



# Spatial and temporal analyses of perfluorooctanoic acid in drinking water for external exposure assessment in the Ruhr metropolitan area, Germany

## The 'PerSpat'-Project

Jonathan Rathjens<sup>1</sup> · Eva Becker<sup>1,2</sup> · Arthur Kolbe<sup>2</sup> · Katja Ickstadt<sup>1</sup> · Jürgen Hölzer<sup>2</sup>

Accepted: 7 November 2020  
© The Author(s) 2020

### Abstract

Perfluorooctanoic acid (PFOA) and related chemicals among the per- and polyfluoroalkyl substances are widely distributed in the environment. Adverse health effects may occur even at low exposure levels. A large-scale contamination of drinking water resources, especially the rivers Möhne and Ruhr, was detected in North Rhine-Westphalia, Germany, in summer 2006. As a result, concentration data are available from the water supply stations along these rivers and partly from the water network of areas supplied by them. Measurements started after the contamination's discovery. In addition, there are sparse data from stations in other regions. Further information on the supply structure (river system, station-to-area relations) and expert statements on contamination risks are available. Within the first state-wide environmental-epidemiological study on the general population, these data are temporally and spatially modelled to assign estimated exposure values to the resident population. A generalized linear model with an inverse link offers consistent temporal approaches to model each station's PFOA data along the river Ruhr and copes with a steeply decreasing temporal data pattern at mainly affected locations. The river's segments between the main junctions are the most important factor to explain the spatial structure, besides local effects. Deductions from supply stations to areas and, therefore, to the residents' risk are possible via estimated supply proportions. The resulting potential correlation structure of the supply areas is dominated by the common water supply from the Ruhr. Other areas are often isolated and, therefore, need to be modelled separately. The contamination is homogeneous within most of the areas.

**Keywords** Perfluorooctanoic acid · Drinking water contamination · Temporal GLM regression · River modelling · Two spatial level data

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00477-020-01932-8>) contains supplementary material, which is available to authorized users.

---

✉ Jonathan Rathjens  
rathjens@statistik.tu-dortmund.de

<sup>1</sup> Chair of Mathematical Statistics with Applications in Biometrics, TU Dortmund University, Dortmund, Germany

<sup>2</sup> Department of Hygiene, Social and Environmental Medicine, Ruhr-University Bochum, Bochum, Germany

## 1 Introduction

### 1.1 Toxicological background

Per- and polyfluoroalkyl substances (PFASs) name a group of man-made persistent organic chemicals produced since the 1940s which are used in a wide variety of products used by consumers and industry (OECD 2004; Buck et al. 2011). Their biochemical persistence and stability have led to pollution of the environment worldwide and an internal exposure of the general human population (cf., e.g., Buck et al. 2011; OECD 2004). Perfluorooctanoic acid (PFOA) and perfluorooctane sulphonic acid (PFOS) are readily absorbed after ingestion or inhalation, not metabolized and

slowly excreted via urine and feces. Mean human half-lives are estimated to be 3.5–3.8 and 4.8–5.4 years, for PFOA and PFOS, respectively (Olsen et al. 2007). The human population's most important source is food, with contaminated drinking water predominating if present. Therefore, an increased water contamination may be used as a surrogate marker of the human internal exposure, where this is considerably above background level: Consumption of contaminated drinking water is associated with increased PFOA-concentrations in human plasma (Hölzer et al. 2008). Tap water contributes to more than 75% of the estimated daily intake in comparison to other exposure pathways (diet, air, consumer articles) in highly affected regions (Vestergren and Cousins 2009).

The toxicity of PFASs is currently being discussed among toxicologists, epidemiologists and regulatory agencies. Developmental and liver toxicity (Lau et al. 2004; Lau 2012), elevated blood lipids (Frisbee et al. 2010; Steenland et al. 2009) and immunotoxicity (NTP 2016) seem to be major health effects associated with exposure to PFOA and PFOS. Other relevant impairments to health, e.g. the incidence of type 2 diabetes in humans (Sun et al. 2018), have been reported. Meta-analyses of animal experimental and human epidemiological data concluded that developmental exposure to PFOA adversely affects human health based on sufficient evidence of decreased fetal growth in both human and non-human mammalian species (Johnson et al. 2014; Koustas et al. 2014; Lam et al. 2014). However, the mechanisms of toxicity have not yet been sufficiently clarified. Detailed toxicological assessments have been published lately (EPA 2016; ATSDR 2018; EFSA CONTAM 2018; UBA HBM Commission 2018; NJDEP 2019).

The European Food Safety Authority (EFSA) lowered the tolerable daily intake from 1500 (150) (EFSA 2008) to 0.9 (1.9) ng per kg body weight for PFOA (PFOS) and concluded that a considerable proportion of the population currently is exceeding these levels (EFSA CONTAM 2018). Only recently, EFSA also evaluated the exposure to the sum of four different PFASs and recommended an even lower tolerable *weekly* intake of 4.4 ng per kg body weight for the sum of PFOA, PFOS, PFHxS and PFNA (EFSA - CONTAM 2020). The German Drinking Water Commission lowered their action guidance value for pregnant women, breastfeeding mothers and infants under 24 months of age to 50 ng PFOA per litre of drinking water (Trinkwasserkommission 2020). Among the effects evaluated by the EFSA and the drinking water commission are lipid metabolism, immunotoxicity and developmental effects.

## 1.2 Contamination event and drinking water data base

A large-scale contamination of drinking water with PFOA and other PFASs has been discovered in the summer of 2006 in North Rhine-Westphalia (NRW), Germany (Skutlarek et al. , 2006). The principal cause has been the use of PFASs-polluted conditioner (fertilizer) on more than 1300 farmlands with subsequent contamination of surface waters. Further PFOA-contaminations of both surface and groundwater by sewage plants and other sources like fire extinguishing foam have occurred in different parts of NRW (LANUV 2011), though with minor effects on drinking water.

PFASs are generally not removed during riverbank filtration and conventional water processing; some water supply stations have subsequently installed activated charcoal filters in order to reduce the PFOA-concentrations in drinking water below the guidance value of the German Drinking Water Commission (Trinkwasserkommission 2016).

Since 2006, the NRW state environmental agency has implemented an extensive PFASs monitoring programme for relevant environmental media, including soil, water and drinking water (LANUV 2011). The resulting drinking water database has been complemented with data collected by the authors from water suppliers. These data constitute the basis for our analyses.

A particularly severe contamination has affected the river Ruhr and its tributary Möhne (Skutlarek et al. 2006; Wilhelm et al. 2008)). Both rivers are important resources of the regional drinking water: More than 20 important water supply stations are installed along them, supplying a large region of NRW, especially the Ruhr metropolitan area with about a quarter of NRW's 18 million inhabitants.

Human biomonitoring studies have been conducted (Hölzer et al. 2008, 2009) with a focus on the town of Arnsberg, supplied by a Möhne-dependent water supply station near the river's mouth, where PFOA concentrations above 500 ng/l have been measured. (The lifelong tolerable guide value is currently set to 100 ng/l, (Trinkwasserkommission 2016). Furthermore, adverse health effects are suspected to occur even in low ranges of PFOA exposure.)

## 1.3 Aims

This article describes the first phase of our **PerSpat** (Perfluoroalkyl Spatial) project, which has been designed to assess the exposure of the residential population to PFOA, in order to analyse possible associations with relevant

health indicators like birth registry data by means of temporal and spatial modelling in a second phase.

This environmental-epidemiological study on the general population, living in a large region of variable and, in parts, very high PFOA concentrations in drinking water, is the first of this kind on state level and, therefore, importantly contributes to the epidemiological research on developmental toxicity of PFASs.

This article contains the first comprehensive description of the collected and processed complex secondary data and accompanying information on PFOA concentrations measured in drinking water following the NRW contamination incident. In this analysis, we focus on PFOA among the PFASs due to higher data variability, measurements considerably above critical values, and its long half-life.

We evaluate statistical models to find adequate representations of these measurements and to predict the PFOA concentration at particular water supply stations and points of time. Thereby, we encounter a new and unusual data situation in terms of the combination of very instationary temporal processes (in terms of both mean and variance), two spatial levels of interest, and the connection via rivers.

After fitting Generalized Linear Models (GLMs) for temporal regression at the stations, we explore the spatial structure of the results along the well-observed river Ruhr.

A second spatial data level—drinking water supply areas—is distinguished. We numerically estimate the proportions of drinking water from different stations supplying the same area, using meta information on the NRW water supply structure, in order to assess the regional average PFOA concentrations.

#### 1.4 Related modelling approaches

We start our empirical data analysis rather open-minded with regard to model classes. Ultimately, we aim at a model in continuous time and discrete space.

As for approaches in similar scenarios and profound model developments, see, e.g., White et al. (2017) for spatial smoothing models for areas being but partially observed: Data are primarily given on area level with a Markov property assumed, whereas we have to focus on the temporal modelling first and also encounter complicated spatial relationships. We consider approaches like these in future stages of analysis.

Bayesian hierarchical discrete-space approaches, along with intrinsic CAR prior choice considerations, can be found in Keefe et al. (2019). Again, the neighbourhood structure is more specific than found in our situation. Grollemund et al. (2019) present a regression model using functional predictors. These could be representations of the temporal structures in a future spatio-temporal analysis in

our situation; before, we have to explore the temporal data patterns.

As a frequentistic alternative, Tang et al. (2019) develop a semi-parametric copula model to cope with the spatio-temporal dependence and to predict values for new points of time and sites. This is rather restrictive, being Markovian in time, whereas we model continuous, irregularly measured temporal data. Wang and Sun (2019) propose a spatial regression model with local polynomials modelling the coefficients as a function of space, but without an explicit temporal component.

