Article:
Jameel, Y, Valle, D and Kay, P (2020) Spatial variation in the detection rates of frequently studied pharmaceuticals in Asian, European and North American rivers. Science of The Total Environment, 724. 137947. ISSN: 0048-9697

https://doi.org/10.1016/j.scitotenv.2020.137947

Spatial variation in the detection rates of frequently studied pharmaceuticals in Asian, European and North American rivers

Yusuf Jameel[1,a], Denis Valle[1] and Paul Kay[2]*

[1] School of Forest Resources and Conservation, University of Florida, Gainesville, Florida, USA.

[2] School of Geography, University of Leeds, Leeds, West Yorkshire LS2 9JT, U.K.

[a] Now at Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.

*Corresponding author: yjameel@mit.edu

Keywords: pharmaceuticals, clustering, big data, water quality

Abstract

Pharmaceutical consumption has expanded rapidly during the last century and their persistent presence in the environment has become a major concern. Unfortunately, our understanding of the distribution of pharmaceuticals in surface water and their effects on aquatic biota and public health is limited. Here, we explore patterns in the detection rate of the most frequently studied pharmaceuticals in 64 rivers from 22 countries using bi-clustering algorithms and subsequently analyze the results in the context of regional differences in pharmaceutical consumption habits, social and environmental factors, and removal-efficiency of wastewater treatment plants (WWTP). To our knowledge, this is the first study to compare several rivers across 3 major continents and systematically analyze them in this framework. We find that 20% of the pharmaceuticals included in this analysis are pervasively present in all the surface waterbodies. Several pharmaceuticals also display low overall positive detection rates; however, they exhibit significant spatial variability and their detection rates are consistently lower in Western European and North America (WEOG) rivers in comparison to Asian rivers. Our analysis suggests the important role of pharmaceutical consumption and population in governing these patterns, however the role of WWTP efficiency appeared to be limited. We were constrained in our ability to assess the role of hydrology, which most likely also plays an important role in regulating pharmaceuticals in rivers. Most importantly though, we demonstrate the ability of our algorithm to provide probabilistic estimates of the detection rate of pharmaceuticals that were not studied in a river, an exercise that could be useful in prioritizing pharmaceuticals for future study.

.

1. Introduction

Pharmaceutical consumption has increased drastically in the last 50 years and is likely to continue increasing in the coming years due to rising population, changing demographic across the globe, and growing availability across the world (Daughton, 2003). The presence of pharmaceuticals and their metabolites in environmental matrices is well established and is a major environmental concern (Beek et al., 2016; Daughton, 2001; Jones et al., 2001; Oaks et al., 2004; Schwarzenbach et al., 2006) . However, there are considerable knowledge gaps on the impacts of pharmaceuticals on aquatic organisms and ecosystems(Botitsi et al., 2007; Brain et al., 2008; Daughton, 2001; Kümmerer, 2009a, 2009b; Santos et al., 2007). With increasing use of gray water in agriculture and in recharging groundwater for future human consumptions , there are also growing concerns on the long-term effects of persistent exposure to pharmaceuticals on public health (de Jesus Gaffney et al., 2015; Grossberger et al., 2014; Jones-Lepp et al., 2012; Webb et al., 2003). Many countries and environmental agencies have recognized their potential detrimental effects and are developing policies to mitigate their impacts (Kaplan, 2013; Peake et al., 2015; Walters et al., 2010).

To evaluate the potential eco-toxicological risks of pharmaceuticals, it is important to measure or model (Amiard-Triquet et al., 2015; Huggett et al., 2003; Johnson et al., 2013; Kehrein et al., 2015; Kostich and Lazorchak, 2008) their concentration in environmental compartments, document their spatiotemporal variability and understand the role of environmental and social factors in determining their presence in the environment. However, there are more than 3000 pharmaceuticals consumed in Europe alone (Donnachie et al., 2016) and exhaustive monitoring of all the pharmaceuticals (and their metabolites) is expensive and impractical. In this regard, statistical analysis (such as meta-analysis, clustering, regression) of large pharmaceutical datasets could be useful in identifying spatiotemporal patterns of pharmaceuticals and their relationship with environmental covariates. This information could then be used to prioritize pharmaceuticals for future studies, assess relationships between pharmaceuticals (for example: which pharmaceuticals always co- occur in a river and which do not), examine pharmaceutical detection patterns across regions, and identify other questions relevant to the risk of pharmaceuticals in surface water (Altenburger et al., 2003; Andrews, 2001; Donnachie et al., 2016; Jones et al., 2002; Kostich and Lazorchak, 2008; Kumar and Xagoraraki, 2010; Rehman et al., 2015). It is however worth mentioning that for each pharmaceutical, a minimum number of analytical measurements is indeed required to understand the relationships between different pharmaceuticals.

