# An integrated framework for the detection and filtration of perfluoroalkyl substances from surface water in the Thames Basin

Wenqi (Jonathan) Zhao[1], Christopher Whitfeld[1]

[1]**Eton College, Windsor, Berkshire, United Kingdom**

**ABSTRACT**

PFAS pollution is a growing concern worldwide, with no equitable solution in the Thames Basin. We developed a geospatial neural network, predicting PFAS values to within 10% of experimentally validated values. With those predictions, we designed and tested a point-of-use filtration device to be installed on taps. Observing a 93% reduction in PFAS concentration, we reduce PFAS to below health limits of 4 ng/l. Further, we optimised the design, where a 10mm depth of activated carbon allows for 2.5 months of usage, with minimal impact on flow rate and introducing no impurities. Not only do we reduce the devastating impacts of PFAS pollution, but we present a potential solution that is accessible for all.

| Abbreviation | Definition | Abbreviation | Definition |
|---|---|---|---|
| PFAS | Perfluoroalkyl Substances | GAC | Granular Activated Carbon |
| PFOA | Perfluorooctanoic Acid | PSDM | Pore Surface Diffusion Model |
| PFOS | Perfluorooctanesulfonic Acid | MFPE | Map Forever Pollution Europe |
| PFHxA | Perfluorohexanesulfonic Acid | RMSE | Root Mean Squared Error |
| PPS | Presumptive Point Source | EBCT | Empty Bed Contact Time |

# Contents

# 1 Introduction

Per and Poly-fluoroalkyl Substances (PFAS) are synthetic organofluorine compounds, characterised by C-F bonds. Due to the low molecular polarity and strong C-F bond energy, PFAS are stable compounds. Thus, they have been used in a variety of industries around the globe, including in firefighting, aerospace and vehicles, and also distributed in our daily consumer products such as food packaging, fabrics and carpets[1]. Therefore, PFAS are prevalent in many water bodies globally, like the River Thames in the UK, especially near industrial sites. However, their solubility means that PFAS can travel into nearby drinking water sources, threatening 13 million lives.

It is impossible to understate the destructive nature of PFAS when unleashed into drinking water sources. Their impacts are demonstrably fatal in humans and animals, impairing immune function, reproductive outcomes, increasing the risk of different cancers, and prompting organ failure. Antibody concentration in young children decreased 49% for every 1 ng/mL of consumed PFAS, total concentrations of >5.7ng/mL indicated a 2.2 times increase in hypothyroid disease[2]. Contemporary research also indicates dramatic increases to the risk of childhood leukaemia (1.39 times)[3], breast cancer (1.47 times)[4], kidney cancer (1.74 times)[5], and thyroid cancer (1.56 times)[6], for every linear doubling of PFAS concentration in blood.

Perfluorooctane sulfonic acid (PFOS) and perfluorooctanoic acid (PFOA) are the most common forms of PFAS, accounting for >50% of PFAS from water samples in Hong Kong and Shenzhen [7] and >80% in Europe (See Materials and Methods). They are long-chained, containing a >8 carbon aliphatic chain. Long chained PFAS exhibit greater toxicity and accumulation, where they can remain in human bodies for 4-5 years[8]. Furthermore, compared to other long chained PFAS such as PFHxS, PFOS and PFOA pose additional harms. Both the International Agency for Research on Cancer (IARC) and the U.S. Environmental Protection Agency (EPA) have designated PFOA as a potential human carcinogen. For example, PFOA exposure indicates a 35% increase in developing testicular cancer, and every increase in 1ng/mL in the mother results in a 30 g reduction in baby mass[2]. 55ng/mL of PFOS resulted in a 350% increase in hepatocellular carcinoma[9], and alanine transaminase enzymes, both indicative of liver cancer, had 3.55 times increased production at high PFOS volumes[10].

Despite recent advancements in trying to phase out PFAS in industry, the occurrence of PFAS in the environment is expected to increase[11]. Though both the US and EU have implemented policies to regulate PFAS production, no such effort have been made in the UK. Thus, policy and technology to tackle PFAS pollution is of utmost importance [12], especially

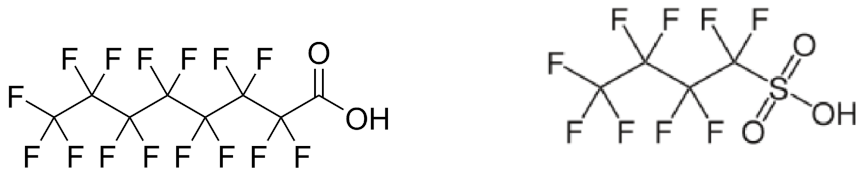


Figure 1: Molecular Structure of PFOA (left) and PFOS (right)

considering that 14.3% of the population have had adverse health effects because of different PFAS[13]. The cost to Europe as a whole thus sits at £46-75 billion[14].

However, the treatment of PFAS pollution in the is hindered in two major ways. First, current approaches to PFAS detection are expensive and highly specialised, with very few facilities in the UK that offer the service. The average cost of concentration measuring is £110. This means that identification of PFAS hot spots - a prerequisite to addressing the issue - is challenging. Therefore, both policies for PFAS regulation and the distribution of PFAS-remediating technology suffers in efficiency. Liberalising access to geospatial data is a struggle that must be overcome. Second, emerging technologies towards PFAS are unclear in their efficacy. Emerging filtration technologies, namely nanofiltration, ion exchange and reverse osmosis, are often limited in efficacy by fouling[15]. Implementation of Reverse Osmosis plants can cost $99 million; single use resins cost $44,164 [16]. Furthermore, they are most suitable for industrial applications, and are thus inaccessible to home settings. Though literature identifies Granular Activated Carbon (GAC) as a potential treatment mechanism, its prominence and efficacy as point-of-use filters are undermined by a lack of research. Specifically, GAC comes in a variety of sizes; there is no available research on the impact of those sizes on the suitability of GAC within such filters and when those filters need to be replaced.

Thus, we present here a novel approach to tackling PFAS contamination in the Thames Valley region: a concurrent detection and filtration mechanism at an affordable price. Detection is achieved with a geospatial neural network and filtration optimised through analysing the size and depth of GAC.

We consider an effective geospatial representation of regional PFAS concentration to be one that corresponds to any known values and allows any user to access data easily. We have thus designed a random forest-KNN ensemble that presents data continuously and interactively. The model has been verified by our own empirical analysis of PFAS concentration at locations along the River Thames, demonstrating concordant results, thus indicating our model's success.

For any point-of-use system design, we identified several important considerations, beyond the efficacy of the filter. First, flow rate must be carefully monitored, both to have minimal impact on customer use and to prevent backwash into the system, clogging the pipes. Second,

the filtration does not introduce additional contaminants. Third, the lifespan of the filter is crucial; it provides information on the churn rate and ensures sufficient filtration efficacy is maximised. Given these parameters, we designed a low-cost, effective, and long-lasting GAC Filter.