The set-up of consistent spatial correlation matrices along more complex river networks than ours is described as a particular challenge (cf., e.g., Peterson et al. 2007). With more data than we have from groundwater-stations not directly adjacent to a river, the latter could be regarded as an inhomogeneous line source of contamination (cf., e.g., Ayub et al. 2019). Another option would be the incorporation of more detailed river network features, such as water amounts, stream velocity and smaller tributaries (cf., Ver Hoef et al. 2006), if corresponding data were available.

Further methods and application scenarios for risk assessment, when measurement data are available at more steps along water supply systems than in our situation, can be found in Roozbahani et al. (2013). Causal remote effect estimation, when both influence and outcome quantities are measured along a river and over time, unlike in our situation, is performed by Saul et al. (2019) via the parametric  $g$ -formula.

#### Article overview

Below, the data are described in more detail in Sect. 2, with an overview on their amount and features (Sect. 2.1), an outline of the NRW river network and its usage for drinking water (Sect. 2.2) as well as the typical temporal structures to be found in the various locations (Sect. 2.3). In Sect. 3, the data are explored with regard to both time and space: Regression models on time are presented in Sect. 3.1, with some findings concerning the spatial structure along the river Ruhr. Section 3.2 deals with the question whether and how to deduce from water supply stations to the contamination of their supplied areas. Section 4 concludes with a discussion, including some preliminary spatial modelling approaches, and perspectives.

## 2 Data description

The PFOA measurements have been conducted by the local drinking water suppliers; the data have been collected by the NRW State Agency for Nature, Environment and Consumer Protection (*Landesamt für Natur, Umwelt und*

*Verbraucherschutz*; LANUV) and by the Department of Environmental Medicine of the Ruhr-University Bochum.

All data preprocessing steps, visualisations and analyses have been conducted using the R environment (R Core Team 2020) along with its basic packages. Processing of geographical data and drawing of maps have been conducted using the R packages *rgdal* (Bivand et al. 2019) and *tmap* (Tennekes 2018).

Some of the maps have been produced using geographical data from the German Federal Agency for Cartography and Geodesy (© GeoBasis-DE / BKG, 2019; license: [www.govdata.de/dl-de/by-2-0](http://www.govdata.de/dl-de/by-2-0); data: wms\_vg250-ew, wfs\_d1m1000. Geographical data on the NRW drinking water supply (Figs. 1 and 2) have been provided by the LANUV.

## 2.1 Data amount and features

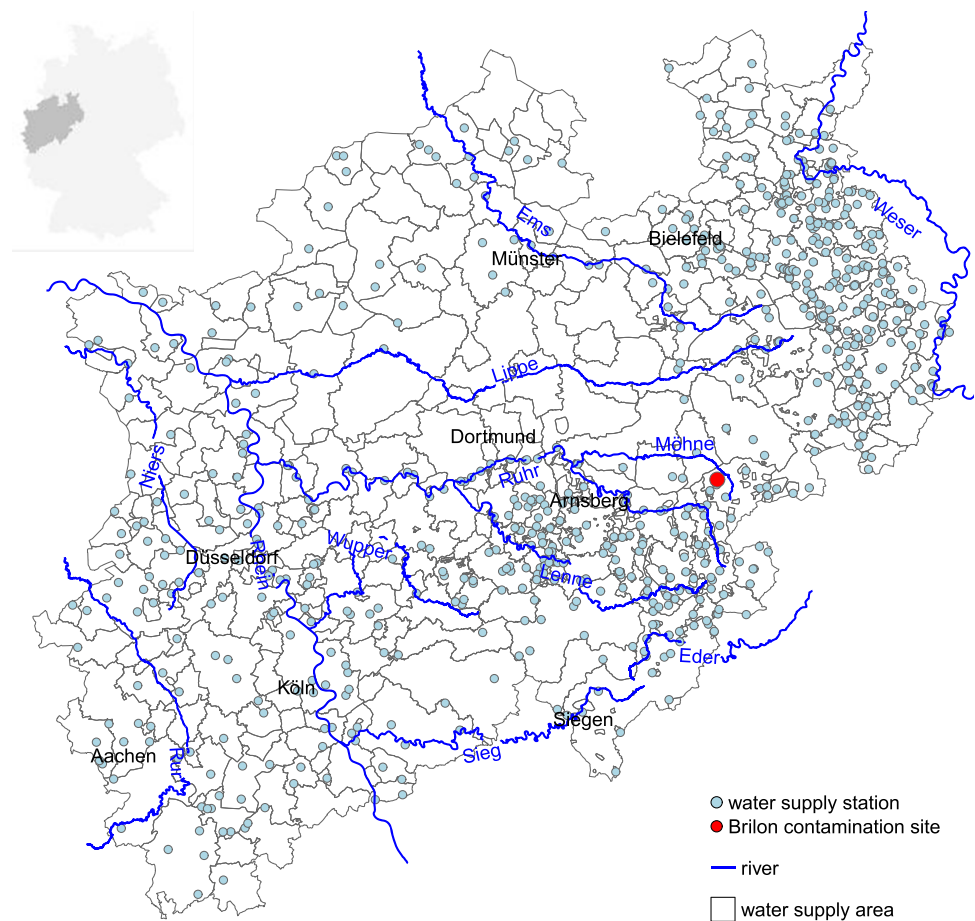
The data consist of measurements from completely purified, ready-to-drink water in both water supply stations (3349 observations) and networks (536) since summer 2006. Observations are available until 2016 and partially

ongoing. Furthermore, there are some samples drawn from raw water (418) and during water treatment (87) at the stations. Many samples, especially the latter, are *non-detects*. See Fig. 2 for a cross-sectional overview.

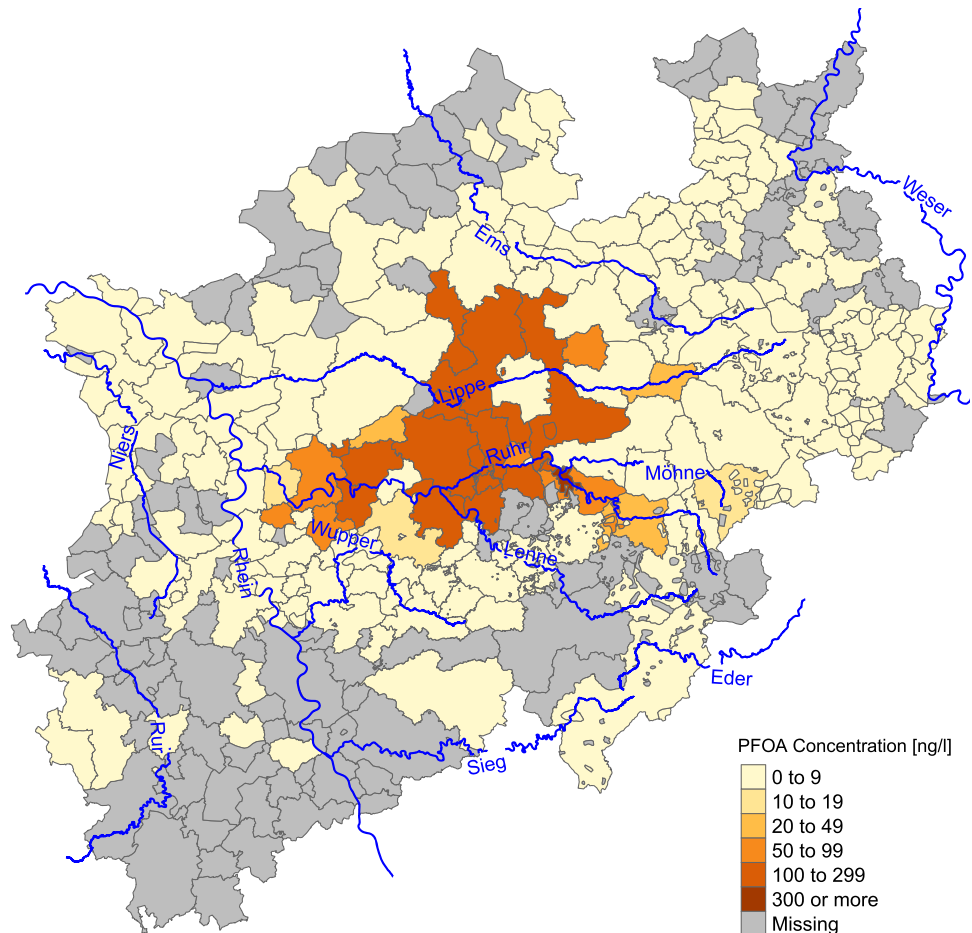
NRW is geographically divided in some 450 water supply areas, with any data available from about 200. For the stations, many of them being very small, any data are available from about 250 of 700. There is a *complex*, neither injective nor surjective *assignment* between areas and stations, with some minor changes in areal division, station running and assignment over the years. The number of *non-equidistant measurements* per time period varies between locations and periods according to the contamination risk.