Here, we systematically analyze the detection rate (how often a pharmaceutical was positively detected when analyzed) of the 112 most commonly studied pharmaceuticals in 64 rivers from 22 countries using a stochastic block model (also known as a co-clustering or bi-clustering model). Briefly, stochastic block model (SBM) is used for clustering high-dimensional data, where the algorithm simultaneously clusters rows and columns of the data to obtain subgroups of rows and subgroups of columns that exhibit a high correlation (Berkhin, 2006; Govaert, 1995; Hartigan, 1972; Tanay and Sharan Ý Ron Shamir, 2004). A salient feature of the algorithm is its ability to perform robustly even with substantial missing data. The algorithm has been used for analyzing high-dimensional data in many fields, including bioinformatics (Tanay and Sharan Ý Ron Shamir, 2004), text-mining (Dhillon, 2001), ecology (Chi et al., 2017; Hill et al., 2013), and social network analysis (Banks and Hengartner, 2008; Hoff et al., 2002). Figure 1 provides a hypothetical example to illustrate how the algorithm works. For detailed information on SBM and/or co-clustering please refer to (Berkhin, 2006; Govaert, 1995.; Hartigan, 1972).

To our knowledge this is the first study to 1) systematically analyze the spatial patterns in the detection rates of the most commonly studied pharmaceuticals, 2) analyze the role of social and environmental factors, such as wastewater treatment plant (WWTP) efficiency, pharmaceutical consumption habits, population density and hydrological factors, in determining the pattern of pharmaceutical detection rates and 3) estimate the occurrence probability of unanalyzed pharmaceuticals to support analyte prioritization for future study.

2. Methods

2.1 Description of the database and data aggregation

We obtained the pharmaceutical data analyzed in this study from the Measured Environmental Concentration (MEC) database maintained by the German Environmental Agency (UBA, https://www.umweltbundesamt.de/en/database-pharmaceuticals-in-the-environment-0). The database, accessed on 10/01/2018, consists of 123,761 entries of pharmaceuticals and/or their transformation products measured in environmental matrices such as surface water, groundwater, drinking water and WWTP effluent across 71 countries. To our knowledge, this is the most comprehensive global dataset on pharmaceuticals available. For details on the database please refer to UBA website and Beek et al., (2016). Majority of the data in the database were from 2001 to October 2013. Only 1281 entries in the database predated 2001 and there were no entries after October 2013.

2.2 Rationale for analyzing detection rates of pharmaceuticals

Instead of analyzing measured concentrations reported in the literature from where the data were obtained, we transformed the data into presence/absence format for several reasons. First, the majority of the studies measuring pharmaceuticals during the last two decades have not followed internationally/regionally established protocols (Ort et al., 2010) with minimal information on uncertainty associated with the measurements. Second, most of the pharmaceuticals included in our analysis have been measured less than 5 times on a river with limited or no information on the prevailing hydrological conditions. As a consequence, using a statistical estimate (such as mean or mode) can lead to incorrect characterization of the concentration if all the measurements were done only within a single hydrologic regime (for e.g. river low-flow season). Finally, several studies often report different summary statistics (e.g., mean, median or maximum concentration), typically based on very different sample sizes, hindering a straight-forward comparison of these concentration values. Due to these limitations, we believe that reducing the data to present/absent format was the most reliable and robust way to minimize measurement uncertainties while capturing the majority of the data published over the last two decades.

2.3 Rationale for analyzing pharmaceutical data on basin scale instead of national scale

While there have been previous global, continental and country level analyses on river systems to identify and understand spatiotemporal variability in pharmaceutical occurrence (Barnes et al., 2008; Hughes et al., 2013; Jiang et al., 2013; Klečka et al., 2009; Loos et al., 2010), none to our knowledge have performed statistical analysis to explore global patterns in pharmaceutical occurrences in surface waterbodies and understand the factors determining these patterns. A primary motivation for basin-scale data analysis was the high variability in data availability between national datasets with some countries (such as Germany or USA) having an order of magnitude or more data than others. Importantly, pharmaceutical measurements when organized by river basins are more evenly distributed and less skewed (supplementary material, Figure S1), thus allowing more robust statistical comparisons.

2.4 Statistical analyses

2.4.1 Pharmaceutical Contamination Index

For each river *i, we* calculated the mean detection rate or River Contamination Index (RCI) using the following formula

$$RCI_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{P_{i,j}}{T_{i,j}}$$

where $P_{i,j}$ and $T_{i,j}$ are the number of times pharmaceutical *j* was positively detected and measured at river *i,* respectively. In this expression, n_i is the number of unique pharmaceuticals measured at river *i*. An RCI value of 1 means that all pharmaceutical analytes assessed in river i were detected and a value of 0 means that none of the pharmaceuticals measured at river *i* were ever detected.

