Knowledge and Strategies for Informing Future Research. *Environmental Toxicology and Chemistry*. **40**(3), pp.606–630.

Field, A., Miles, J. and Field, Z. 2012. *Discovering statistics using R*. London: SAGE Publications Ltd.

Filzmoser, P., Hron, K. and Templ, M. 2019. *Applied compositional data analysis: with worked examples in R* (P. Diggle, U. Gather, & S. Zeger, eds.). Springer.

Finnish Ministry of Social Affairs and Health 2015. Decree 1352/2015 on the quality requirements and monitoring of drinking water *In*: Helsinki. Available from: https://www.finlex.fi/eli?uri=http://data.finlex.fi/eli/sd/2015/1352/ajantasa/2023-01-03/fin.

FPP 2023. The Forever Pollution Project: Journalists tracking PFAS across Europe. Available from: https://foreverpollution.eu/.

Gallen, C., Drage, D., Eaglesham, G., Grant, S., Bowman, M. and Mueller, J.F. 2017. Australia-wide assessment of perfluoroalkyl substances ( PFASs ) in landfill leachates. *Journal of Hazardous Materials*. **331**, pp.132–141.

Gallen, C., Eaglesham, G., Drage, D., Nguyen, T.H. and Mueller, J.F. 2018. A mass estimate of perfluoroalkyl substance (PFAS) release from Australian wastewater treatment plants. *Chemosphere*. **208**, pp.975–983.

Gamlin, J., Newell, C.J., Holton, C., Kulkarni, P.R., Skaggs, J., Adamson, D.T., Blotevogel, J. and Higgins, C.P. 2024. Data Evaluation Framework for Refining PFAS Conceptual Site Models. *Groundwater Monitoring and Remediation*. **44**(4), pp.53–66.

Geosyntec 2018. Assessment of the Chemical And Spatial Distribution of PFAS in the Cape Fear River. . (September), p.82.

Guelfo, J.L. and Adamson, D.T. 2018. Evaluation of a national data set for insights into sources, composition, and concentrations of per- and polyfluoroalkyl substances (PFASs) in U.S. drinking water. *Environmental Pollution*. **236**, pp.505–513.

Güler, C., Thyne, G.D., McCray, J.E. and Turner, A.K. 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeology Journal*. **10**(4), pp.455–474.

Houtz, E.F., Sutton, R., Park, J.S. and Sedlak, M. 2016. Poly- and perfluoroalkyl substances in wastewater: Significance of unknown precursors, manufacturing shifts, and likely AFFF impacts. *Water Research*. **95**, pp.142–149.

Hu, X.C., Andrews, D.Q., Lindstrom, A.B., Bruton, T.A., Schaider, L.A., Grandjean, P., Lohmann, R., Carignan, C.C., Blum, A., Balan, S.A., Higgins, C.P. and Sunderland, E.M. 2016. Detection of Poly- and Perfluoroalkyl Substances (PFASs) in U.S. Drinking Water Linked to Industrial Sites, Military Fire Training Areas, and Wastewater Treatment Plants. *Environmental Science and Technology*.

Husson, F., Josse, A.J., Jérôme, A. and Agrocampus, P. 2010. *Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data?* [Online]. Available from: http://www.agrocampus-ouest.fr/math/.

IARC 2024. Carcinogenicity of perfluorooctanoic acid and perfluorooctanesulfonic acid. *The Lancet. Oncology*. **25**(1), pp.16–17.

Johansson, J.H., Shi, Y., Salter, M. and Cousins, I.T. 2018. Spatial variation in the atmospheric deposition of perfluoroalkyl acids: source elucidation through analysis of isomer patterns. *Environmental Science: Processes and Impacts*. **20**(7), pp.997–1006.

Joseph, N.T., Schwichtenberg, T., Cao, D., Jones, G.D., Rodowa, A.E., Barlaz, M.A., Charbonnet, J.A., Higgins, C.P., Field, J.A. and Helbling, D.E. 2023. Target and Suspect Screening

Integrated with Machine Learning to Discover Per- and Polyfluoroalkyl Substance Source Fingerprints. *Environmental Science and Technology*. **57**(38), pp.14351–14362.

Junttila, V., Vähä, E., Perkola, N., Räike, A., Siimes, K., Mehtonen, J., Kankaanpää, H. and Mannio, J. 2019. PFASs in Finnish rivers and fish and the loading of PFASs to the Baltic Sea. . **11**(4), pp.1–15.

Junttila, V., Virkkunen, H., Vahtera, H., Männynsalo, J., Högmander, P., Perkola, N. and Mehtonen, J. 2021. Vantaanjoen PFAS-hanke Loppuraportti.

Karamat, A., Tehrani, R., Foster, G.D. and Van Aken, B. 2024. Plant responses to per- and polyfluoroalkyl substances (PFAS): a molecular perspective. *International Journal of Phytoremediation*. **26**(2), pp.219–227.