More detailed information on the amount of data per station and their ranges and temporal patterns can be found in Table 1. According to this, the data availability is satisfying for all affected stations, i.e., along the river Ruhr (a total of 2484 observations at all of the 27 stations), and all of them feature at least rather high values; stations not depending on the river Ruhr are much less observed, typically with low values, although there are exceptions. Table 2 details the data availability with respect to water supply areas: Where some of the stations supplying a

**Fig. 1** NRW water supply areas, water supply stations and main rivers used as drinking water resources. An important contamination discovered in 2006 is located near the Möhne spring. (Above: NRW within Germany)



**Fig. 2** PFOA concentration per water supply area measured in the respective network. The data closest to 1 July 2006 are used. Where no network samples are available, the respective supplying stations' measurements are averaged with weights according to Sect. 3.2.1. (With the limit of quantification at 10 ng/l, the respective areas in light yellow are non-detects)



**Table 1** Number of water supply stations by data availability and temporal data pattern of the PFOA concentrations. (In this rough classification of typical observed situations, 'very high' means PFOA concentrations about and above the guide value of 100 ng/l, 'rather high' values are about and above 50 ng/l. LoQ: limit of quantification)

Data Availability / Information	Temporal Data Pattern	Number of Stations	
		Ruhr	others
many (more than 25 data)	decrease, with very high values	17	1
	decrease, with rather high values	6	0
	diffuse, with rather high values	4	1
	all values below LoQ	0	1
few	single very high value	0	2
	decrease, with rather high values	0	3
	diffuse, with rather high values	0	1
	low values	0	16
	all values below LoQ	0	142
raw water / treatment	low values	0	4
	all values below LoQ	0	53
expert statement	(no contamination risk)	0	56
none	—	0	≈ 400
	total	27	≈ 675

**Table 2** Number of water supply areas by contamination risk and data availability per station: Considering the water supply stations with a relevant supply for a particular area (estimated share of more than 5%, Sect. 3.2.1), we distinguish whether all or some of these stations are potentially affected by the Möhne-Ruhr contamination incident, and whether any data are available from all or some of these stations (right the respective numbers of areas with only one supplying station)

Ruhr Contamination – Data Availability	Cases (Areas)	Trivial Cases (1 Station)
all stations affected	23	14
– all with data	23	14
– some with data	0	—
– none with data	0	0
some stations affected	9	—
– all with data	8	—
– some with data	1	—
– none with data	0	—
no station affected	423	271
– all with data	181	110
– some with data	27	—
– none with data	215	161

particular area are Ruhr dependent, there are typically enough data to assess the area's contamination risk by deduction from the stations.

An important issue with regard to modelling are *outstanding values* at the beginning of the observed period of time: The most affected supply station of Mönnebogen at the Ruhr's tributary Mönne near Arnsberg features a PFOA maximum of 640 ng/l. On the other hand, values *below the limit of quantification* (LoQ, mostly 10 ng/l) are frequent in less affected regions and later periods of time. A certain discretisation is caused by a data accuracy of 10 ng/l frequently given.

Apart from the measurement data, there is *further information* from water supply operators on the most likely non-existence of PFOA contamination at certain stations or areas over longer periods of time. For some places, where there are but a few values, or even just one, the measurement usually confirms the water supplier's assumption of no relevant PFASs contamination.

With data being on *two spatial levels*, stations and areas, there is also information on their *assignment* and the supplied and demanded water amounts.

In summary, the PFOA concentrations of all water supply stations along the river Ruhr are well observed, beside water network data from many water supply areas depending on them. For other regions, there are fewer measurements per station—usually non-detects— or none. The relationship of stations and areas is explored in Sect. 3.2.1.

## 2.2 Spatial connectivities via the river network

The main rivers of NRW, with water supply stations along them, are shown in Fig. 1. (A graph of their directions of flow and their junctions is shown in the Online Resource, Fig. 1: a simplified spatial structure, comprising of river segments. As explored in Sects. 3.1 and 3.2.2, a certain homogeneity may be presumed for river-depending stations within the same segment.)

For the river Ruhr, there are five water supply stations upstream beyond the Mönne mouth, the station of Mönnebogen on the river Mönne, and twenty-one stations downstream from both rivers' junction. With the main PFOA source drained off by the Mönne, the former five stations are less affected and may be neglected in some analyses, yielding a one-dimensional spatial structure, if needed for simplified models.

Apart from the main rivers, there are many stations taking water from smaller tributaries. A weak dependence between these and the respective main river segments may be presumed.

## 2.3 Temporal data patterns for individual water supply stations

We focus on water supply stations along the river Ruhr, with numerous measurements covering a longer period of time. For data availability and patterns, cf., Table 1. According to this, there are two main types of temporal data patterns, a rather diffuse one (Fig. 3) and a striking decrease (Fig. 5).

The first typical temporal structure is found at stations less affected by the contamination incident, being upstream beyond the Mönne mouth (like Mengesohl, Fig. 3 left), partly supplied by groundwater (like Volmarstein, centre) or far downstream at the river Ruhr (like Styrum West, right). Measurements observed here feature no or hardly detectable trends and comparatively small variability. A certain decline in background pollution may be reflected, too.

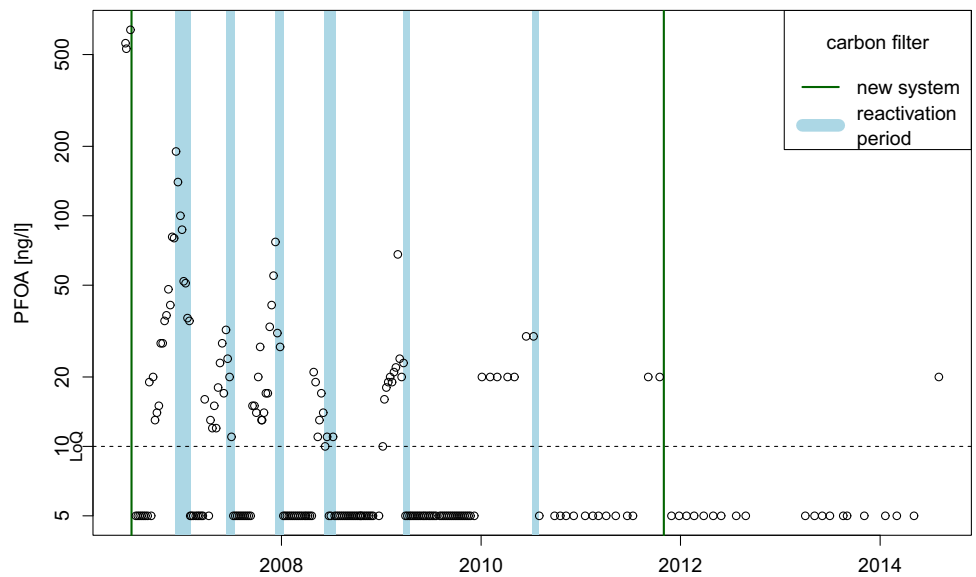
A second type is a more or less smooth decline over the years, though some short-term trends may be distinguished. The decrease is considerably stronger for the first couple of years (e.g., Essen-Überruhr, Fig. 5).

This decrease at the beginning is the strongest at the *outstanding station of Mönnebogen*, where some water filtering measures have taken place after discovery of the contamination. Filters eliminate all PFASs after installation, become gradually less efficient and have to be reactivated after a while (Fig. 4). Thus, some outliers, *change points*, and even *seasonality* may be included in analyses of



**Fig. 3** Measurement series of PFOA of exemplary water supply stations with comparatively low PFOA values and no or a weak trend. Left: Mengesohl; centre: Volmarstein; right: Styrum West

**Fig. 4** Measurement series of PFOA of Möhnebogen with times of carbon filtering interventions



this station—and of some others downstream which have installed filters too, if filtering data were available.

Apart from these differing observed types of temporal structures, there are periods of missing data in some locations: The density of measurements in the course of time and space varies widely. Thus, extrapolation should be performed with caution, especially when using rather simple regression or interpolation methods.

In summary, there are two typical trends in the PFOA values of water supply stations at the river Ruhr—beside the special case of Möhnebogen station. In particular, a characteristic form of ‘decreasing decrease’ can be distinguished. Apart from the Ruhr-dependent stations, there are but a few with sufficient data to identify a trend (one example are the stationary measurements around the limit of quantification in Haltern, see Fig. 9).

### 3 Data exploration

The first part of this Section deals with the temporal modelling of the PFOA concentration data per water supply station along the river Ruhr: The suitability of various regression models is studied and the prediction results of the most appropriate one are discussed with regard to spatial structures along the river. The water supply areas are considered in Sect. 3.2, especially whether and how statements on their PFOA contamination can be deduced from their respective supplying stations.

Let  $X_{it}$  denote the PFOA concentration at water supply station  $i = 1, \dots, m$  (in our situation,  $m \approx 700$ ) for time  $t = 1, \dots, T$ . Reasonable units of time are months or days. Let  $j = 1, \dots, n$  ( $n \approx 450$ ) denote the water supply area.

### 3.1 Temporal regression for individual water supply stations

#### 3.1.1 Model choice

The data are continuous with respect to time with an irregular measurement process: There are periods of time with a high ‘data density’ (up to two measurements within a few days) and others without any data for several month. Therefore, we use regression models with the time as a covariate instead of autoregressive models. We fit models individually to the data of a water supply station  $i$ , but are mainly interested in obtaining a consistent approach along the river Ruhr.

An appropriate model has to cope with the varying shapes of decrease (cf., Figs. 3 and 5), which may also

imply some heteroscedasticity. We consider several Generalized Linear Models (GLMs, Table 3) with various distribution families for  $X_{it}$  and link functions

$$g(\mathbb{E}X_{it}) = g(\mu(t)) = \eta(t) = \beta_0 + \beta_1 t$$

as well as data transformations.

For the latter, a pre-transformation like  $\ln(X_{it})$ ,  $\sqrt{X_{it}}$  or  $\frac{1}{\sqrt{X_{it}}}$  and usage of a simple linear regression leads to very large residuals for periods of high measurements. For the pre-transformation  $\frac{1}{X_{it}}$ , the instability of the estimation is even worse. Additionally, results are difficult to interpret, when re-transformed after the model fit.