2.4.2 Stochastic Block Model

For each river, we determine the number of times a pharmaceutical was analyzed and positively detected. We arranged our data in a format where each row represents a river and each column represents a unique pharmaceutical. The model groups together rivers and pharmaceuticals that have similar detection rate and output subgroups (also called blocks) that are similar. We used SBM in our analysis as it not only allows us to identify rivers groups and pharmaceutical clusters with similar detection rates but also provides information on their covariation that can be used for prediction. Additionally, the generative nature of SBM allows computing the mean probability (together with the associated uncertainty) of positively detecting pharmaceuticals for each" river and pharmaceutical block". In other words, the model provides us the probability (with uncertainty) of detecting unmeasured pharmaceuticals in a river. The detailed process of sub-setting data from the MEC database, its subsequent manipulation for analysis and a complete description of our algorithm are provided in the supplementary material. We provide an illustrative example of our data formatting and its subsequent rearrangement by SBM in Figure 1. Since the algorithm groups rivers as well as pharmaceuticals (see Figure 1), we refer to pharmaceutical groups as 'pharmaceutical clusters' to avoid confusion with river groups.

Similar to the river, we determined the number of times a pharmaceutical was analyzed and positively detected in WWTPs (Influent and effluent). Pharmaceuticals that were measured in WWTP but were not part of our river subset samples were discarded. To explore continental scale differences, we subdivided the WWTP detection rates in three UN groups (Asia, Eastern Europe and Western Europe and others) and summarized them based upon pharmaceutical clusters.

2.5 Social and environmental variables

We explored the effect of environmental and anthropogenic factors (e.g., watershed size, river length, flowrate and population density) on the degree of contamination for the different rivers. We specifically chose these variables as it has been shown that they can play an important role in governing the degree of contamination of the rivers (Acuña et al., 2015; Burns et al., 2018; Kaushal and Belt, 2012; Osorio et al., 2016, 2012a; Peng et al., 2008). We obtained the corresponding information for each river basin from published literature and reports from national agencies. For the few rivers with no published data on population, we estimated basin population by clipping the global population estimates, obtained from the Center for International Earth Science Information Network (Columbia University), with river shape files obtained from HydroSHED (Lehner et al., 2008) and European Environmental agency.

3. Results

Our methodology resulted in 2202 measurements of 112 pharmaceuticals across 64 rivers (Figure S2) with 1324 positive observations resulting in a mean detection rate of 60%. The range of RCI varied between 0 and 1. Except for 1 river with measurements between 30-50 samples (Figure 2), very low RCI values were generally associated with rivers with a lower number of measurements (Figure 2) suggesting that sample size might play a role in governing the RCI. Indeed, for rivers with less than 50 measurements, the range of RCI was large (0 to 1). On the other hand, for rivers, with greater than 50 measurements RCI ranged from 0.3 to 0.85 (Figure 2), revealing that as the number of measurements increases, extreme low RCI values are unlikely and thus every river would exhibit some degree of contamination if pharmaceuticals are measured with adequate intensity. This suggest that the limited monitoring of pharmaceuticals in waterbodies, compared to a more traditional pollutants, may lead to inaccurate conclusions on their presence or absence, and concentrations, and that further, more spatially and temporally intensive, monitoring is needed.

The stochastic block model (SBM) resulted in 6 pharmaceutical clusters and 5 river groups respectively (Figure 3) i.e. 30 (6 multiplied by 5) blocks of rivers and pharmaceuticals. Each block consists of a set of rivers that have similar detection rates for a set of pharmaceuticals. Each block can also be considered as a set of pharmaceuticals that have similar detection rates for a set of rivers. The effectiveness of the model in grouping surface waterbodies as well as pharmaceuticals with similar detection rates is best realized by visually comparing the data before and after clustering (see Figure S3 for the raw un-clustered data). The pharmaceutical clusters and the river groups are arranged in increasing order of the detection rates.