# 2 Materials and Methods

## 2.1 Materials

**Chemicals**

- Granular Activated Carbon - purchased from Intra Laboratories
- *Type 1* - Av. length: 4.60mm; 4 on Sieve N°(ASTM E11-87) scale
- *Type 2* - Av. length: 3.71mm; 6 on Sieve N°(ASTM E11-87) scale
- *Type 3* - Av. length: 0.67mm; 20 on Sieve N°(ASTM E11-87) scale; equivalent size to Calgon F400.

**Equipment**

- LC-MS Machine: Mass spectral resolution 70,000 FWHM
- Infrared Spectroscopy: Nicolet iS 10 FTIR Spectrometer
- pH probe: ETI 8000 pH meter
- Bunsen Burner + thermometer

**Software**

- Environmental Technologies Design Option Tool (ETDOT), AdDesign - PSDM
- Python + Libraries (MPL, NumPy, SciPy, GeoPy, Pandas, Folium)
- OMNIC Spectroscopy Standard tools

## 2.2 Machine Learning Detection

### 2.2.1 Model Summary

Building our geospatial model was a crucial first step in identifying high PFAS concentration areas. In summary, we used an ensemble of Random Forest (RF) and K-Nearest Neighbours (KNN) architecture for our model and trained on a data set compiled by the Map of Forever Pollution in Europe (MFPE) project. We then integrated the model's predictions into an interactive map that used inverse distance weighting (IDW) formulae to plot values continuously.

### 2.2.2 Dataset

The MFPE project was founded by Le Monde newspaper to track PFAS contamination across Europe and was released in 2023[17]. The data set includes 19500 instances of PFAS concentration samples, and 25,000 locations that are presumptive point sources (PPS) for PFAS. There were 5 distinct types of samples: surface water, ground water, sediment, biota, and soil. We primarily utilised the 6697 instances of soil sampling locations, as these are the most indicative of concentration in surrounding bodies of water [18][19]. The PFAS PPS's include 12,200 fire-fighting incidents, 1000 airports, 3200 industrial plants, 740 military locations, and 4800 waste management sites. For all sampling locations, the readings were measured with resolution ng/L or the equivalent ng/Kg. For soil samples, there was a range of 17 - 5,351,000 ng/Kg and a mean of 32,859 ng/kg. The samples measured 6 different types of PFAS, although only PFOS and PFOA were consistently sampled. The values most significant in training our model were the PFAS summation values for each sampling location, and their longitude and latitude.

### 2.2.3 Training Process

The relationship we sought to model was one that mapped the concentration of a sample to its distance from the closest PPS. This was an intuitive decision, as PFAS are mainly produced in an industrial capacity. Thus, their prevalence in soil is contingent on their relative location to producers. To analyse this relationship, the geo-coordinates of the samples were first processed with respect to their haversine distance from the closest industrial site. Where $\phi$ and $\lambda$ represent latitude and longitude, the haversine equation is given by:

$$d = 2r \sin^{-1}(\sqrt{\sin^2 \frac{\phi_1 - \phi_2}{2} + \cos(\phi_1)\cos(\phi_2)\sin^2(\frac{\lambda_2 - \lambda_1}{2})})$$

We took this approach because the samples and PPS's were often sufficiently far from each other to justify accounting for the earth's curvature. Given these distances, which were computed in Python, we were able to plot graphs relating sample concentration and distance to nearest PPS. To model the relationship and allow predictions, we ran the data through a variety of classical machine learning architectures. We tested a geographic-weighted regression model, KNN, and RF. We ultimately settled on the RF-KNN package, since it is highly robust when dealing with statistical noise and outliers like much of the MFPE data [20], We partitioned our data in a standard 80:20, training-testing split. To optimise performance, we ran the model through a ten-fold cross validation and the results of these validations were averaged. Significant outliers, defined as values greater than 400,000 ng/Kg, were excluded since they

distorted predictions dramatically. This had only a minor impact on the model's efficacy, as this represented less than 4% of data.

### 2.2.4 Geospatial prediction for UK

To generate our map, we chose 15,000 locations within the UK, and predicted their value. We then converted the soil concentration values to water concentration figures using the following equation:[19]

$$C_w = \frac{C_S}{\frac{\phi_w}{\rho} + K_d}$$

Where $\phi_w$ is the water content of soils in the Thames Valley, $K_d$ is the soil-water distribution coefficient for PFOS/PFOA, and $\rho$ is the soil bulk density. Values for these constants were attained from our communications with Thames Water. Though there is relative variance in soil type in London, these constants were sufficiently similar such that modelling with these constants could be applied homogeneously. We give the values for these constants in *Table 1.* Values were either taken from direct communication with Thames Water or through research.

| Bulk Density (kg/L) | Water Content of Soil | Soil-Water Coefficient (L/kg) |
|---|---|---|
| 0.95-1.30 [21] | 0.62 [22] | 1895 [18] |

Table 1: Table of soil concentration equations constants

To make our model more continuous, we applied the inverse distance weighting technique to smooth gradients on our map.

## 2.3 GAC Filtration

We constructed the CAD model of the filter. It is a simple point-of-use add-on to existing taps and pumps in homes.

- We used PBS (Polybutylene succinate) for the filament, for its biodegradable properties and low media migration, which stops micro plastic pollution.
- The design was double open end - the influent end open to maximise flow rate; the effluent end contains a perforated surface of diameter 1.4mm (corresponding to mesh size 8x14)[23], ensuring the GAC media cannot pass into the effluent.

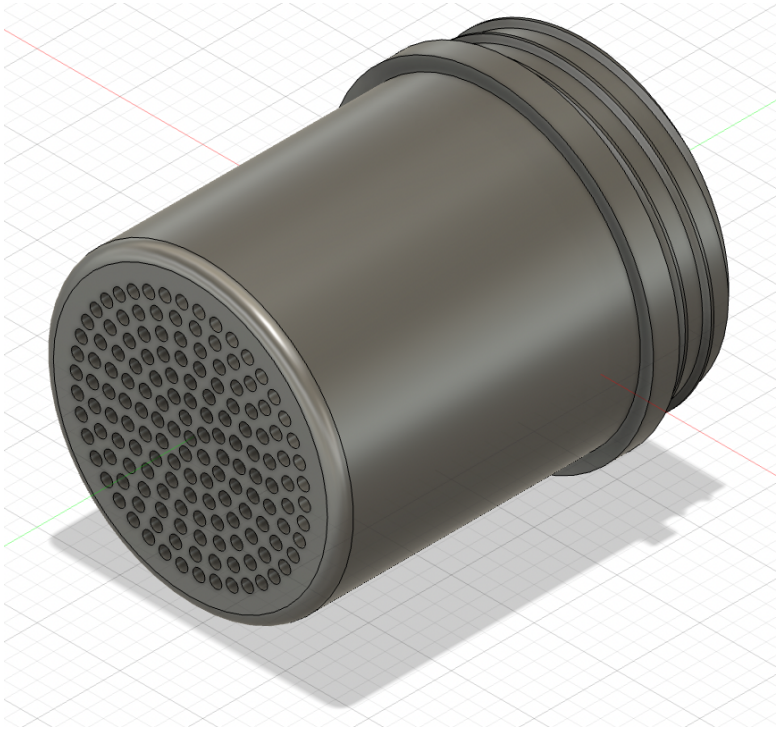

Figure 2: Autodesk Fusion 360 CAD Design for filter

### 2.3.1 Flow Rate Testing

A point-of-use experiment was set up in the intended environment - taps and pumps in homes. We filled the filter to a stratified depth of GAC (increasing, regular intervals of 10 or 20 mm). For each, we calculated the maximal permissible flow rate, defined as the flow rate at which the height of water within the filter remained constant. Any greater, and the filter will gradually fill up over time and backlog the pipes. Assuming a constant tap flow rate, the rate was then measured by the time taken to fill a given volume of water - set at 1 litre. The time was measured by a stopwatch. To account for inaccuracies in human response time, we took the average across 10 readings of flow rate and discarded outliers, defined as values more than 1.5 * IQR away from the 1st and 3rd quartile.