Instead of that, while sticking with the simple linear regression, we experimentally transform the time covariate: Among the usual fractional polynomials, the model

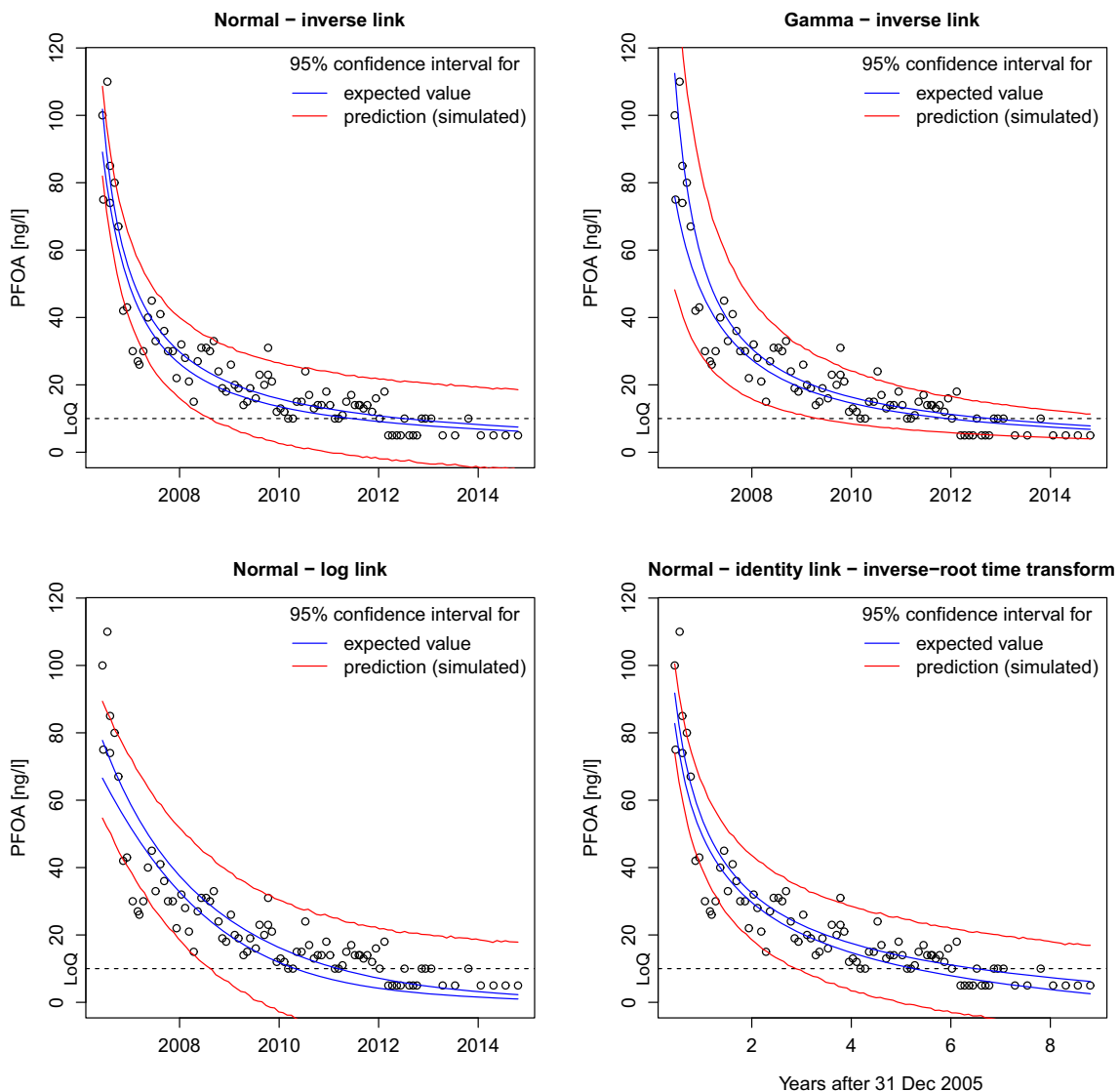


Fig. 5 (GLM) Regression on time: confidence and prediction intervals of various models fitted to the data of Essen-Überruhr



**Table 3** Goodness of fit of various GLMs (distribution families with link functions) in terms of the residual sum of squares by water supply station along the river Ruhr

Station	Normal			Gamma	Normal, identity
	Identity	Log	Inverse	Inverse	Time transform
<i>Upper Ruhr</i>					
Hennenohl	<b>2509</b>	2594	2969	3492	2797
Insel	<b>3960</b>	5205	6991	10,436	8761
Mengesohl	<b>4822</b>	6403	8012	12,069	9204
Stockhausen	9309	<b>8349</b>	8448	8655	8913
Langel	<b>14,313</b>	15,219	18,680	26,437	20,501
<i>Middle Ruhr</i>					
Echthausen	88504	45,406	<b>14,310</b>	21,009	27,617
Warmen	113,172	35,174	<b>27,297</b>	92,980	33,468
Fröndenberg	20,550	<b>14,131</b>	16,808	23,934	14,807
Ruhrtal	42,814	30,823	<b>23,334</b>	23,354	23,793
Halingen	52,522	35,329	<b>21,342</b>	24,756	22,985
Hengsen	45,081	26,048	<b>20,154</b>	20,515	20,240
Villigst	117,266	54,163	<b>32,504</b>	33,629	39,224
Ergste	52,451	34,579	<b>22,754</b>	25,543	25,000
Westhofen 1	83,985	39,238	<b>21,610</b>	22,606	29,586
Westhofen 2	52,549	30,258	<b>21,761</b>	21,861	22,315
<i>Lower Ruhr</i>					
Hengstey	131,908	<b>79,197</b>	84,974	130,458	81,677
Volmarstein	3501	3207	<b>3110</b>	<b>3110</b>	3122
Ruhrstrasse	6356	5454	<b>5061</b>	5132	5096
Witten	3926	<b>3007</b>	3342	3740	3336
Stiepel	12,004	8328	7003	7146	<b>6886</b>
Essen-Horst	17,017	7938	5645	6071	<b>5572</b>
Essen-Überruhr	16,607	8839	<b>4616</b>	4710	5053
Essen-Kettwig	11,580	6569	<b>4859</b>	5016	5358
Dohne	13,437	11,544	<b>10,594</b>	10,679	10,912
Styrum Ost	<b>7256</b>	7386	7499	7682	7823
Styrum West	5738	5499	5103	5272	<b>5087</b>

Best fitting model per station in bold

$$\mathbb{E}X_{it} = \beta_0 + \beta_1 \frac{1}{\sqrt{t}}$$

results in a particularly good fit. However, it is not translation invariant and we have to choose a starting point  $t_0$ , preferably close to the beginning of the measurements—in our situation the beginning of the year 2006, consistently for all stations. From an environmental scientific point of view, the results are difficult to interpret in terms of the change of concentration during a given span of time.

For GLMs without pre-transformation of data, at first we identify all combinations of continuous distribution families and link functions such that the model fit converges for all stations (shown in Table 3). Among these, a Normal distribution with an inverse link

$$(\mathbb{E}X_{it})^{-1} = \beta_0 + \beta_1 t$$

turns out to yield the best fit in terms of the residual sum of squares for most of the stations, especially those with the typical ‘decreasing decrease’. But also the Gamma distribution with its natural inverse link is often close to it in that sense. Table 3 shows the goodness of fit of the various models. Fig. 5 illustrates their suitability with regard to the estimated mean function as well as their capability to reflect the data’s variance.

While the variance  $\text{Var}(\hat{\mu})$  of the mean response estimator is always a function of time and is further affected in case of non-identity link functions, a proper choice of the data variance function turns out to be more relevant in terms of scale:  $\text{Var}(X_{it})$  is constant for all  $\mu$  in case of the Normal family, leading to a certain probability of negative predictions, but also to a closer prediction interval in the period of high values. For the Gamma family, it holds

$\text{Var}(X_{it}) = \phi^2 \mu^2(t)$ , so we are able to interpret and estimate the measurements' standard error as a constant multiple of the mean, in accordance with some assumptions on measurement technique and natural fluctuation (LANUV 2011); on the other hand, the prediction interval is very large in the period of high values (perhaps corresponding with inferior measurement technique in the early period) and rather small later on (notwithstanding that an increased uncertainty is assumed for measurements close the limit of quantification). However, all models sufficiently cover the data variance.

The inverse link is very suitable for modelling the 'decreasing decrease' type of temporal data patterns; for less definite trends, especially along the upper Ruhr, others would do (cf. Table 3). In some cases, a slight heteroscedasticity remains. Outlying data at the measurements' beginning have a strong leverage that may lead to a certain instability, especially when 'predicting the past'. With the inverse link being superior compared to the log link, the decrease turns out to be steeper than exponential for highly contaminated places.

Any such models are applicable for situations without change points or outliers as described in Sect. 2.3. Where carbon filtering interventions are known, a piece-wise regression is applied.

With regard to easy interpretability of the parameters, the goodness of the predictions, and the perspective model extension (say, by additional spatial terms), we focus on the Normal GLM with an inverse link.