Kibbey, T.C.G., Jabrzemski, R. and O'Carroll, D.M. 2021. Source allocation of per- and polyfluoroalkyl substances (PFAS) with supervised machine learning: Classification performance and the role of feature selection in an expanded dataset. *Chemosphere*. **275**, Article 130124.

Kibbey, T.C.G., Jabrzemski, R. and O'Carroll, D.M. 2020. Supervised machine learning for source allocation of per- and polyfluoroalkyl substances (PFAS) in environmental samples. *Chemosphere*. **252**, Article 126593.

Kivimäki, A.-L., Rautio, A., Korkka-Niemi, K., Brander, M., Nygård, M., Vahtera, H., Karhu, J., Salonen, V., Kiirinki, M. and Lahti, K. 2013. *Vantaanjoen ja sen sivujokien hydrauliset yhteydet pohjavesimuodostumiin ja vaikutukset veden laatuun*.

Kurwadkar, S., Dane, J., Kanel, S.R., Nadagouda, M.N., Cawdrey, R.W., Ambade, B., Struckhoff, G.C. and Wilkin, R. 2022. Per- and polyfluoroalkyl substances in water and wastewater: A critical review of their global occurrence and distribution. *Science of the Total Environment*. **809**.

Langberg, H.A., Arp, H.P.H., Breedveld, G.D., Slinde, G.A., Høiseter, Å., Grønning, H.M., Jartun, M., Rundberget, T., Jenssen, B.M. and Hale, S.E. 2021. Paper product production identified as the main source of per- and polyfluoroalkyl substances (PFAS) in a Norwegian lake: Source and historic emission tracking. *Environmental Pollution*. **273**, Article 116259.

Lê, S., Husson, F. and Josse, J. 2008. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*. **25**(1).

Li, F., Duan, J., Tian, S., Ji, H., Zhu, Y., Wei, Z. and Zhao, D. 2020. Short-chain per- and polyfluoroalkyl substances in aquatic systems: Occurrence, impacts and treatment. *Chemical Engineering Journal*. **380**(June 2019).

Lin, Y., Capozzi, S.L., Lin, L. and Rodenburg, L.A. 2021. Source apportionment of perfluoroalkyl substances in Great Lakes fish. *Environmental Pollution*. **290**(April), p.118.

Liu, Y., D'Agostino, L.A., Qu, G., Jiang, G. and Martin, J.W. 2019. High-resolution mass spectrometry (HRMS) methods for nontarget discovery and characterization of poly- and per-fluoroalkyl substances (PFASs) in environmental and human samples. *TrAC - Trends in Analytical Chemistry*. **121**, pp.115–121.

Lyu, X., Xiao, F., Shen, C., Chen, J., Park, C.M., Sun, Y., Flury, M. and Wang, D. 2022. Per- and Polyfluoroalkyl Substances (PFAS) in Subsurface Environments: Occurrence , Fate , Transport , and Research Prospect.

Ma, X., Shan, G., Chen, M., Zhao, J. and Zhu, L. 2018. Riverine inputs and source tracing of perfluoroalkyl substances (PFASs) in Taihu Lake, China. *Science of the Total Environment*. **612**, pp.18–25.

Martin, D., Munoz, G., Mejia-Avendaño, S., Duy, S.V., Yao, Y., Volchek, K., Brown, C.E., Liu, J. and Sauvé, S. 2019. Zwitterionic, cationic, and anionic perfluoroalkyl and polyfluoroalkyl substances integrated into total oxidizable precursor assay of contaminated groundwater. *Talanta*. **195**(September 2018), pp.533–542.

Morethe, M.F., Mpenyana-Monyatsi, L., Daso, A.P. and Okonkwo, O.J. 2024. Unveiling the hidden threat: spatiotemporal trends and source apportionments of per-and polyfluorinated alkyl substances in wastewater treatment plants in South Africa. *Water Science and Technology*. **89**(1), pp.71–88.

Murphy, B. and Morrison, R. 2014. *Introduction to Environmental Forensics*.

Nguyen, M.A., Wiberg, K., Ribeli, E., Josefsson, S., Futter, M., Gustavsson, J. and Ahrens, L. 2017. Spatial distribution and source tracing of per- and polyfluoroalkyl substances (PFASs) in surface water in Northern Europe. *Environmental Pollution*. **220**, pp.1438–1446.

O'Carroll, D.M., Jeffries, T.C., Lee, M.J., Le, S.T., Yeung, A., Wallace, S., Battye, N., Patch, D.J., Manefield, M.J. and Weber, K.P. 2020. Developing a roadmap to determine per- and polyfluoroalkyl substances-microbial population interactions. *Science of the Total Environment*. **712**.