Pharmaceuticals in clusters D to F were positively detected in all the river groups and pharmaceuticals in clusters A and B were mostly undetected in river groups 1 to 3 (Figure 3). We also observe regional differences in the river groups. All but two Asian rivers were assigned to river groups 4 and 5 which exhibited high detection rates, suggesting highest level of contamination in Asian Rivers. European and North American rivers were present in all the groups, however our model also revealed important differences within the European rivers. Only German and Slovenian rivers belonged to river groups 1 and 2, with very low detection rates of cluster A pharmaceuticals (<10%, Figure 3). In contrast, the detection rate of cluster A pharmaceuticals for Italian, Spanish and French rivers (belonging mostly to river groups 3, 4 and 5, Figure 3) were ~35% which, although lower than the detection rate in Asian rivers (>80%), was still higher than the rivers flowing in Germany, Slovakia and Netherlands (<20%). None of the cluster A pharmaceuticals (more than 20 different pharmaceuticals) that were measured multiple times in the River Rhine (flows through Switzerland, Germany and the Netherlands) were positively detected (Figure 3).

Our result suggests that for all the rivers groups, the mean probability of positively detecting the pharmaceuticals in cluster F was high (Figure 4). As a result, pharmaceuticals in cluster F are likely to be positively detected in all of the studied rivers. Similarly, except for rivers in group 1, the mean likelihood of positively detecting clusters D and E pharmaceuticals in unmeasured rivers is greater than 50%. In contrast, the detection rates of clusters A to C pharmaceuticals in river groups 1 and 2 is low (Figure 4).

The estimated 95% credible intervals provide confidence in interpreting the mean detection rate associated with each river and pharmaceutical block. The narrow 95% credible intervals (CIs, ranging mostly from 0.6 to 1) associated with cluster F for all the river groups (Figure 4) suggests high confidence in the likelihood of positively detecting cluster F pharmaceuticals at all the rivers. On the other hand, the 95% CI associated with clusters C and D are large (Figure 4) (due to limited number of measurements) indicating substantial uncertainty associated with these probabilities (Figure 4).

4. Discussion

4.1 Pattern in pharmaceutical detection rates

The high detection rates of 22 pharmaceuticals in clusters D to F (Figure 3) suggests that these pharmaceuticals were present ubiquitously in all the rivers included in this study. Many pharmaceuticals in clusters D to F are among the most widely consumed in the USA, UK and several other countries (Fuentes et al., 2018; Letsinger and Kay, 2019) and have exhibited high detection frequencies in previous global analyses of pharmaceuticals in surface water bodies (Fekadu et al., 2019; Hughes et al., 2013). The pharmaceuticals included by our model in clusters D to F do not belong to a single therapeutic group but come from diverse classes including analgesics, antibiotics, estrogens and beta-blockers (Table S1).

Even though the overall detection rate of pharmaceuticals in clusters A, B and C (Figure 3) was lower, the detection rate for pharmaceuticals in these clusters were not similar across the river groups. Blocks 4-A, 5-A 3-B, 4-B and 5-B had much higher positive detection than blocks 1-A, 2-A, 3-A, 1-B and 2-B (see figure 3). Most of the rivers with high detection rates of pharmaceuticals in clusters A and B were Asian. Among the European rivers, only Italian, French and Spanish exhibited high detection rates. Rivers from other European countries including England, Germany, Netherlands and Slovakia exhibited low detection rates for pharmaceuticals in clusters A and B. Our model output suggests, there are systematic country level differences in the rivers for clusters A and B pharmaceuticals. These differences might be attributable to multiple factors (e.g., pharmaceutical consumption pattern, WWTP removal processes, hydrological and social factors and/or a combination of these factors), that we discuss below.

4.2 Factors governing the regional differences among the rivers

To explore the patterns observed above, we combined the rivers into their official UN regional group resulting in 13, 10 and 41 rivers belonging to Asia, Eastern Europe (EE) and Western Europe and others (WEOG) regional groups, respectively. For the WEOG group, 33 rivers were Western European and 8 were North American. We also combined pharmaceuticals in clusters A to C and D to F respectively in 2 groups as pharmaceuticals in clusters A to C and D to F have similar detection rates. We restrict our discussion to Asian and WEOG groups as the majority of the rivers in the EE group are from a single country (Slovakia, see Figures 3 and S2).