### 3.1.2 Discussion of regression results

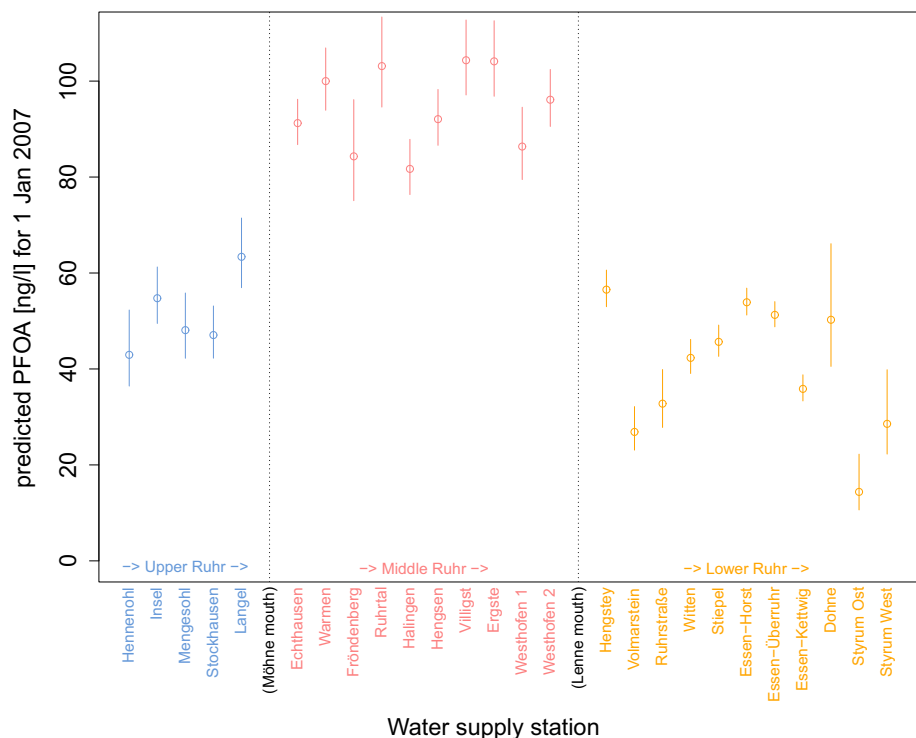
Figure 6 shows a cross-section  $\hat{\mu}(t^*)$  of the mean response estimation results along the river Ruhr. (In the Online Resource, further points of time can be found Fig. 2, a longitudinal overview in Fig. 3 at the same place.)

The results show a clear distinction between the river's segments: The abrupt increase of PFOA concentrations below the junction of Ruhr and Möhne is reflected as well as a decrease below the mouth of the Ruhr's tributary Lenne. The estimations for most of the stations along the lower Ruhr tend to feature smaller uncertainty intervals.

Within the segments, no definite trend is recognizable; local effects of the individual stations seem to overlay a trend along the river. Potential causes may be the varying parts of groundwater in relation to surface water used by the water supply stations as well as their filtering measures. Another aspect are further minor PFOA discharges, especially from sewage plants affecting the river Ruhr (cf. LANUV 2011). In fact, there seems to be a slightly increasing trend along parts of the lower Ruhr.

To conclude, the GLM with Normal distribution and an inverse link function is a convenient way of functional representation of temporal trends in PFOA data of the water supply stations along the river Ruhr. This also opens a possibility for integrating spatial trends in the model. The latter should respect the river segments.

**Fig. 6** Predictions (with approximate 95% confidence intervals) of the PFOA values on 1 January 2007, from GLMs with Normal distribution and inverse link, individually for each water supply station along the river Ruhr



### 3.2 Water supply area assessment

#### 3.2.1 Station-to-area supply proportions

With data and models on two spatial levels—water supply stations and water supply areas—we are strongly interested in the proportions, how much of an area’s water stems from the respective stations, in order to mediate between the two levels. However, there is no direct information on this question—apart from rough estimations by a few local drinking water suppliers.

Instead of this, we estimate these supply proportions using data on the stations’ water supply amounts and the areas’ water demands together with the qualitative information which stations supply which areas.

Let  $\mathbf{a} = (a_1, \dots, a_m)^T$  be the supply and  $\mathbf{b} = (b_1, \dots, b_n)^T$  the demands. The data (say, cubic metres per year) are somewhat vague and, therefore, standardized to  $a_i := \frac{a_i}{\mathbf{a}^T \mathbf{1}_m}$  and  $b_j := \frac{b_j}{\mathbf{b}^T \mathbf{1}_n}$ , respectively, in order to obtain matching totals.

Formally, we are in search of a weight matrix

$$\mathbf{W} = \begin{pmatrix} w_{11} & \dots & w_{1m} \\ \vdots & & \vdots \\ w_{n1} & \dots & w_{nm} \end{pmatrix} \in [0, 1]^{n \times m},$$

such that area  $j$  gets the  $w_{ji}$ ’s part of its water from station  $i$ , which means that the water amounts  $w_{ji}b_j$  of the areas sum up to the stations total supply

$$a_i = \sum_{j=1}^n w_{ji}b_j, \quad i = 1, \dots, m, \tag{1}$$

with the area-wise restrictions

$$\sum_{i=1}^m w_{ji} = 1, \quad j = 1, \dots, n. \tag{2}$$

Such a deconvolution problem (cf., e.g., Cutler 1978) is generally not uniquely solvable.

In our situation, the problem is simplified by several facts:

- Only a small number of stations and areas are connected at all. And these relations are known at least qualitatively, albeit not in numbers. Therefore, it is known for far the most of the cases, that a station  $i^*$  does not supply an area  $j^*$  and we have many further restrictions  $w_{j^*i^*} = 0$  when calculating  $\mathbf{W}$ .
- If an area is supplied by a single station, the respective weight  $w_{j^*i^*} = 1$  is known prior to solving the system of (1) and (2), even if this relation is not bijective.

**Table 4** Water supply relations: relevant (> 5%) shares

Supplying Stations per Area Cases (Areas)	1	2	3	4	5	$\geq 6$
	285	81	46	19	8	16
Supplied Areas per Station Cases (Stations)	none	1	2	3	4	$\geq 5$
	22	578	73	16	5	8

- Due to the many zeros in  $\mathbf{W}$ , the equation system of (1) and (2) is actually split in, often very small, partial problems

$$\mathbf{a}_{(l)} = \mathbf{W}_{(l)}^T \mathbf{b}_{(l)}, \quad \mathbf{W}_{(l)} \mathbf{1}_{m_l} = \mathbf{1}_{n_l}, \quad l = 1, \dots, L, \tag{3}$$

solving the non-zero blocks  $\mathbf{W}_{(l)} \in [0, 1]^{m_l \times n_l}$  of  $\mathbf{W}$  concerning  $m_l$  stations and  $n_l$  areas. Moreover, a large number of supply relations are unique, i.e.,  $m_l = n_l = 1$  (cf., Table 4).

Transforming  $\mathbf{W}$  to a diagonal structure of non-zero blocks – with the help of the R packages `igraph` (Csardi and Nepusz 2006) and `Matrix` (Bates and Mächler 2019)—we obtain  $L = 322$  linear equation systems like (3). Of these, 27 are non-trivial, i.e.,  $m_l, n_l \geq 2$ . Each comprises of  $m_l + n_l$  independent equations to solve a certain number of unknown quantities  $w_{ji} \in (0, 1)$ .

In 6 cases, the system is determined. In 20 systems, there are more equations than unknowns and we can find a numerical solution for the least squares problem

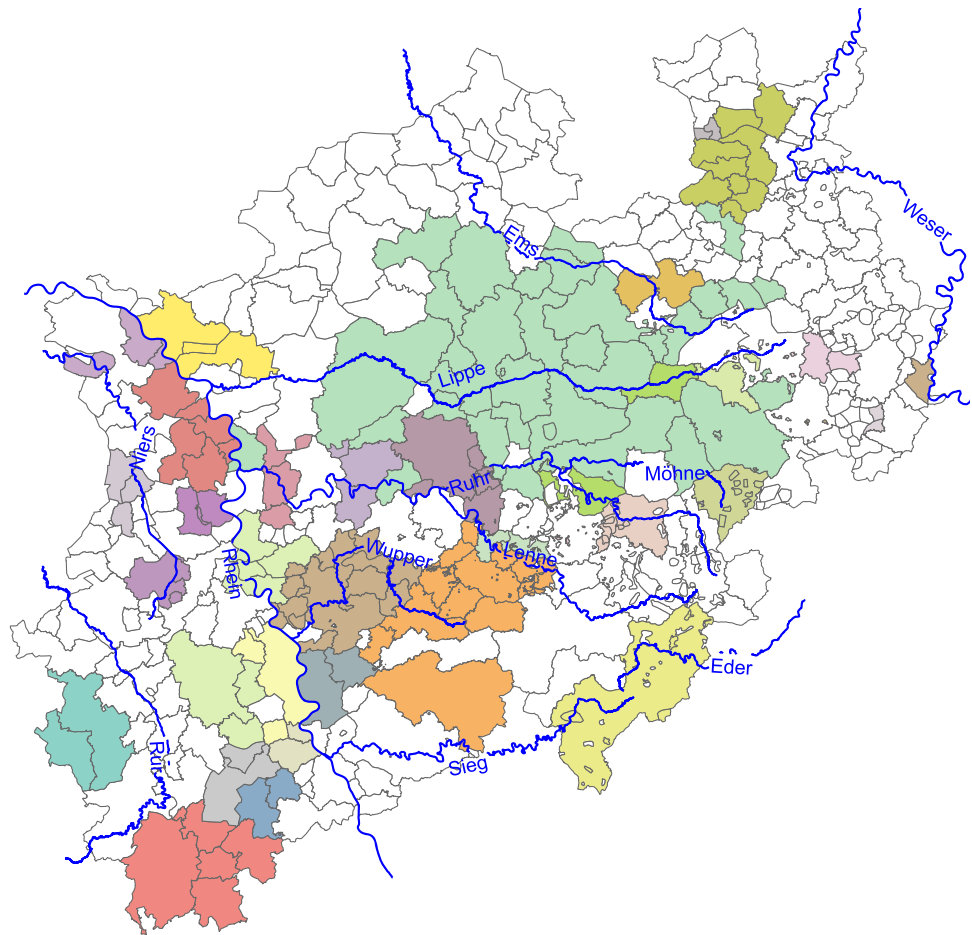
$$\left\| \mathbf{a}_{(l)} - \mathbf{W}_{(l)}^T \mathbf{b}_{(l)} \right\|_2^2 \rightarrow \min_{\mathbf{W}_{(l)}}!, \quad \mathbf{W}_{(l)} \mathbf{1}_{m_l} = \mathbf{1}_{n_l}$$

using the R package `NlcOptim` (Chen and Yin 2019) to perform optimisation with constraints; the obtained estimated proportions are exact enough given the vague data; in particular, the by far largest system ( $m_l = 72, n_l = 82$ ), containing many of the Ruhr dependent stations, is solvable. Only one system is under-determined: it concerns a small region ( $m_l = 3, n_l = 6$ ) not related to the Ruhr and is not considered further.