### 2.3.2 Filtration Efficacy

At the optimum depth we calculate later, we ran tap water through our device at that flow rate. We measured the concentrations before and after the filtration with an LC-MS machine (since PFAS/PFOA exist at extremely low concentrations, we employed an LC-MS machine). For preprocessing, Trizma (5.0 g/L) was added as a buffer reagent and removed chlorine in the water. PFOS and PFOA were extracted from an aqueous matrix with solid phase extraction, utilising Thermo Scientific's EQuan Liquid Chromatography (LC) system[24]. The compounds

are then backflushed from the pre-concentration column within the LC system via a gradient run through Hypersil Gold aQ pre-concentration column (140 x 46mm). At a flow rate of 1 mL/min, the mobile phase solvent was 0.1% ammonia at pH 10.6; the solid phase solvent was methanol + 0.1% ammonia.

With this method, 250 mL of the sample is collected and re-suspended in 1 mL after extraction, creating a 250:1 preconcentration. For the LC-MS, we determined the concentration and volume. For concentration, gentle stream of nitrogen in a heated water bath (60-65 °C) removed all the water/methanol mix. Then, the IS PDS (Internal Standard Primary Dilution Standard) - in methanol with 4 moles of NaOH - was added to reach a final volume of 1mL. A small aliquot was transferred with a plastic pipette to a polypropylene autosampler vial. For volume, the weight of sample was measured (subtracting container weight) and volume calculated given a density of 1g/ml[25].

Calibration was conducted with stock solution of PFOS/PFOA (Sigma Aldrich). Subsequently, mass spectrometry was performed. Electrospray ionization (ESI) was used to ionise the sample.

### 2.3.3 Lifespan Estimation

Given the physical and cost constraints of a long-term experiment, we modelled the experiment with AdDesignS [26], a software developed by the US Environmental Protection Agency. We inputted the adsorbent length based on measurements of *Type 1* GAC length (average of 10 pieces, outliers removed). Most of the parameters were universal. For the device, the bed diameter and bed length were measured with a vernier gauge. The Freundlich isotherms, $K$ and $\frac{1}{n}$ were obtained from literature [27].

There were two main parameters - independent variable - that we changed to observe lifespan: the bed mass and the flow rate. Both were calculated from the Flow Rate Experimentation graphs, assuming a linear relationship (See Discussion). A given flow rate (in time taken to fill 1 litre) corresponds to a particular bed depth.

The density $\rho$, of the GAC was calculated by the mass, measured with a mass balance, divided by the volume of a known container. Thus, the bed mass, which we input, is $\rho$ times the the bed volume. To estimate lifespan, we ran a Pore Surface Diffusion Model (PSDM) experiment, outputting the C/C0 ratio, which represents the filtration efficacy, where $C$ =concentration and $C_0$ = initial concentration, against the run time.

### 2.3.4 Water Quality

There were three major experiments performed to ensure GAC filtration does not introduce additional contaminants to the water. We ran tap-water through the filtration device, measuring the pH with an ETI 8000 Series pH metre of the water before and after. Similarly, we tested the boiling rate through a simple Bunsen Burner set-up with a thermometer for the temperature; the boiling point was defined as the point where temperature stays constant.

Separately, we used a standard FTIR spectrometer to provide a thorough analysis of the chemical composition of water before and after filtration. The spectrometer shines IR light at the sample and records the intensity at different wavelengths to determine how much IR is absorbed by the sample. O-H bonds for example, are detected when the wavelength of IR corresponds to the O-H bonds specific vibrational mode. The results were interpreted using OMNIC software.

Samples were placed on the deuterated triglycine sulfate (DTGS) detector and left for 25 seconds. Results were then saved and the DTGS detector cleaned with ethanol to remove traces of previous samples. We then interpreted the results by layering each spectra onto one another.

# 3 Results

## 3.1 Machine Learning

Our model demonstrates remarkable accuracy. We judge this through standard statistical analyses, and our own empirical experimentation that confirms the model's validity.

### 3.1.1 Visualisations

We produced visualisations (*Figure 3*) following exploratory data analysis to understand its properties better.

The map demonstrates the breadth of our data. For our training we relied on soil values, which cover a more limited but still relevant area.

### 3.1.2 Statistical analyses

We considered three metrics to judge our efficacy, root mean squared error (RMSE), mean absolute error (MAE) and our $R^2$ value. Our results are summarised below in *Table 2*. For our MAE and RMSE, values were stated in the summation of concentration of all PFAS in soil. To convert values to water, we apply the equation in 2.2.4. For this we use the constants

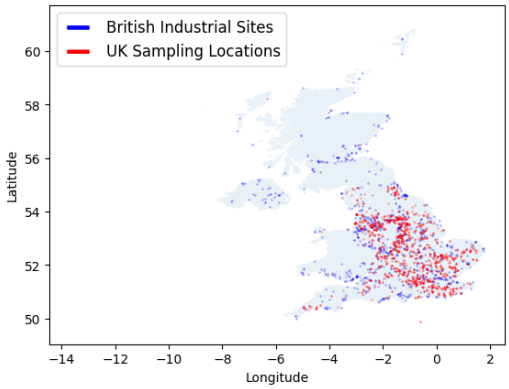


Figure 3: Relationship between industrial sites and PFAS sampling locations in the UK

determined in *Table 1.* When we consider the mean (32,859 ng/kg = 17.34 ng/L) value for soil concentrations, we can thus derive an average percentage error in the model's predictions

| Metric | Conc. soil (ng/kg) | Conc. water (ng/l) | Error (%) |
|---|---|---|---|
| RMSE | 3653 | 1.92 | 11.07% |
| MAE | 2119 | 1.12 | 6.46% |

Table 2: Model error in terms of soil and water concentration

### 3.1.3 Interactive map and comparative analysis with experimental findings

We present a map of the UK that has the predicted values at 15,000 locations. This map is continuous and user can find concentrations at any longitude and latitude. *Table 3* shows the values measured at 4 diverse locations by ourselves, and compares these samples to predictions from the model for the types PFOS and PFOA.