4.2.1 Wastewater treatment plants

In developed countries, WWTP effluent is considered a primary source of pharmaceuticals to aquatic environments (Andreozzi et al., 2003; Letsinger and Kay, 2019; Petrovic et al., 2002) and the degree of contamination of a river is linked to the pharmaceutical removal efficiency of WWTPs. In developing countries, untreated effluent could also be discharged directly due to absence of WWTPs and/or limited connectivity between houses and WWTPs. The removal rate of pharmaceuticals in WWTP varies significantly (Khamis et al., 2011; Verlicchi et al., 2012). Many of the clusters D to F pharmaceuticals such as diclofenac, acetylsalicylic acid, naproxen, and gemfibrozil are in ionic state at neutral pH, and therefore difficult to remove during waste water treatment processes (Khamis et al., 2011). In an extensive review, (Verlicchi et al., 2012) showed that the removal rate of several clusters D to F pharmaceuticals such as carbamazepine, sotalol, sulfamethoxazole, metoprolol, erythromycin and others are as low as 40% even post-secondary treatment. In contrast, many of the pharmaceuticals in clusters A to C including doxycycline, chlortetracycline, estradiol, paroxetine, sulfamethizole etc. have been shown to have higher removal rates (Verlicchi et al., 2012). The median removal rate of clusters A to C and D to F pharmaceuticals complied in Verlicchi et al (2012) is 62% and 48% respectively (see Figure S4). It is therefore possible that the patterns observed in our pharmaceutical clusters are related to their removal efficiency by WWTP. Since WWTP are more extensive and up to date in WEOG (includes secondary and tertiary treatment processes), we hypothesized that the differences in the detection rate for cluster A to C pharmaceuticals between Asian and WEOG rivers could be due to more efficient removal of clusters A to C pharmaceuticals in WEOG.

For Asia as well as WEOG groups, the detection rates of pharmaceuticals in clusters D to F were high for both WWTP influent and effluent, with little difference between Asian and WEOG effluents (Figure 5c and 5d). This was not surprising as clusters D to F pharmaceuticals are difficult to remove using conventional WWTP processes(Verlicchi et al., 2012). As expected, for pharmaceuticals in clusters A to C, the median detection rates in WWTP effluent were lower than the influent detection rates for both Asia and WEOG groups (Figures 5a and 5b) suggesting that WWTP processes are more successful in removing these pharmaceuticals than D to F pharmaceuticals. However, the decrease in the detection rate from influent to effluent were not statistically different (t-test, $p>0.05$) for Asian and WEOG WWTP effluents. Therefore, our first order comparative analysis does not provide any compelling indication that there are systematic differences between the WWTPs in Asia and WEOG, or that WEOG WWTPs are removing pharmaceuticals more effectively compared to the Asian WWTPs. It is possible that WEOG WWTPs are better at lowering the concentration; however, our analysis suggests that even in that case, the concertation are high enough for the pharmaceutical to be detected in

WWTP effluents. A meta-analysis of pharmaceutical concertation in WWTP influent and effluent across the different countries can provide more detailed insight into these differences.

We observe a substantial decrease in the detection rates of cluster A to C pharmaceuticals between WWTP effluents and downstream river sites for WEOG (Figure 5a) but not for Asia (Figure 5b). The higher detection rates in rivers compared to the WWTP effluent for Asia suggests additional input through combined sewer overflows and/or direct discharge of untreated sewage water to the rivers. Indeed, the degree of connectivity of households to WWTP in Asia are significantly lower compared to the WEOG and the observed pattern is not surprising and highlights the need of reducing discharge of untreated wastewater in rivers and other surface waterbodies in Asia (Isobe et al., 2004; Shrestha and Pandey, 2016; Thomes et al., 2019).

It would have been interesting to divide European WWTPs in two subgroups that included Germany, Netherlands, Austria, Switzerland, Belgium and England in one group and France, Italy, Spain, Portugal and Greece in another, as the countries in latter group had less than 40% of the population served by WWTP with tertiary treatment process before 2005 (https://www.eea.europa.eu/data-and-maps/indicators/urban-waste-water-treatment/urban-waste-water-treatment-assessment-4) whereas more than 80% of the population in Germany, Netherlands, Austria, Switzerland, Belgium and England were served by WWTPs with tertiary treatment processes by 2005. However, due to limited WWTP samples, we did not further subdivide WEOG WWTPs data in subgroups. Given the fact that most European WWTPs have upgraded to tertiary treatment in recent years, and there have been large number of studies in recent years an analysis comparing detection rates in WWTP pre and post 2010 in Europe can help to understand and document the effectiveness of the advanced techniques in removing pharmaceuticals and perhaps explain the differences in degree of contamination of European rivers.

4.2.2 Regional variation in pharmaceutical consumption

The majority of the pharmaceuticals in clusters A to C are antibiotics (48 out of 85, Table S2) and their consumption varies significantly across the globe. Indeed, antibiotics are used less often and are generally more difficult to obtain without prescription in WEOG whereas their consumption in Asia is widespread and they are easily available and often unregulated (Komori et al., 2013; Shimizu et al., 2013). Between 2000 and 2010, global antibiotic consumption increased by 35%, fueled dominantly by Asian countries (Van Boeckel et al., 2014) with India and China being the largest consumers. In comparison, the consumption of antibiotics was not only lower in European countries, but also declined ("Antimicrobial consumption - Annual Epidemiological Report for 2017,"; Van Boeckel et al., 2014).