The obtained proportions of water supply from stations to areas are important to estimate the stations’ relevance for the NRW water supply in general. They are also used to assess the areas’ PFOA burden and thereby the residents’ risk by weighted averaging. This is particularly relevant if the PFOA concentration is inhomogeneous within an area, especially one supplied from both the Ruhr and other rivers or groundwater (Sect. 3.2.2).

The supply matrix  $\mathbf{W}$  also illustrates to what extent the spatial units (water supply areas) are contiguous in terms of shared water: Fig. 7 shows one large region of potentially correlated areas (with the river Ruhr in their south) and

**Fig. 7** NRW water supply areas combined according to their potential correlation (i.e., two areas have relevant ( $> 5\%$ ) shares of water from the same station): A coloured area is correlated with at least one other area of the same colour and not correlated with any area of other colours. White coloured areas are not correlated with any other area



several smaller clusters. On the other hand, there is a large number of single isolated areas.

Therefore, the obtained supply matrix is a key to understand the unusual discrete spatial structure of NRW's drinking water supply. It is the basis for models on supply area level, including analyses of network data.

### 3.2.2 Data homogeneity of water supply areas

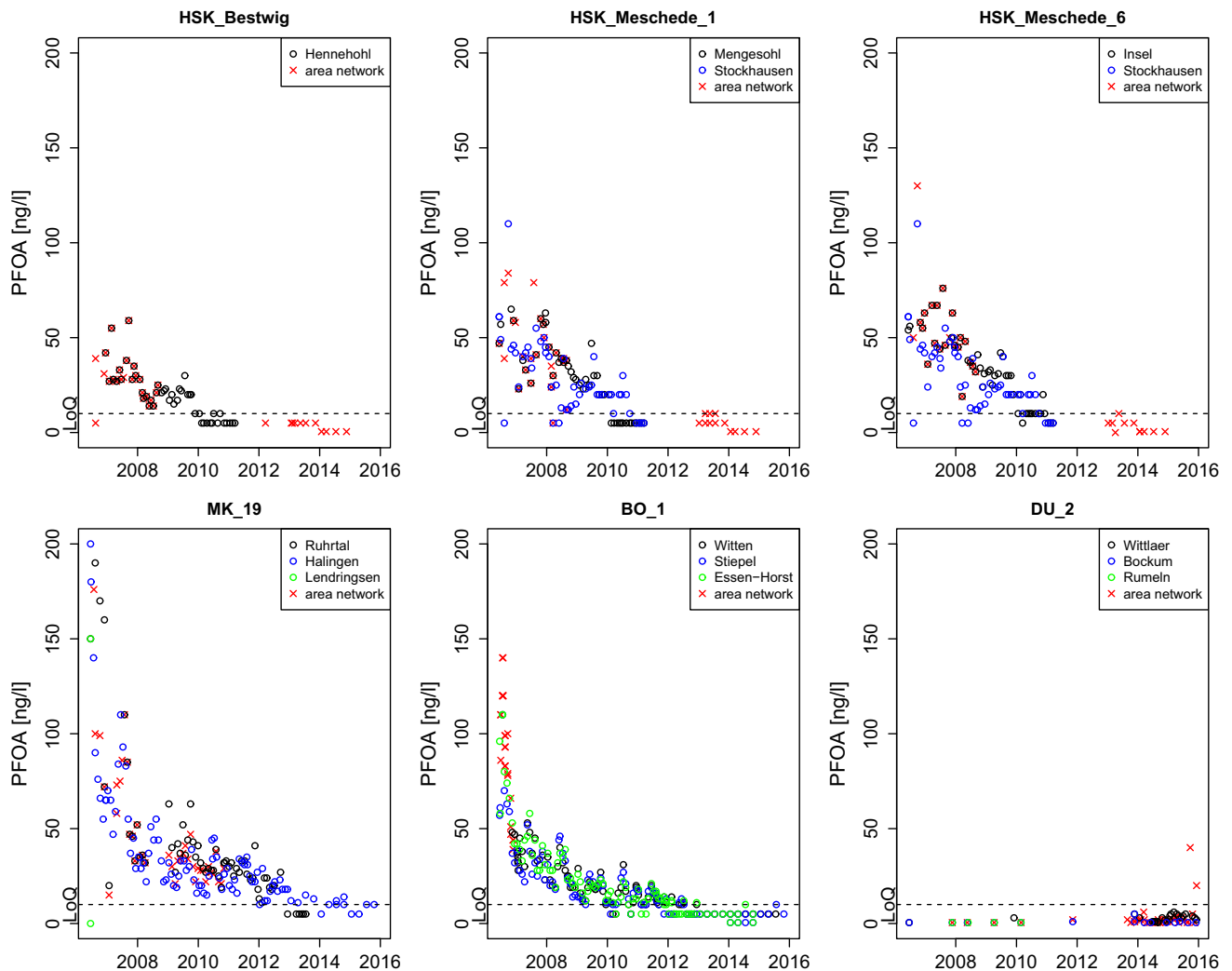
In many cases, PFOA values for water supply areas have to be deduced from the relevant water supply stations' data in order to assess the residents' exposure. We consider whether the involved stations' temporal trends are similar enough ('homogeneous') for joint modelling and whether these data and the respective network data, if any, are sufficiently homogeneous in the course of time.

Homogeneity or inhomogeneity are presumed on the basis of the exploratory results from above, particularly on the two distinguishable main types of temporal data patterns at the stations (Sect. 2.3), the stations' locations with respect to the affected river Ruhr (cf., Sect. 2.2) and the structures found in predictions for Ruhr-dependent stations (Sect. 3.1.2).

As illustrated in Table 1, the occurrence of the typical data patterns is strongly correlated with the stations' dependence from the river Ruhr, where high values with a steep decrease in the course of time can be found. Moreover, these patterns are most similar for stations within the same segment of the river, regarding both data and predictions, where prediction intervals often overlap or are near to each other, along the course of time, as shown in Fig. 6 and in the Online Resource, Fig. 2. In other locations, there are typically low values and few measurements.

Therefore, we presume an area as homogeneous, if it is essentially supplied by stations, which are in the same situation with regard to their location and their temporal data pattern. Examples for the homogeneity of data, especially of such with a decreasing pattern, are shown in Fig. 8.

On the other hand, there are a few areas with some supplying stations depending on the contaminated river Ruhr and some being non-affected (cf., the middle part of Table 2). An example of an inhomogeneously supplied area is shown in Fig. 9: Three stations supply the COE\_3 area, two of them are dependent on the contaminated river



**Fig. 8** Comparison of network and relevant supplying station data for some areas

Ruhr, the third (with a water share of about 40%) is non-affected.

When reliably double-checking the homogeneity, we are limited to cases with sufficient amounts of data from the respective stations or from the area's water network, representing the same periods of time. Larger amounts of station data are mostly given along the river Ruhr (cf., Tables 1 and 2). Network data are rare, especially for the most interesting period of time. (E.g., there are none for the interesting case of the inhomogeneous COE\_3 area.) Where available, the range of the network data seems not to substantially differ from the respective concurrent station samples (cf., Fig. 8).

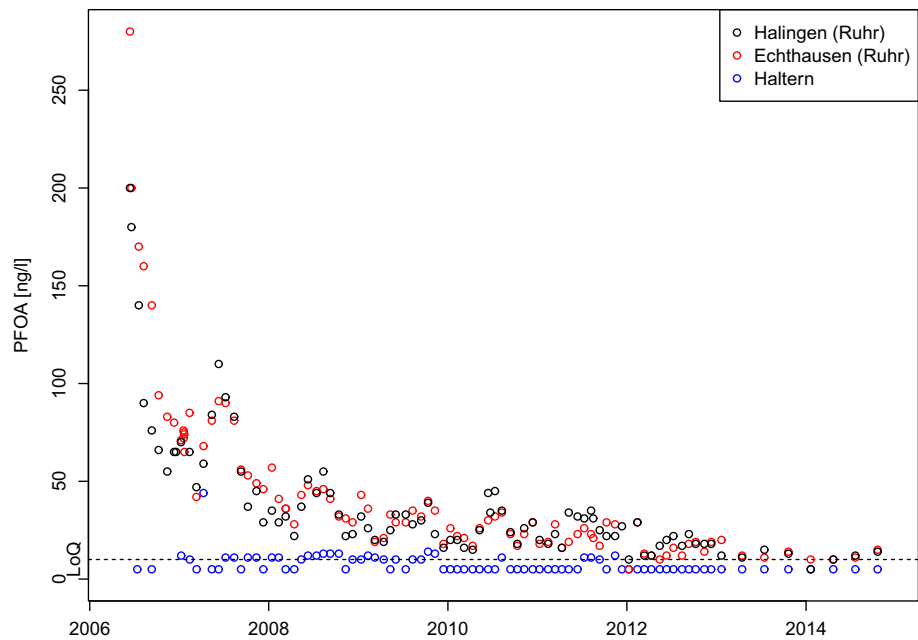
For Ruhr-dependent areas, we combine data from the respective stations, to search for clusters in the space of time and values (using the R package *EMCluster*, Chen and Maitra 2015). The same is done with network data,

where available. It is not possible to distinguish the data and (re-)discover the stations in this way.