| Locations | Own Sampling PFOS/PFOA (ng/L) | | Model Prediction PFOS/PFOA (ng/L) | | Error PFOS/PFOA (ng/L) | |
|---|---|---|---|---|---|---|
| 51.4589, -0.3092 (Richmond) | 6.44 | 2.94 | 5.61 | 2.65 | -12.9% | -9.86% |
| 51.4763, -0.1760 (Battersea) | 6.09 | 3.14 | 5.55 | 3.29 | -8.87% | +4.78% |
| 51.5015, -0.0170 (Isle of Dogs) | 9.04 | 3.79 | 8.82 | 3.72 | -2.43% | -1.85% |
| 51.4851, -0.6119 (Windsor) | 6.16 | 1.74 | 5.49 | 1.91 | -10.88% | +9.77% |

Table 3: Experimental values vs Theoretical values

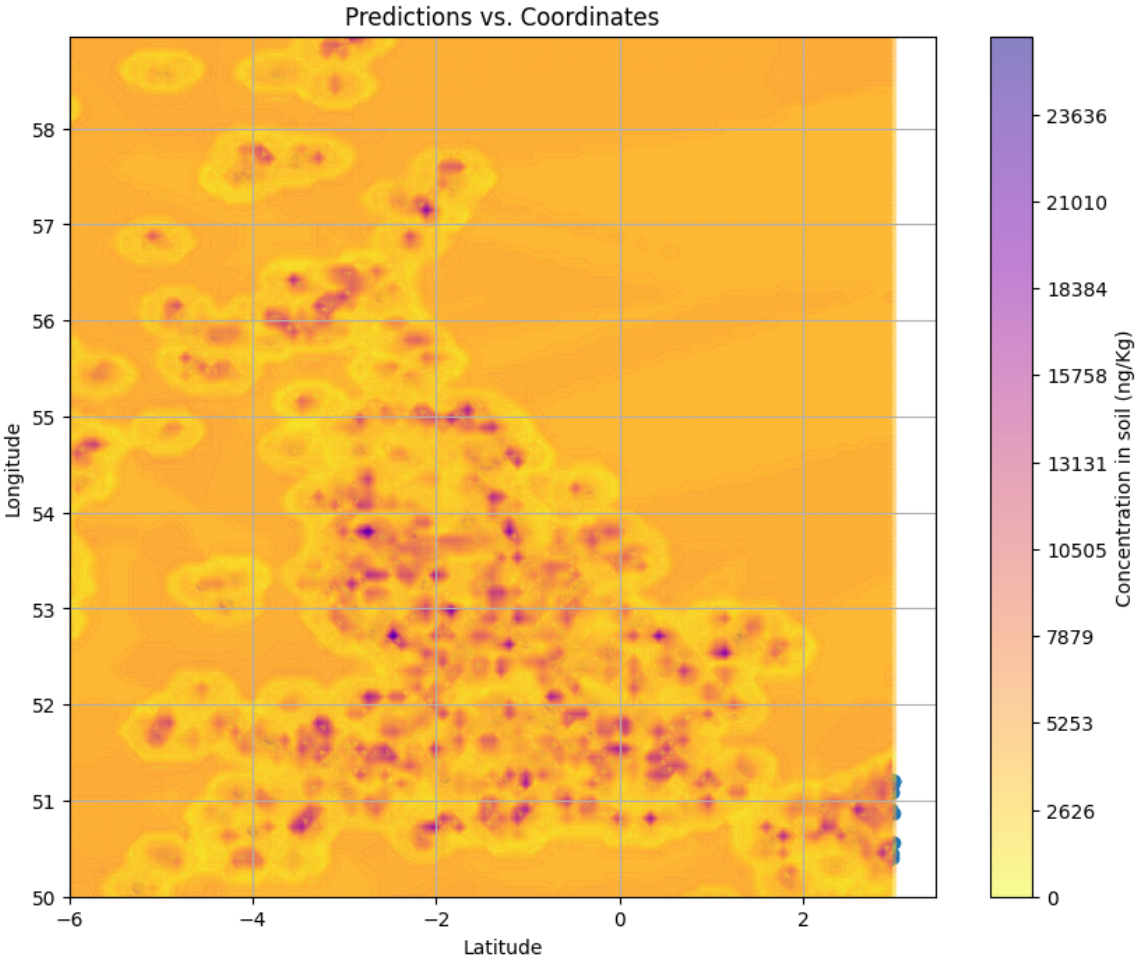


Figure 4: Interactive map of UK for Soil Concentrations

## 3.2 Filtration

### 3.2.1 Flow Rate

We plotted the relationships between flow rate (time taken to fill 1 litre) against depth for all three granularities of GAC and fitted the relationship. The $R^2$ was $> 90\%$ for all three relationships, and generally positive relationship was observed.

### 3.2.2 Filtration Efficacy

Having interpreted the outputs from the LC-MS machine (*Figure 6*), the observed % reduction in PFAS after filtration by GAC is displayed in *Table 4*.

| | Control | Type 1 | Type 2 | Type 3 |
|---|---|---|---|---|
| PFOS | 2.34 | 0.27 | 0.53 | 0.33 |
| PFOA | 5.66 | 0.29 | 0.84 | 0.56 |
| Total | 8.00 | 0.56 | 1.37 | 0.89 |
| % Reduction | 0 | 93.00 | 82.725 | 88.725 |

Table 4: Filtration efficacy of GAC for PFOS and PFOA

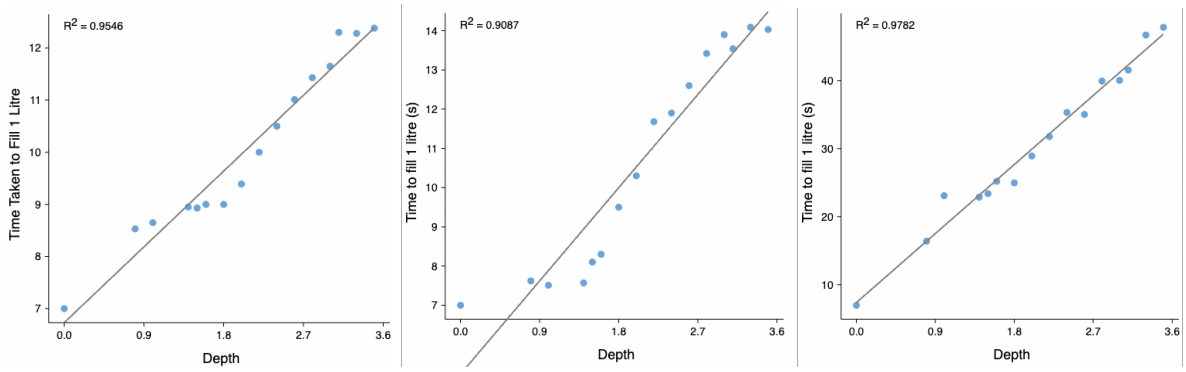


Figure 5: Results of Depth of Filter filled with GAC vs Flow Rate. Linear Regression for optimal best fit line.