As mentioned previously, the majority of the rivers in groups 1 and 2 were German and Slovenian, whereas rivers in France, Italy and Spain belonged to groups 3 to 5. According to the latest OCED (Organization for Economic Co-operation and Development) report (2017), Italy and France are among the highest consumers of antibiotics in Europe. The defined daily dose (DDD) of antibiotics in Italy and France are approximately three times higher than Netherlands and twice that of Germany and Slovenia. For this reason, we believe that the pattern observed for pharmaceuticals in clusters A to C with much higher detection rate in Asia and some

European countries in part reflect the regional and country level variation in consumption of these pharmaceuticals.

4.2.3 Effects of hydrologic and socio-environmental factors

The differences observed in the detection rates among the rivers could also be due to local hydrological factors. The presence of pharmaceuticals will vary in rivers due to the prevailing hydrological conditions at the time of sampling. For instance, high river flows may dilute pharmaceutical residues emanating from wastewater treatment plants. Conversely, untreated effluent could be released from combined sewer overflows during storm events. Unfortunately, these hydrological characteristics are seldom described in published reports and scientific articles. Although pharmaceutical measurements in rivers are traditionally taken during low flow summer conditions close to the WWTP effluent outlet, many pharmaceutical datasets comprise a small number of samples taken with no consideration of flow conditions. As a result, our study, which focuses on general trends at large spatial scales based on a meta-analysis, unfortunately cannot account for how flow conditions may have affected the presence of pharmaceuticals in rivers. Nevertheless, it is important to note that it would be unlikely that high flow events would have discriminately diluted pharmaceuticals in clusters A to C in WEOG to an extent that they were not detected with a similar dilution effect missing for pharmaceuticals in clusters D to F.

Keeping in mind the limitations of data available and the lack of detailed information associated with sampling events, we analyzed the relationship between basin size, river length and mean flow rates and river contamination index (RCI) of the river. These hydrologic metrics were available (or obtained) for most of the basins, however our analysis did not result in any statistically meaningful relationship between RCI and these metrics. Indeed, many of the rivers in group 1 (most contaminated) and group 5 (least contaminated) were rivers with comparable mean flow rate and size. Whereas hydrology is a critical factor in determining the degree of contamination of a river as highlighted by several studies (Kay et al., 2017; Keller et al., 2014; Kolpin et al., 2004), the lack of relationship between mean flow rate and the detection rates highlights the complexity of interaction between hydrology and pharmaceuticals in water and the inability of seasonally and basin averaged mean flow values to capture this relationship. Our analysis highlights the need for long-term catchment scale spatiotemporal studies to understand these relationships.

We observe an increasing trend in RCI with increasing population density within the basin (Figure 6) albeit with significant variability. Most of the pharmaceuticals analyzed in our study were used primarily for human consumption and the positive trend between population density and pharmaceutical detection was expected. The effect of population on the degree of contamination was appropriately highlighted for the rivers Ebro, LLobregat and Ter. These rivers are comparable in size, situated within the Iberian Peninsula, Spain (thus experiencing similar climatic regime and country level pharmaceutical policies) and have more than 30 unique measurements on each river. In our analysis, the detection of pharmaceuticals was much lower for the River Ter (RCI = 0.25) compared to the Llobregat (RCI = 0.78) and Ebro (RCI = 0.80) which might be due to the lower population density of the River Ter (Céspedes et al., 2006). A recently conducted independent study (Osorio et al., 2016) within the same region comparing four rivers (Llogregat, Ebro, Jucar and Guadalquivir) also highlighted the positive correlation between human population and pharmaceutical concentration in these rivers and showed that the degree of contamination of the LLobregat and Ebro were higher than Jucar and Guadalquivir, most likely due to their higher population density (Osorio et al., 2016).

The presence of substantial scatter around the relationship between RCI and population density in our analysis could be due to multiple factors that can vary on basin, local and national scales including access to pharmaceuticals, pharmaceutical consumption habits, mean age of the population, seasonal variability, per capita domestic water consumption, and sampling strategies (Murata et al., 2011; Osorio et al., 2012b). Our result highlights the relationship between contamination and population and the growing need to quantify the presence of pharmaceuticals in densely populated areas especially in developing countries where public health and aquatic ecosystems might be acutely affected due to elevated presence of several pharmaceuticals.