Another approach to check whether concurrent station data are distinguishable is a regression model similar to those in Sect. 3.1 for the combined measurement data of an area's supplying stations, with additional dummy variables representing the stations. For areas completely dependent on (a section of) the river Ruhr, the resulting station effects are not significant or at most very weak. For inhomogeneous areas, we can find such effects.

Both approaches let us conclude for the vast majority of water supply areas (cf., the upper and lower part of Table 2), that there is no evidence against a hypothesis of homogeneity, where adequate data are available (i.e., some of the about 30 stations with many data according to Table 1 are involved). This avoids uncertainty by averaging PFOA concentrations. Perspectively, we are able to fit joint models to the respective water supply stations,

**Fig. 9** Measurement series of PFOA of the three water supply stations supplying the COE\_3 water supply area



respecting those without any data, to obtain estimates for a whole area's exposure. Where waters from the river Ruhr and other sources are involved (about 9 areas according to Table 2), this situation is more difficult.

#### 4 Discussion and conclusions

Given the potential health risks of PFASs, especially PFOA, and the contamination incident affecting the rivers Möhne and Ruhr, we explore the measurement data from water supply stations and within the drinking water network.

The data comprise of quite complete series along the river and only few measurements each at other water supply stations. For the Ruhr, we observe a steeply decreasing temporal data pattern at mainly affected stations and a more diffuse structure upstream; later after the contamination, the decrease is weaker. Elsewhere, there are usually non-detects.

We are interested in a consistent temporal modelling approach for comparison of results along the river Ruhr and later building of joint models. A Gaussian Generalized Linear Model with an inverse link function turns out to have the best fit to the data among regression approaches, hinting to a more than exponential contamination decrease after the incident. If there are crucial events such like change points or extreme outliers, a piece-wise regression is applied.

When exploring the spatial structure of the predictions from this model, we find the important role of the river segments (between the main junctions): The PFOA

concentration considerably varies between the Ruhr segments, but less and without clear trends within, where station-related effects and minor additional PFOA discharges may prevail.

We are ultimately interested in the residents' PFOA risk assessment and, therefore, have to deduce from station level data and predictions to the water supply areas. We use an approximate supply proportion matrix obtained from rough information of the water amounts. Furthermore, from our exploratory results, we can presume homogeneity of data from stations along the river Ruhr, supplying the same areas, and find no evidence against it. Therefore, the uncertainty when aggregating data in such areas is low. For those supplied by both the river Ruhr and unaffected stations, it is higher, but aggregation is possible using weights from the proportion matrix.

The potential state-wide spatial correlation structure is explored using the supply proportions. We find one large region of potentially correlated areas near the river Ruhr, but also many other supply areas which are isolated or grouped to small clusters. This information is important to set up a covariance matrix for usage in spatial models, but there are still relevant additional dependencies, induced by river or groundwater connections, which are not directly included.

Useful predictions are not possible for largely unobserved regions or periods of time. Extrapolations are challenges for all methods, especially for the early period prior to the first measurements in 2006. Yet, the duration of the Brilon contamination may be estimated to several years from the fertilizer's quantity, the drainage, and the

residents' internal exposure (Hölzer et al. 2009; Skutlarek et al. 2006).

A lack of data in vast regions also affects the applicability of comprehensive state-wide smoothing models. As a preliminary spatio-temporal approach, we have used a Bayesian conjugate-prior Gamma-Gamma model focussing on the mean surface to predict PFOA values for any water supply station, water supply area and point of time; its uncertainty varies widely according to the presence of data. So, such models are restricted to temporal analyses and to stations along the well-observed river Ruhr. Moreover, water supply areas which are isolated within the supply relationship structure and additional spatial dependencies would have to be respected.

Kriging approaches—using the R packages `Com-pRandFld` (Padoan and Bevilacqua 2015) and `geoR` (Ribeiro et al. 2020)—along the river turn out to be inappropriate, too, since they do not respect the river segments, which dominate the spatial data structure. Furthermore, even for the well observed river Ruhr, the data are too sparse to analyse proper variograms, and sometimes the assumption of isotropy seems violated. Ordinary and universal Kriging lead to sufficient local estimates when cross-validated along the river Ruhr, but perform worse near the Möhnebogen water supply station. It is questionable whether the continuous Kriging solution is an efficient way to merely estimate concentrations at discrete, definite stations, as in our situation.

Topological (Top-) Kriging (using the R package `rTop`, Skøien et al. 2014; cf., e.g., Laaha et al. 2014) is designed for prediction on the level of the rivers' catchment areas. This would be useful to model measurements along the NRW river network, thereby summarizing several stations each. Nonetheless, with the stations' values connected to water supply areal units, it does not seem productive to introduce a third spatial level. Furthermore, we have no geographical data on the actual catchment areas.

In general and state-wide, the problems of a varying data density, up to entirely non-observed regions, and of a regionally different strength of dependence remain hindering for all Kriging approaches. For the river Ruhr, the rather complicated models, especially Top-Kriging, seem exaggerated for an 'linear' spatial river structure. Therefore, we drop these approaches because of the non-suitable spatial data structure. For problems of the performance of spatial methods, especially (Top-)Kriging, compared to regression approaches for river-related catchment area data, see, e.g., Brunner et al. (2018).

Our station-wise regression results lead to the perspective to combine the temporal regression models for the Ruhr data and add a spatial structure. Spatial effects of the river segments and of the individual water supply

stations—to quantify their influence—may be included either in a single linear predictor or hierarchically.

Comprehensive spatial models should respect the complex potential state-wide correlation structure. Given the results in Sect. 3.1, it seems reasonable to focus on the river segments when building up a neighbourhood structure of the water supply stations based on their locations along the NRW river network.

Within our **PerSpat** project, a fixed correlation matrix of the water supply areas based on river dependency and the estimated station-to-area-supply proportions is initially used to apply an algorithm solving a realignment problem (Kohlenbach 2019) and also for the preliminary smoothing approaches. In the future, river-induced correlations may be estimated from the Ruhr data.

Spatial correlations caused by groundwater may be relevant, too, and, therefore, are possible extensions of the neighbourhood structure. The existence of several contamination sources should be taken into account, too.

With our data amount, all such models are reliable only for the river Ruhr, but generally applicable to all main rivers of the state. Our modelling approaches are transferable to other or more complex river systems and to other substances solute in water or to similar scenarios.

Predicted PFOA concentrations in drinking water will be used in further phases of our **PerSpat** project to estimate PFOA concentrations in plasma of NRW residents, depending on time and place of residence, and for the analysis of health related data from the state-wide birth registry.

**Acknowledgements** The authors thank Katharina Olthoff (d. 2019) and Sabine Bergmann at the Department of Water Management and Water Conservation of the NRW State Agency for Nature, Environment and Consumer Protection (LANUV) for the very helpful cooperation and consultation as well as the provision of data on PFASs measurements and water supply structures along with geographical data. We thank Stiftung Mercator for funding parts of our work.

**Funding** Open Access funding enabled and organized by Projekt DEAL. The research has been partly funded by the MERCUR (Mercator Research Center Ruhr) programme of Stiftung Mercator, Essen, Germany, under the project title "Räumliche Modelle zur Analyse und Bewertung von Gesundheitsindikatoren am Beispiel der Trinkwasserbelastung perfluorierter Verbindungen" (spatial models for the analysis and evaluation of health indicators on the example of drinking water contamination by perfluorinated compounds) with the Project Number An-2015-0001.

**Data availability** Measurement data from the relevant drinking water supply stations along the river Ruhr can be obtained from the LANUV at [https://www.elwasweb.nrw.de/elwas-hygrisc/twbericht/pft\\_tw.php?exhibit-use-local-resources](https://www.elwasweb.nrw.de/elwas-hygrisc/twbericht/pft_tw.php?exhibit-use-local-resources) (08/10/2020).