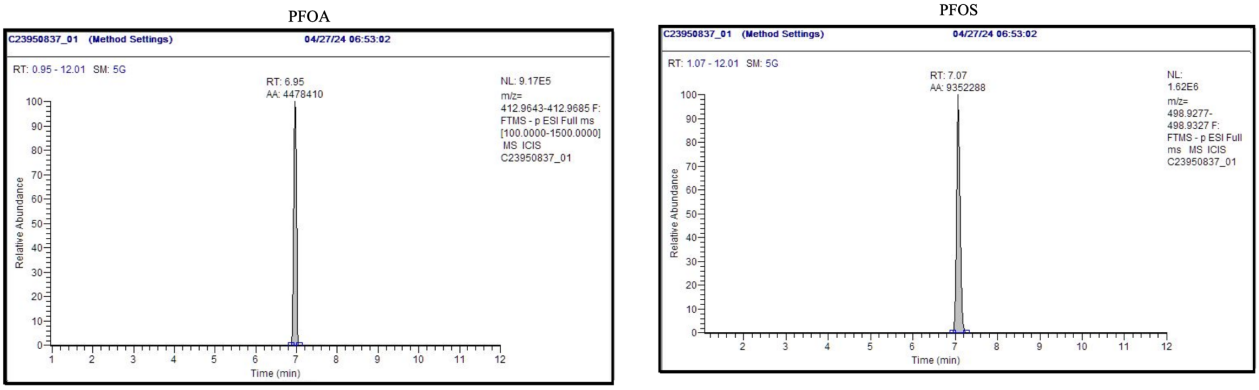


Figure 6: Outputs of LC-MS calculation for PFOA and PFOS concentration for Control

### 3.2.3 Lifespan

We find a clear best granularity for both filtration efficacy and flow rate - 4.6mm average size. Given the size of this media, we modelled the lifespan of filter based on the depth we fill the filter with. The minimum C/C0 required is 50%: PFAS (PFOS + PFOA) values pre-filter sum to 8 ng/l, which is double the recommended limit of 4 ng/l. Therefore, we find the $x$ value when the $y$ value is 0.5; the following graph models a lifespan of 320 minutes, or 0.22 days, of continuous water flow. The final lifespan predictions (*Figure 8*) given flow rate and mass of media shows a linear relationship to both flow rate and depth.

### 3.2.4 Device Optimisation

Given equal importance for both flow rate and lifespan, the optimal is found to be the intersection of the graphs. Here, the depth is 10.0mm, allowing for a flow rate of 7.12 litres per minute, while being effective for an estimated 368 minutes. A home tap runs continuous for around 5 minutes a day (See Discussion), and therefore this corresponds to 2.5 months of usage before a replacement is needed.

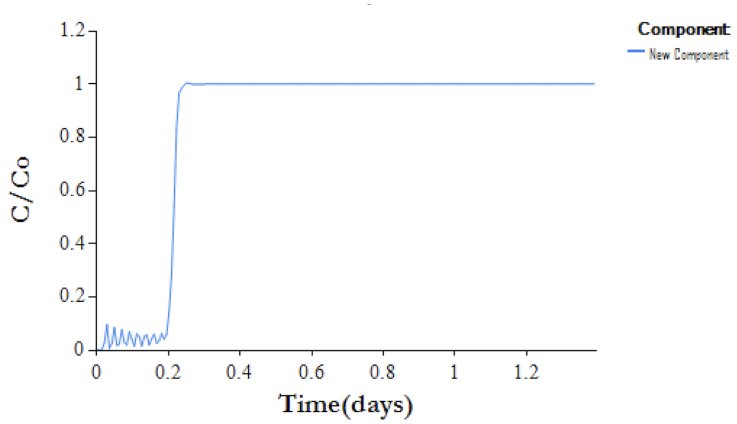


Figure 7: Lifespan prediction for a depth of 0.45 of GAC

| Time taken to fill 1 litre (s) | Depth (cm) | Lifespan (minutes) |
|---|---|---|
| 10 | 2.01 | 450 |
| 9.75 | 1.85 | 435 |
| 9.5 | 1.75 | 420 |
| 9.25 | 1.60 | 408 |
| 9 | 1.44 | 397 |
| 8.75 | 1.25 | 384 |
| 8.5 | 1.10 | 375 |
| 8.25 | 0.91 | 365 |
| 8 | 0.80 | 353 |
| 7.75 | 0.65 | 337 |
| 7.5 | 0.45 | 320 |
| 7 | 0.2 | 300 |

Table 5: Lifespan of GAC based on depth and flow rate

### 3.2.5 Adverse Effects

| Before | Type 1 | Type 2 | Type 3 |
|---|---|---|---|
| 7.41 | 7.01 | 6.84 | 7.29 |
| 7.40 | 7.00 | 6.87 | 7.28 |
| 7.40 | 6.98 | 6.84 | 7.29 |

Table 6: pH of tap water before and after filtration with GAC

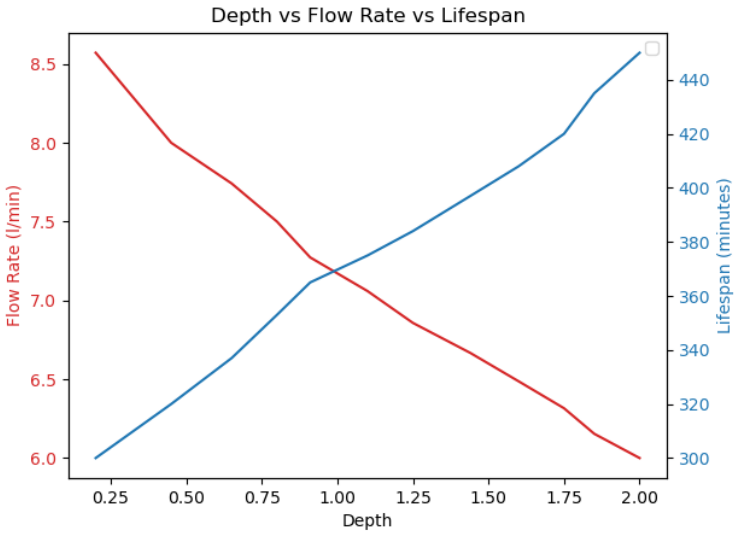


Figure 8: Impact of depth of media on lifespan and flow rate

| Before | Type 1 | Type 2 | Type 3 |
|---|---|---|---|
| 100.7 | 100.6 | 100.6 | 100.6 |
| 100.7 | 100.6 | 100.6 | 100.7 |
| 100.7 | 100.6 | 100.7 | 100.7 |

Table 7: Boiling point of tap water before and after filtration with GAC

# 4 Discussion

## 4.1 Machine Learning

### 4.1.1 Key findings and literature context

The empirical and theoretical results of our ML model demonstrate a substantial degree of accuracy and applicability. For the two most common and harmful PFAS compounds, PFOS and PFOA, our model consistently achieves values that vary from the true value by an absolute value of less than 10%. Even the RMSE metric, that penalises data sets like ours that have significant range, we achieve a remarkably low 11.07% average error. Our own sampling demonstrates irrefutably that our model can be used in the real world, which has allowed the development of our interactive map for greater accessibility to this data

Contextually, our model provides a novel approach to predicting PFAS concentration. Some have considered ML approaches using biological sampling, such as fish tissue [28]. Though this approach yields positive MAE values (8.32%, at 7.26ng/g), which we note is still greater than our results, RMSE values are comparatively inaccurate (164.65 ng/g, where the mean is 6.70 ng/g)[28]. Similarly, results from this research are derived from a data set containing only 45 samples of types of PFAS, where we have trained on >6600. We believe our model is thus more