4.4 A novel approach for selecting pharmaceuticals to be studied in rivers

Currently, more than 3000 pharmaceuticals are being used globally (Donnachie et al., 2016) and the list is growing. Given that our understanding of the eco-toxicological effects of most pharmaceuticals in surface water is not fully developed(Fent et al., 2006) , it is important to determine their environmental concentration. However, monitoring or modelling concentration of pharmaceuticals in surface water is challenging due to limited resources, time and costs associated with these studies. Most monitoring efforts have been limited to fewer than 10 pharmaceuticals per study (Gros et al., 2006). To circumvent these challenges, researchers have complemented field measurements with estimated concentrations in surface water using pharmaceutical sales and wastewater production rates and have developed ranking schemes to prioritize pharmaceuticals for analysis in a given location (Al-Khazrajy and Boxall, 2016; Berninger et al., 2016; Bu et al., 2020; De Voogt et al., 2009; Fick et al., 2010; Huggett et al., 2003; Kostich and Lazorchak, 2008; Kumar and Xagoraraki, 2010; Sui et al., 2012), the output of such models varies substantially (Roos et al., 2012) limiting their utility for analytical prioritization purposes.

The SBM enables the identification of pharmaceuticals with similar occurrence patterns in surface water. For example, in our dataset for all the rivers where both diclofenac and carbamazepine were measured, they were positively detected 90% of the time. Similar patterns were also observed for pharmaceuticals that were not detected when measured concurrently. Our model provides a probabilistic estimate of positively detecting unstudied pharmaceuticals in rivers (Figure 4), which can complement existing mechanistic/process-based models such as those proposed by (Huggett et al., 2003; Kumar and Xagoraraki, 2010; Roos et al., 2012) to choose the pharmaceuticals needed to be included in a study. For example, if diclofenac is positively detected in a river, it might not be useful to measure carbamazepine in the same river as it is very likely to be positively detected. A cross-validation exercise (results not shown) suggests that, by grouping pharmaceuticals with similar co-occurrence pattern in rivers, we can make reasonable predictions on the presence/absence of all the pharmaceuticals within a group by performing field measurement of few ‘selected’ pharmaceuticals, a very useful feature given the high costs associated with measuring concentration of these pharmaceuticals. As an example, we provide estimates of the probability of detecting few selected pharmaceuticals that were not studied in River Colorado, Elbe and Rhine (Table 1).Indeed

Diclofenac was positively detected in all the water ways of Elbe River catchment (Marsik et al., 2017; Meyer et al., 2016). Although this example is for illustrative purposes, the goal is to highlight the applicability of statistical analyses of big pharmaceutical datasets in providing useful information on environmental pharmaceutical contamination. We encourage researchers to validate the robustness and accuracy of the method by comparing the pharmaceutical detection rates to our model output. We hope that this paper would motivate users to use our method or develop newer statistical methods that can be applied to the emerging field of environmental pharmaceutical contamination. Our analysis has highlighted and confirmed some of the patters (effect of population, consumption patterns) that have been suggested before but never explored globally.

As the number of studies measuring pharmaceuticals in environmental matrix using standardized protocol, composite sampling and concurrent measurements of WWTP and receiving water is increasing rapidly, for example see (Challis et al., 2018; Cui et al., 2019; Grill et al., 2016; Kay et al., 2017), in future we plan to perform similar analysis using concentration rather than presence/absence data yielding results that are more useful from eco-toxological and policy point of view. Such analysis will be especially appropriate for comparing river basins within a country as in-country variation in pharmaceutical consumption behavior and WWTP efficiency is likely to be smaller than between country variation. We believe that combining process-based rankings with results from sophisticated statistical model would maximize the information that can be obtained on the toxicity of pharmaceuticals in different environmental matrices and could help in developing sustainable strategies to minimize the effects of pharmaceuticals on aquatic ecosystems.

5. Conclusions

Previous works have suggested the presence of numerous pharmaceuticals from a wide spectrum of therapeutic classes in environmental waters (Beek et al., 2016; Daughton, 2001; Hughes et al., 2013; Loos et al., 2010). However, to our knowledge, none of them except (Loos et al., 2010) have conducted a systematic assessment of the detection rate of pharmaceuticals across multiple rivers. Our meta-analysis highlights the differences in the detection rate of 112 pharmaceuticals and their variation across Asia, Europe and North America. We identify some of the possible factors including consumption rate, local hydrology and population that could be driving this pattern. Whereas we could detect a first order relationship between pharmaceutical detection rates and pharmaceutical use, the effect of hydrological factors could not be resolved in this analysis. Importantly, our approach informs the probability of detecting unanalyzed pharmaceuticals and supports analyte prioritization for future.

Many of our findings have been suggested before, however here we show these empirically using a large dataset analyzed within a statistical framework. Future analysis could leverage much larger datasets and more sophisticated statistical techniques to acquire more detailed and improved information on pharmaceutical contamination in surface water.