**Code availability** R-Code for the analyses can be provided by the authors upon request.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- ATSDR (2018) Toxicological Profile for Perfluoroalkyls—Draft for Public Comment. Technical report, Agency for Toxic Substances and Disease Registry, <https://www.atsdr.cdc.gov/toxprofiles/tp200.pdf>. Accessed 30 Jan 2020
- Ayub R, Messier KP, Serre ML, Mahinthakumar K (2019) Non-point source evaluation of groundwater nitrate contamination from agriculture under geologic uncertainty. *Stoch Environ Res Risk Assess* 33(4):939–956
- Bates D, Mächler M (2019) Matrix: sparse and dense matrix classes and methods. R package version 1.2-18
- Bivand R, Keitt T, Rowlingson B (2019) RGDAL: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-8
- Brunner MI, Furrer R, Sikorska AE, Viviroli D, Seibert J, Favre AC (2018) Synthetic design hydrographs for ungauged catchments: a comparison of regionalization methods. *Stoch Environ Res Risk Assess* 32(7):1993–2023
- Buck RC, Franklin J, Berger U, Conder JM, Cousins IT, de Voogt P, Jensen AA, Kannan K, Mabury SA, van Leeuwen SPJ (2011) Perfluoroalkyl and polyfluoroalkyl substances in the environment: terminology, classification, and origins. *Integr Environ Assess Manage* 7(4):513–541
- Chen WC, Maitra R (2015) EMcluster: EM algorithm for model-based clustering of finite mixture Gaussian distribution. R package version 0.2-5
- Chen X, Yin X (2019) NlOptim: solve nonlinear optimization with nonlinear constraints. R package version 6
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Syst* 1695(5):1–9
- Cutler DJ (1978) Numerical deconvolution by least squares: Use of prescribed input functions. *J Pharmacokinet Biopharm* 6(3):227–241
- EFSA (2008) Perfluorooctane sulfonate (PFOS), perfluorooctanoic acid (PFOA) and their salts scientific opinion of the panel on contaminants in the food chain. *EFSA J* 653:1–131
- EFSA CONTAM (2018) Risk to human health related to the presence of perfluorooctane sulfonic acid and perfluorooctanoic acid in food. *EFSA J* 16(12):1–295
- EFSA CONTAM (2020) Risk to human health related to the presence of perfluoroalkyl substances in food. *EFSA J* 18(9):1–391
- EPA (2016) Health Effects Support Document for Perfluorooctanoic Acid (PFOA). Technical Report 822-R-16-003, U.S. Environmental Protection Agency, [https://www.epa.gov/sites/production/files/2016-05/documents/pfoa\\_hesd\\_final-plain.pdf](https://www.epa.gov/sites/production/files/2016-05/documents/pfoa_hesd_final-plain.pdf). Accessed 30 Jan 2020
- Frisbee SJ, Shankar A, Knox SS, Steenland K, Savitz DA, Fletcher T, Ducatman AM (2010) Perfluorooctanoic acid, perfluorooctanesulfonate, and serum lipids in children and adolescents: results from the c8 health project. *Arch Pediatr Adolesc Med* 164(9):860–869
- Grolemund PM, Abraham C, Baragatti M, Pudlo P (2019) Bayesian Functional Linear Regression with Sparse Step Functions. *Bayesian Anal* 14(1):111–135
- Hölzer J, Midasch O, Rauchfuss K, Kraft M, Reupert R, Angerer J, Kleeschulte P, Marschall N, Wilhelm M (2008) Biomonitoring of perfluorinated compounds in children and adults exposed to perfluorooctanoate-contaminated drinking water. *Environ Health Perspect* 116(5):651–657
- Hölzer J, Göen T, Rauchfuss K, Kraft M, Angerer J, Kleeschulte P, Wilhelm M (2009) One-year follow-up of perfluorinated compounds in plasma of German residents from arnsberg formerly exposed to PFOA-contaminated drinking water. *Int J Hyg Environ Health* 212(5):499–504
- Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, Robinson KA, Axelrad DA, Woodruff TJ (2014) The Navigation Guide - Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth. *Environ Health Perspect* 122(10):1028–1039
- Keefe MJ, Ferreira MAR, Franck CT (2019) Objective Bayesian analysis for Gaussian hierarchical models with intrinsic conditional autoregressive priors. *Bayesian Anal* 14(1):181–209
- Kohlenbach J (2019) Methodenvergleich zum Realignment räumlicher Daten am Beispiel von Perinataldaten und Trinkwasserbelastung. Master's thesis, TU Dortmund University
- Koustas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S, Robinson KA, Axelrad DA, Woodruff TJ (2014) The navigation guide—evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. *Environ Health Perspect* 122(10):1015–1027
- Laaha G, Sköien JO, Blöschl G (2014) Spatial prediction on river networks: comparison of top-kriging with regional regression. *Hydrol Processes* 28:315–324
- Lam J, Koustas E, Sutton P, Johnson PI, Atchley DS, Sen S, Robinson KA, Axelrad DA, Woodruff TJ (2014) The navigation guide—evidence-based medicine meets environmental health: integration of animal and human evidence for PFOA effects on fetal growth. *Environ Health Perspect* 122(10):1040–1051
- LANUV (2011) Verbreitung von PFT in der Umwelt: Ursachen – Untersuchungsstrategie – Ergebnisse – Maßnahmen. LANUV-Fachbericht 34, Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen, [https://www.lanuv.nrw.de/fileadmin/lanuvpubl/3\\_fachberichte/30034.pdf](https://www.lanuv.nrw.de/fileadmin/lanuvpubl/3_fachberichte/30034.pdf). Accessed 04 Dec 2019
- Lau C (2012) Perfluorinated compounds. In: Luch A (ed) *Molecular, clinical and environmental toxicology, experientia supplementum*, vol 101. Springer, Basel, pp 47–86
- Lau C, Butenhoff JL, Rogers JM (2004) The developmental toxicity of perfluoroalkyl acids and their derivatives. *Toxicol Appl Pharmacol* 198(2):231–241
- NJDEP (2019) Interim Specific Ground Water Criterion for Perfluorooctanoic Acid (PFOA, C8). Technical support document, New Jersey Department of Environmental Protection—Division of Science and Research, [https://www.nj.gov/dep/dsr/support/docs/PFOA\\_TSD.pdf](https://www.nj.gov/dep/dsr/support/docs/PFOA_TSD.pdf). Accessed 12 Feb 2020
- NTP (2016) Immunotoxicity Associated with Exposure to Perfluorooctanoic Acid or Perfluorooctane Sulfonate. NTP Monograph, Office of Health Assessment and Translation—Division of the National Toxicology Program



- OECD (2004) Results of Survey on Production and Use of PFOS, PFAS and PFOA, Related Substances and Products/Mixtures Containing these Substances, OECD Environment, Health and Safety Publications - Series on Risk Management, vol 19. Inter-Organization Programme for the Sound Management of Chemicals, Environment Directorate - Organisation for Economic Co-Operation and Development
- Olsen GW, Burriss JM, Ehresman DJ, Froehlich JW, Seacat AM, Butenhoff JL, Zobel LR (2007) Half-life of serum elimination of perfluorooctanesulfonate, perfluorohexanesulfonate, and perfluorooctanoate in retired fluorochemical production workers. *Environ Health Perspect* 115(9):1298–1305
- Padoan SA, Bevilacqua M (2015) Analysis of random fields using *CompRandFld*. *J Stat Softw* 63(9):1–27
- Peterson EE, Theobald DM, Ver Hoef JM (2007) Geostatistical modelling on stream networks: developing valid covariance matrices based on hydrologic distance and stream flow. *Freshw Biol* 52(2):267–279
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Ribeiro Jr PJ, Diggle PJ, Schlather M, Bivand R, Ripley B (2020) *geoR: analysis of geostatistical data*. R package version 1.8-1
- Roozbahani A, Zahraie B, Tabesh M (2013) Integrated risk assessment of urban water supply systems from source to tap. *Stoch Environ Res Risk Assess* 27(4):923–944
- Saul BC, Hudgens MG, Mallin MA (2019) Downstream effects of upstream causes. *J Am Stat Assoc* 114(528):1493–1504
- Skøien JO, Blöschl G, Laaha G, Pebesma E, Parajka J, Viglione A (2014) *RTOP: an R package for interpolation of data with a variable spatial support, with an example from river networks*. *Comput Geosci* 67:180–190
- Skutlarek D, Exner M, Farber H (2006) Perfluorinated surfactants in surface and drinking waters. *Environ Sci Pollut Res Int* 13(5):299–307
- Steenland K, Tinker S, Frisbee S, Ducatman A, Vaccarino V (2009) Association of perfluorooctanoic acid and perfluorooctane sulfonate with serum lipids among adults living near a chemical plant. *Am J Epidemiol* 170(10):1268–1278
- Sun Q, Zong G, Valvi D, Nielsen F, Coull B, Grandjean P (2018) Plasma concentrations of perfluoroalkyl substances and risk of type 2 diabetes: a prospective investigation among U.S. Women. *Environ Health Perspect* 126(3):037001
- Tang Y, Wang HJ, Sun Y, Hering AS (2019) Copula-based semiparametric models for spatiotemporal data. *Biometrics* 75(4):1156–1167
- Tennekes M (2018) *tmap: thematic maps in R*. *J Stat Softw* 84(6):1–39
- Trinkwasserkommission (2016) Fortschreibung der vorläufigen Bewertung von Per- und polyfluorierten Chemikalien (PFC) im Trinkwasser – Begründungen der vorgeschlagenen Werte im Einzelnen. Tech. rep., Trinkwasserkommission des Bundesministeriums für Gesundheit, [https://www.umweltbundesamt.de/sites/default/files/medien/374/dokumente/bewertung\\_der\\_konzentrationen\\_von\\_pfc\\_im\\_trinkwasser\\_-\\_wertebegrueudungen.pdf](https://www.umweltbundesamt.de/sites/default/files/medien/374/dokumente/bewertung_der_konzentrationen_von_pfc_im_trinkwasser_-_wertebegrueudungen.pdf). Accessed 30 Jan 2020
- Trinkwasserkommission (2020) Empfehlung des Umweltbundesamtes: Umgang mit per- und polyfluorierten Alkylsubstanzen (PFAS) im Trinkwasser. Tech. rep., Trinkwasserkommission des Bundesministeriums für Gesundheit, [https://www.umweltbundesamt.de/sites/default/files/medien/5620/dokumente/twk\\_200826\\_empfehlung\\_pfas\\_final.pdf](https://www.umweltbundesamt.de/sites/default/files/medien/5620/dokumente/twk_200826_empfehlung_pfas_final.pdf). Accessed 19 Oct 2020
- UBA HBM Commission (2018) Ableitung von HBM-I-Werten für Perfluoroktansäure (PFOA) und Perfluoroktansulfonsäure (PFOS) - Stellungnahme der Kommission “Humanbiomonitoring” des Umweltbundesamts. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 61(4):878–885
- Ver Hoef JM, Peterson E, Theobald D (2006) Spatial statistical models that use flow and stream distance. *Environ Ecol Stat* 13(4):449–464
- Vestergren R, Cousins IT (2009) Tracking the pathways of human exposure to perfluorocarboxylates. *Environ Sci Technol* 43(15):5565–5575
- Wang W, Sun Y (2019) Penalized local polynomial regression for spatial data. *Biometrics* 75(4):1179–1190
- White P, Gelfand A, Utlaut T (2017) Prediction and model comparison for areal unit data. *Spatial Stat* 22:89–106
- Wilhelm M, Kraft M, Rauchfuss K, Hölzer J (2008) Assessment and management of the first german case of a contamination with perfluorinated compounds (PFC) in the Region Sauerland, North Rhine-Westphalia. *J Toxicol Environ Health Part A* 71(11–12):725–733

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.