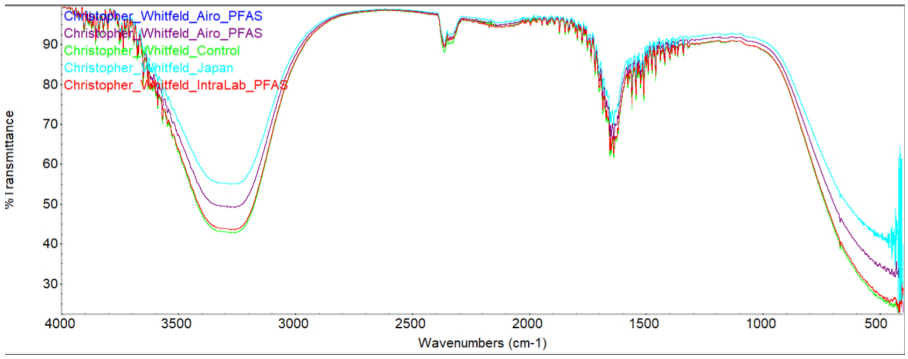


Figure 9: Infrared Spectroscopy transmission of water before and after GAC filtration

robust, and therefore more helpful for application. Equally, the geographic parameters that convert predictions to water concentrations are uniquely centred on the Thames and England. Our results more generally are derived from Western European geographic data, and no other [20] research intersecting ML and PFAS detection share our bounds. Since there is a well-defined relationship between soil matrices and PFAS diffusion into water sources [19][21][29], our results are thus a remarkably strong approximation for PFAS concentration, owing to their geographic relevance.

Some results have prompted surprise, however. Though there was concordance between our own sampling and the model's predictions, why they varied is not fully clear. However, we notice a subtle negative correlation between the sum total PFAS and error; as total PFAS increases, the error seems to decrease as well. The error from the true value for our largest sample (Isle of Dogs) is only -2.43% for PFOS and -1.85% for PFOA. We also find the predictions for our map intriguing. PFAS are industrial compounds, and their prevalence in the less industrialised[30] North and South West of London are counter-intuitive. This is perhaps driven from a misinformed hypothesis on our part about the nature of London's industry; clearly PFAS production is not constrained to traditionally industrial areas, further highlighting the necessity for our model. However, our hypothesis that there is a geospatial relationship between PFAS PPS's and PFOS/PFOA concentration at a given point, has been generally met more strongly than expected.

### 4.1.2 Limitations

Though we are generally satisfied with our methodology and subsequent results, we see some potential problems in our execution. We have not considered the impact of time and PFAS half-life on our model's performance. There is insufficient temporal data within the data set, and thus we cannot make any inferences about the model's capabilities in this regard. However,

the data was collected in the 2019-2022 period, and our experimentation was in early 2024. Given also the minimum 1000 year half life of PFOS and PFOA [1], there is perhaps cause for optimism that the model can perform well into the future. There was also some unevenness in the data set. We primarily considered soil concentration samples, which though the matrix most reflective of PFAS concentration, only existed in large quantities in some of the UK, Belgium and Denmark. Thus, there is less range to the data than hoped. Our experimental results help refute this to an extent, but economical reasons meant we only had access to 4 samples, which perhaps makes it more challenging to rely on this information.

#### 4.1.3 Recommendations

Thus, we see scope for potential future research to refine our initial model proposal. We would like to see a more complete picture of soil samples across Western Europe by pioneering organisations like MFPE. Overlaying this data with the current data set could produce temporal inferences, if only for the sake of confirming the "foreverness" of PFAS. We would like to conduct more analyses of samples from across the Thames ourselves, to verify our model further. Finally, we would like to refine the map by integrating into it reservoirs and river abstraction points, to make immediate inferences about the relative impact of PFAS contamination on health. Not including this information does not detract from the model's efficacy though.

Ultimately, our results show much promise in detecting PFAS computationally, indicating only small divergences from experimental truths, and maintaining consistency in spite of the varied nature of our data. Our visualisation is indicative of an accessible interface that helps bridge the current knowledge and data gap that pervades even the state of the art of PFAS detection.

With our model, we bring equity into solving the PFAS problem. The results are accessible to anyone, liberalising access to data about PFAS. We remove constraints on cost, which limit sample points, and demographics, which leave certain regions ignorant about pressing environment issues. Indeed, through improving knowledge about the problem distribution and detail the risk to more people, we raise more awareness for the cause.

### 4.2 Filtration

#### 4.2.1 Suitability

The results for filtration efficacy are most important. It is in line with literature suggesting GAC's ability to filter PFAS. The optimal granularity is observed at *Type 1*, with average length

4.6mm. Although the control exhibited volumes for PFOS/PFOA of 8 ng/l, well above the recommended safety limit of 4 ng/l, all three sizes of GAC were able to filter PFOS and PFOA to far below that limit. The results are above satisfactory compared to industry standards. Belkouteb et al., using GAC as traditional media, found a 82% filtration rate for PFAS [31]. An 84% filtration efficacy was observed by single-stage under-the-sink filters[32]. Furthermore, whole house GAC systems are only 67% effective for PFOS and 19% for PFOA. Other treatment media, such as anion exchange membranes, only observe a $60-80\%$ filtration efficacy [33]. This suggests our filtration device can be a crucial addition to current home filtration technologies.

Our efficacy is not as high as some emerging physical technologies, such as reverse osmosis, achieving 100% and $> 92\%$ filtration of PFOS and PFOA, respectively[31]. However, these solutions are extremely expensive to run and operate. The reverse osmosis plant in Brunswick County, North Carolina cost $99 million, with additional $2.9 million annual costs [14]. For the same population, our costs are at $1.4 million (See 4.2.2).

Our flow rate results also align with our hypothesis: that the size of the GAC has a statistically significant (p-value = 0.0157) impact on its permeability and maximal permissible flow rate. The flow rates allowed by our device suggest its suitability in a commercial setting. The optimal granularity for flow rate was again observed for *Type 1*. Though the maximum flow rate of conventional taps are up to 12 l/min, actual usage is around 5 l/min [34]. This is below the flow rate (7.12 l/min) of our proposed optimal depth, meaning we don't impede on the daily life of users. Similarly, it is possible to achieve a flow rate of 8.5 l/min, albeit with a reduced lifespan to 300 minutes (2 months). A similar relationship is observed by Belkouteb et al., where an increase in flow rate decreases efficacy (and hence lifespan), as it lowers the EBCT, and hence the time PFOS/PFOA molecules are in contact with GAC. This research differentiates us from existent filtration devices such as Brita - which can only filter small quantities of water at a given time on account of its slow flow rate.

The lifespan analysis were promising. The optimal balance between flow rate and lifespan produced allowed for 368 minutes of continuous usage. This corresponds to a 2.5 month usage time. The lifespan of large scale point-of-entry systems, which use 100x the mass of GAC, is only 7 months [16].