OECD 2021. Reconciling Terminology of the Universe of Per- and Polyfluoroalkyl Substances: Recommendations and Practical Guidance Series on Risk Management No.61. *Organisation for Economic Co-operation and Development*. (61), pp.1–44.

OECD 2018. SUMMARY REPORT ON UPDATING THE OECD 2007 LIST OF PER- AND POLYFLUOROALKYL SUBSTANCES (PFASs). *Series on Risk Management*. (39).

Pereira, H., Ullberg, M., Kleja, D.B., Gustafsson, J.P. and Ahrens, L. 2018. Sorption of perfluoroalkyl substances (PFASs) to an organic soil horizon – Effect of cation composition and pH. *Chemosphere*. **207**, pp.183–191.

Pétré, M.A., Salk, K.R., Stapleton, H.M., Ferguson, P.L., Tait, G., Obenour, D.R., Knappe, D.R.U. and Genereux, D.P. 2022. Per- and polyfluoroalkyl substances (PFAS) in river discharge: Modeling loads upstream and downstream of a PFAS manufacturing plant in the Cape Fear watershed, North Carolina. *Science of the Total Environment*. **831**, Article 154763.

Pyy, O., Haavisto, T., Niskala, K. and Silvola, M. 2013. *Pilaantuneet maa-alueet Suomessa* [Online]. [Accessed 6 August 2025]. Available from: www.syke.fi/julkaisut.

Qi, Y., Huo, S., Xi, B., Hu, S., Zhang, J. and He, Z. 2016. Spatial distribution and source apportionment of PFASs in surface sediments from five lake regions, China. *Scientific Reports*. **6**(February), pp.1–11.

Rafiei, V. and Nejadhashemi, A.P. 2023. Watershed scale PFAS fate and transport model for source identification and management implications. *Water Research*. **240**(May), p.120073.

Reinikainen, J., Perkola, N., Äystö, L. and Sorvari, J. 2022. The occurrence, distribution, and risks of PFAS at AFFF-impacted sites in Finland. *Science of the Total Environment*. **829**.

Reinikainen, J., Perkola, N. and Takala, M. 2019. Perfluorattujen alkyyliyhdisteiden ympäristötutkimukset ja riskinarviointi. *Suomen ympäristökeskuksen raportteja*.

Rosen, E.M., Kotlarz, N., Knappe, D.R.U., Lea, C.S., Collier, D.N., Richardson, D.B. and Hoppin, J.A. 2022. Drinking Water–Associated PFAS and Fluoroethers and Lipid Outcomes in the GenX Exposure Study. *Environmental Health Perspectives*. **130**(9), pp.1–9.

Russell, M.H., Berti, W.R., Szostek, B. and Buck, R.C. 2008. Investigation of the biodegradation potential of a fluoroacrylate polymer product in aerobic soils. *Environmental Science Technology*. **42**, pp.800–802.

Ruyle, B.J., Pickard, H.M., LeBlanc, D.R., Tokranov, A.K., Thackray, C.P., Hu, X.C., Vecitis, C.D. and Sunderland, E.M. 2021. Isolating the AFFF signature in coastal watersheds using oxidizable PFAS precursors and unexplained organofluorine. *Environmental Science and Technology*. **55**(6), pp.3686–3695.

Ruyle, B.J., Thackray, C.P., McCord, J.P., Strynar, M.J., Mauge-Lewis, K.A., Fenton, S.E. and Sunderland, E.M. 2021. Reconstructing the Composition of Per- And Polyfluoroalkyl Substances in Contemporary Aqueous Film-Forming Foams. *Environmental Science and Technology Letters*. **8**(1), pp.59–65.

Salim, I., Sajjad, R.U., Paule-Mercado, M.C., Memon, S.A., Lee, B.Y., Sukhbaatar, C. and Lee, C.H. 2019. Comparison of two receptor models PCA-MLR and PMF for source identification and apportionment of pollution carried by runoff from catchment and sub-watershed areas with mixed land cover in South Korea. *Science of the Total Environment*. **663**, pp.764–775.

Schaider, L.A., Ackerman, J.M. and Rudel, R.A. 2016. Septic systems as sources of organic wastewater compounds in domestic drinking water wells in a shallow sand and gravel aquifer. *Science of the Total Environment*. **547**, pp.470–481.

Siimes, K., Vähä, E., Junttila, V., Lehtonen, K.K. and Mannio, J. 2019. *Haitalliset aineet Suomen vesissä: tilanne ja seurannan suuntaviivat*.

Silver, M., Phelps, W., Masarik, K., Burke, K., Zhang, C., Schwartz, A., Wang, M., Nitka, A.L., Schutz, J., Trainor, T., Washington, J.W. and Rheineck, B.D. 2023. Prevalence and Source Tracing of PFAS in Shallow Groundwater Used for Drinking Water in Wisconsin, USA. *Environmental Science and Technology*. **57**(45), pp.17415–17426.