Finally, we verified that GAC filtration results in little negative impacts on other metrics for water potability. The pH, slightly alkaline with an average of 7.40, was even broadly neutralised (to varying degrees) by each type of GAC. The size demonstrating the optimal filtration efficacy and allowing maximal flow rate, *Type 1*, indeed resulted in an almost completely neutral pH - safe for human consumption. Second, the boiling point of water, indicative of impurities [35],

remained constant. A two-tailed $p$-value of 0.11 suggests no statistical significance. Moreover, the results from the IR spectroscopy suggest that no impurities were introduced. The three troughs observed: at 3300-3700 cm$^{-}$1, 1600-1650 and 650-700 belong to the O-H stretching, O-H bending and H-O-H bending effect. Therefore, no additional organic compounds were introduced by GAC filtration. This is supported by other pilot studies that list GAC as a means of commercial water filtration [36].

### 4.2.2 Cost, Sustainability and Wider Social Application

Our filtration device has been designed to minimise costs, such that it is an accessible tool for all socio-economic brackets. The cost to 3D print is only £5.72, and 25g of *Type 1* GAC (sufficient for one device across a year) is only 47p, totalling £6.19. Given the £14,700 average income for the poorest decile in London, we hope our device remains an affordable option. Since PFAS contamination occurs in a range of geographies, this price versatility is one of the most important features of our project. To encourage implementation of our device, we are setting up stalls at community events to spread awareness about the impacts of PFAS. Obviously, we have ensured the device and GAC disposal remain environmentally friendly as well. PBS filament stops micro-plastic media migration, and there is substantial evidence [37] suggesting that disposing of GAC after use does not release PFAS into soil. This is because PFAS binds to the carbon, rendering it harmless. To coordinate the release of our filter into the wider consumer market, we have already contacted environmental organisations like the Soil Association, Thames Water, and the US Environmental Protection Agency for advice. Consumer interest has already been shown by the Wandsworth Council and investors at the 1517 Fund.

### 4.2.3 Limitations and Further Developments

The success of the final optimised product is limited by several factors.

First, on the flow rate analysis. We were initially interested in finding the degree of the function - whether the relationship between the depth and the flow rate is linear, quadratic, cubic etc. However, the data points gathered were insufficient to give a significant insight. The R-squared values for the different bit lines has a t-test value of 0.34, suggesting it is impossible to determine which fit is best. Therefore, our assumption of a linear fit meant that further predictions of depth for a corresponding growth rate could have been marginally different, potentially affecting the lifespan modelling.

Second, on the lifespan modelling. First, having observed *Type 1* to allow the maximum flow rate and maximum filtration efficacy, we used this granularity for every further calculation. We were surprised to find a linear relationship between the depth, flow rate and lifespan, given there were two independent variables simultaneously changed to result in the output for lifespan. This may be the way in which the AdDesign PSDM software is set up, where the EBCT (influenced by depth and flow rate) has high weighting on the final lifespan prediction. This paves way for further research, where the accuracy of the long term prediction and breakthrough time can be verified experimentally.

Third, it is important to note that the optimised results we reached - utilising *Type 1* GAC at a depth of 10mm - assumes an equal importance between flow rate impact and the replacement rate of the filter. Customers may well favour one aspect over another. That's where our geospatial neural network can come in. For example, in the Isle of Dogs, where the concentration in water is highest at 12.83 ng/L, a depth of 2cm for the GAC may be optimal in ensuring maximal efficacy. Customers may then be encouraged to use taps more sparingly and refrain from using the highest pressure.

Finally, we specifically target PFOS and PFOA, the most common forms of PFAS. Therefore, the applicability of our methodology to other forms of PFAS, such as PFNS, PFHxS are unclear. However, there is rationale behind our choice. GAC is specifically effective for long-chained, higher-prevalence forms of PFAS. Additional experiments may be carried out for a broader spectrum of PFAS to verify whether the device is as effective for those. However, the concentration is generally at far lower levels and pose less urgent threat to human health.

The technology and pipeline we developed can be scaled globally: not only is the cost-of-implementation low, but the research is robust. However, further research and development is required into varying diameters of filters. We designed the filter to fit on sinks prevalent among the Thames region, but industrial applications may involve calculations with a different diameter (110mm), whose influence on PSDM modelling is still unclear.

# 5 Conclusion

PFAS continue to pose an existential threat to drinking water in communities across Western Europe. Their destructive footprint is a painful reminder that industrial regulation is still insufficient to prevent their diffusion. Our research thus hopes to provide a grassroots solution for their removal from water sources using our single pass filtration process. We bridge the information gap that has left the wider public largely unaware of the consequences of consuming

such insidious chemicals. We are able to predict areas of high PFAS concentration with an average error of only 6.46%, allowing limited resources to be directed to most polluted regions and policies to limit PFAS production. For every household, we provide a sustainable, cost-effective solution to a problem whose consequences are potentially fatal. Not only do we filter $> 90\%$ of the most prevalent and harmful forms of PFAS, we ensure it has no impact on water usage in homes and maintains efficient for months of usage.

Beyond the Thames Basin, the scalability is global. Whether it is the Biden Administration banning PFOS/PFOA, or the European Union proposing regulations on PFAS, "forever chemicals" are becoming a worldwide problem. Governments and consumers worldwide require a concurrent pipeline to understand their prevalence - even in parts per trillion - and act to remove them from our planet forever.

# References


[1] CHEM Trust UK. *The 'Forever Chemicals*. Sept. 2019. URL: https://chemtrust.org/wp-content/uploads/PFAS_Brief_CHEMTrust_2019.pdf (visited on 05/12/2024).

[2] Suzanne E. Fenton et al. "Per- and Polyfluoroalkyl Substance Toxicity and Human Health Review: Current State of Knowledge and Strategies for Informing Future Research". In: *Environmental Toxicology and Chemistry* 40.1 (2021). First published: 05 October 2020, pp. 113–133. DOI: 10.1002/etc.4890. URL: https://doi.org/10.1002/etc.4890.

[3] RR Jones et al. "Maternal serum concentrations of per-and polyfluoroalkyl substances and childhood acute lymphoblastic leukemia". In: *Journal of the National Cancer Institute* 116.5 (May 2024), pp. 728–736. DOI: 10.1093/jnci/djad261.

[4] X Li et al. "Perfluoroalkyl substances (PFASs) as risk factors for breast cancer: a case-control study in Chinese population". In: *Environmental Health* 21.1 (Sept. 2022), p. 83. DOI: 10.1186/s12940-022-00895-3.

[5] MS Seyyedsalehi and P Boffetta. "Per-and Poly-fluoroalkyl Substances (PFAS) Exposure and Risk of Kidney, Liver, and Testicular Cancers: A Systematic Review and Meta-Analysis". In: *Medicina del Lavoro* 114.5 (Oct. 2023), e2023040. DOI: 10.23749/mdl.v114i5.15065.

[6] Li Jing and Zhixiong Shi. "Per- and polyfluoroalkyl substances (PFAS) exposure might be a risk factor for thyroid cancer". In: *EBioMedicine* 72 (Oct. 2023). Open Access, p. 104866. DOI: 10.1016/j.ebiom.2023.104866. URL: https://doi.org/10.1016/j.ebiom.2023.104866.