Smaili, H. and Ng, C. 2022. Adsorption as a remediation technology for short-chain per- and polyfluoroalkyl substances (PFAS) from water - a critical review. *Environmental Science: Water Research and Technology*. **9**(2), pp.344–362.

Smith, S.J., Keane, C., Ahrens, L. and Wiberg, K. 2023. Integrated Treatment of Per- and Polyfluoroalkyl Substances in Existing Wastewater Treatment Plants- Scoping the Potential of Foam Partitioning.

Stults, J., Higgins, C. and Helbling, D. 2023. Integration of Per- and Polyfluoroalkyl Substance (PFAS) Fingerprints in Fish with Machine Learning for PFAS Source Tracking in Surface Water. *Environmental Science & Technology Letters*. **10**(11), pp.1052–1058.

SYKE 2018. Maaperän tilan tietojärjestelmä (MATTI). [Accessed 8 August 2025]. Available from: https://ckan.ymparisto.fi/dataset/maaperan-tilan-tietojarjestelma-matti.

Terveysministeriö 2023. *Sosiaali- ja terveysministeriön asetus talousveden laatuvaatimuksista ja valvontatutkimuksista annetun sosiaali- ja terveysministeriön asetuksen muuttamisesta*.

Turtiainen, H. 2023. PFAS-yhdisteiden esiintyminen taajama-alueiden pohjavesissä.

Turtiainen, H., Vahtera, H., Seppälä, A., Ulmanen, H. and Oksanen, A. 2025. *PFAS-yhdisteiden levinneisyys ja poistokokeilu Savion jätehuoltoalueella Ikuisuuskemikaalit pois kierrosta -hanke*.

Ulmanen, H. 2024. *Temporal and spatial variation of PFAS , metal and nutrient load in Vantaa River*. University of Helsinki.

Valvira, S. terveysalan lupa-ja valvontavirasto 2024. *Talousvesisäännösten soveltamisohje Osa II Enimmäisarvojen perusteet* [Online]. [Accessed 18 September 2025]. Available from: https://valvira.fi/documents/152634019/172742999/Talousvesisäännösten+SO+OSA+II+2024.pdf/ca6a4add-b39b-8546-9675-a9caa16c9b9d/Talousvesisäännösten.pdf?t=1709730064568.

Wang, Z., Dewitt, J.C., Higgins, C.P. and Cousins, I.T. 2017. A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)? *Environmental Science and Technology*. **51**(5), pp.2508–2518.

Weber, A.K., Barber, L.B., Leblanc, D.R., Sunderland, E.M. and Vecitis, C.D. 2017. Geochemical and Hydrologic Factors Controlling Subsurface Transport of Poly- and Perfluoroalkyl Substances, Cape Cod, Massachusetts. *Environmental Science and Technology*. **51**(8), pp.4269–4279.

WFD 2015. The Water Framework Directive (Standards and Classification) Directions. *Water Framework Directive*. **40**(2), 66 pp.

Woodward, E.E., Senior, L.A., Fleck, J.A., Barber, L.B., Hansen, A.M. and Duris, J.W. 2024. Using a Time-of-Travel Sampling Approach to Quantify Per- and Polyfluoroalkyl Substances (PFAS) Stream Loading and Source Inputs in a Mixed-Source, Urban Catchment. *ACS ES and T Water*.

Xiao, F., Deng, B., Dionysiou, D., Karanfil, T., O'Shea, K., Roccaro, P., Xiong, Z.J. and Zhao, D. 2023. Cross-National Challenges and Strategies for PFAS Regulatory Compliance in Water Infrastructure. *Nature water*. **1**(12), pp.1004–1015.

Zhang, X., Lohmann, R., Dassuncao, C., Hu, X.C., Weber, A.K., Vecitis, C.D. and Sunderland, E.M. 2016. Source Attribution of Poly- and Perfluoroalkyl Substances (PFASs) in Surface Waters from Rhode Island and the New York Metropolitan Area. *Environmental Science and Technology Letters*. **3**(9), pp.316–321.

Zhang, Y., Lai, S., Zhao, Z., Liu, F., Chen, H., Zou, S., Xie, Z. and Ebinghaus, R. 2013. Spatial distribution of perfluoroalkyl acids in the Pearl River of Southern China. *Chemosphere*. **93**(8), pp.1519–1525.

Zhong, H., Zheng, M., Liang, Y., Wang, Yingjun, Gao, W., Wang, Yawei and Jiang, G. 2021. Legacy and emerging per- and polyfluoroalkyl substances (PFAS) in sediments from the East China Sea and the Yellow Sea: Occurrence, source apportionment and environmental risk assessment. *Chemosphere*. **282**(March), Article 131042.