[7] Ziyad Abunada, Motasem YD Alazaiza, and Mohammed JK Bashir. "An overview of per-and polyfluoroalkyl substances (PFAS) in the environment: Source, fate, risk and regulations". In: *Water* 12.12 (2020), p. 3590.

[8] Fan Li et al. "Short-chain per-and polyfluoroalkyl substances in aquatic systems: Occurrence, impacts and treatment". In: *Chemical Engineering Journal* 380 (2020), p. 122506.

[9] JA Goodrich et al. "Exposure to perfluoroalkyl substances and risk of hepatocellular carcinoma in a multiethnic cohort". In: *JHEP Reports* 4.10 (Aug. 2022), p. 100550. DOI: 10.1016/j.jhepr.2022.100550.

[10] Elizabeth Costello et al. "Exposure to per- and Polyfluoroalkyl Substances and Markers of Liver Injury: A Systematic Review and Meta-Analysis". In: *Environmental Health Perspectives* 130.4 (Apr. 2022), p. 046001. DOI: 10.1289/EHP1009. URL: https://doi.org/10.1289/EHP1009.

[11] URL: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/pfoa-stewardship-program-baseline-year-summary-report.

[12] Emiliano Panieri et al. "PFAS molecules: a major concern for the human health and the environment". In: *Toxics* 10.2 (2022), p. 44.

[13] *Risks of PFAS for human health in Europe (Signal)*. Published 16 Apr 2024, Modified 17 Apr 2024. European Environment Agency. Apr. 2024. URL: https://www.eea.europa.eu/themes/human/chemicals/pfas-risk-human-health-europe.

[14] Alissa Cordner et al. "The True Cost of PFAS and the Benefits of Acting Now". In: *Environmental Science & Technology* 55.14 (July 2021). Erratum in: Environ Sci Technol. 2021 Sep 21;55(18):12739, pp. 9630–9633. DOI: 10.1021/acs.est.1c03565.

[15] Nour AlSawaftah et al. "A comprehensive review on membrane fouling: Mathematical modelling, prediction, diagnosis, and mitigation". In: *Water* 13.9 (2021), p. 1327.

[16] Anderson C Ellis et al. "Life cycle assessment and life cycle cost analysis of anion exchange and granular activated carbon systems for remediation of groundwater contaminated by per-and polyfluoroalkyl substances (PFASs)". In: *Water Research* 243 (2023), p. 120324.

[17] G. Dagorn et al. "Map of Forever Pollution in Europe (MFPE), Journalists tracking forever pollution across Europe". In: *Le Monde, The Guardian, Süddeutsche Zeitung, Reporters United et al.* (2023).

[18] Nika Nordanstorm. "Evaluation of distribution coefficients (KOC and Kd) for per- and polyfluoroalkyl substances". Examination project work, Environmental risk analysis, 30 ECTS. Master's Thesis. Sweden: Linnaeus University, 2020–2021.

[19] Bundesamt für Umwelt BAFU. *Derivation of concentration values and solid limits*. Environmental Enforcement UV-1333-D. Federal Office for the Environment FOEN, 2013, p. 21.

[20] Adewale Ajao. "Application Of Machine Learning To Understand Pfas Occurrence, Distribution, Transport And Removal In Water". In: (2024).

[21] J. Evans et al. "Soil water content in southern England derived from a cosmic-ray soil moisture observing system - COSMOS-UK: Soil Water Content in Southern England - COSMOS-UK". In: *Hydrological Processes* 30 (May 2016). DOI: 10.1002/hyp.10929.

[22] *COSMOS UK*. UK Centre for Ecology & Hydrology. n.d. URL: https://cosmos.ceh.ac.uk/ (visited on 05/12/2024).

[23] Jiuzhou Chemicals. *Conversion of particle size of molecular sieve (mesh and mil)*. July 2020. URL: https://www.jiuzhouchemicals.com/NEWS/Industry/2020/0703/137.html (visited on 05/12/2024).

[24] ALS. *Method Statement*. ALS Services. Apr. 2024.

[25] J. Shoemaker and Dan Tettenhorst. *Method 537.1: Determination of Selected Per- and Polyfluorinated Alkyl Substances in Drinking Water by Solid Phase Extraction and Liquid Chromatography/Tandem Mass Spectrometry (LC/MS/MS)*. Tech. rep. Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, 2018.

[26] Jonathan B Burkhardt et al. "Modeling PFAS removal using granular activated carbon for full-scale system design". In: *Journal of Environmental Engineering* 148.3 (2022), p. 04021086.

[27] Dongqing Zhang et al. "Sorption of perfluoroalkylated substances (PFASs) onto granular activated carbon and biochar". In: *Environmental technology* 42.12 (2021), pp. 1798–1809.

[28] Nicole M. DeLuca et al. "Using Geospatial Data and Random Forest To Predict PFAS Contamination in Fish Tissue in the Columbia River Basin, United States". In: *Environmental Science & Technology* 57.37 (2023), pp. 14024–14035. DOI: 10.1021/acs.est.3c03670.

[29] ML Brusseau, RH Anderson, and B Guo. "PFAS concentrations in soils: Background levels versus contaminated sites". In: *Science of the Total Environment* 740 (Oct. 2020). Epub 2020 Jun 6, p. 140017. DOI: 10.1016/j.scitotenv.2020.140017.

[30] Jessica Ferm, Dimitrios Panayotopoulos-Tsiros, and Sam Griffiths. "Planning Urban Manufacturing, Industrial Building Typologies, and Built Environments: Lessons From Inner London". In: *Urban Planning* 6 (2021), pp. 350–367. DOI: 10.17645/up.v6i3.4357.

[31] Nadine Belkouteb et al. "Removal of per-and polyfluoroalkyl substances (PFASs) in a full-scale drinking water treatment plant: Long-term performance of granular activated carbon (GAC) and influence of flow-rate". In: *Water Research* 182 (2020), p. 115913.

[32] Nicholas J Herkert et al. "Assessing the effectiveness of point-of-use residential drinking water filters for perfluoroalkyl substances (PFASs)". In: *Environmental Science & Technology Letters* 7.3 (2020), pp. 178–184.

[33] Minkyu Park et al. "Magnetic ion-exchange (MIEX) resin for perfluorinated alkylsubstance (PFAS) removal in groundwater: Roles of atomic charges for adsorption". In: *Water research* 181 (2020), p. 115897.

[34] AM Fidar, FA Memon, and D Butler. "Performance evaluation of conventional and water saving taps". In: *Science of the total environment* 541 (2016), pp. 815–824.

[35] Walter P White. "Estimatinig Impurities by Means of the Melting Point Curve". In: *The Journal of Physical Chemistry* 24.5 (2002), pp. 393–416.

[36] P Servais et al. "A pilot study of biological GAC filtration in drinking water". In: *Aqua* 41.3 (1992), pp. 163–168.

[37] US Environmental Protection Agency. "Treatment Options for Removing PFAS from Drinking Water: Fact Sheet". In: (Apr. 2024). URL: https://www.epa.gov/system/files/documents/2024-04/pfas-npdwr_fact-sheet_treatment_4.8.24.pdf.