6. Supporting Information

Data sub-setting and aggregation; description of the stochastic block model; model implementation; full conditional distributions; tables of pharmaceutical clusters; data matrices

7. Acknowledgements

This work was funded by NSF Grant 1458034. We thank the scientists and agencies who provided the data to the MEC database. We would like to thank Tim A D Beek for discussion on the manuscript. The authors have no conflicts of interest in relation to publication of this manuscript. We thank seven anonymous reviewers for their insightful comments and suggestions.

7. References

Acuña, V., von Schiller, D., García-Galán, M.J., Rodríguez-Mozaz, S., Corominas, L., Petrovic, M., Poch, M., Barceló, D., Sabater, S., 2015. Occurrence and in-stream attenuation of wastewater-derived pharmaceuticals in Iberian rivers. Sci. Total Environ. 503–504, 133–141. https://doi.org/10.1016/j.scitotenv.2014.05.067

Al-Khazrajy, O.S.A., Boxall, A.B.A., 2016. Risk-based prioritization of pharmaceuticals in the natural environment in Iraq. Environ. Sci. Pollut. Res. 23, 15712–15726. https://doi.org/10.1007/s11356-016-6679-0

Altenburger, R., Nendza, M., Schüürmann, G., 2003. Mixture toxicity and its modeling by quantitative structure-activity relationships. Environ. Toxicol. Chem. https://doi.org/10.1897/01-386

Amiard-Triquet, C., Amiard, J.C., Mouneyrac, C., 2015. Predictive Ecotoxicology and Environmental Assessment, in: Aquatic Ecotoxicology: Advancing Tools for Dealing with Emerging Risks. Elsevier Inc., pp. 463–496. https://doi.org/10.1016/B978-0-12-800949-9.00019-X

Andreozzi, R., Marotta, R., Paxéus, N., 2003. Pharmaceuticals in STP effluents and their solar photodegradation in aquatic environment. Chemosphere 50, 1319–1330. https://doi.org/10.1016/S0045-6535(02)00769-5

Andrews, J.M., 2001. Determination of minimum inhibitory concentrations. J. Antimicrob. Chemother. 48, 5–16. https://doi.org/10.1093/jac/48.suppl_1.5

Antimicrobial consumption - Annual Epidemiological Report for 2017 [WWW Document], n.d. URL https://www.ecdc.europa.eu/en/publications-data/antimicrobial-consumption-annual-epidemiological-report-2017 (accessed 12.26.19).

aus der Beek, T., Weber, F.-A., Bergmann, A., Hickmann, S., Ebert, I., Hein, A., Küster, A., 2016. Pharmaceuticals in the environment-Global occurrences and perspectives. Environ. Toxicol. Chem. 35, 823–835. https://doi.org/10.1002/etc.3339

Banks, D.L., Hengartner, N., 2008. Social Networks, in: Encyclopedia of Quantitative Risk Analysis and Assessment. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/9780470061596.risk0667

Barnes, K.K., Kolpin, D.W., Furlong, E.T., Zaugg, S.D., Meyer, M.T., Barber, L.B., 2008. A national reconnaissance of pharmaceuticals and other organic wastewater contaminants in the United States - I) Groundwater. Sci. Total Environ. 402, 192–200. https://doi.org/10.1016/j.scitotenv.2008.04.028

Berkhin, P., 2006. A survey of clustering data mining techniques, in: Grouping Multidimensional Data: Recent Advances in Clustering. Springer Berlin Heidelberg, pp. 25–71. https://doi.org/10.1007/3-540-28349-8_2

Berninger, J.P., Lalone, C.A., Villeneuve, D.L., Ankley, G.T., 2016. Prioritization of pharmaceuticals for potential environmental hazard through leveraging a large-scale mammalian pharmacological dataset. Environ. Toxicol. Chem. 35, 1007–1020. https://doi.org/10.1002/etc.2965

Botitsi, E., Frosyni, C., Tsipi, D., 2007. Determination of pharmaceuticals from different therapeutic classes in wastewaters by liquid chromatography-electrospray ionization-tandem mass spectrometry, in: Analytical and Bioanalytical Chemistry. pp. 1317–1327. https://doi.org/10.1007/s00216-006-0804-8

Brain, R.A., Hanson, M.L., Solomon, K.R., Brooks, B.W., 2008. Aquatic plants exposed to pharmaceuticals: Effects and risks. Rev. Environ. Contam. Toxicol. 192, 67–115. https://doi.org/10.1007/978-0-387-71724-1_3

Bu, Q., Cao, Y., Yu, G., He, X., Zhang, H., Sun, J., Yun, M., Cao, Z., 2020. Identifying targets of potential concern by a screening level ecological risk assessment of human use pharmaceuticals in China. Chemosphere 246. https://doi.org/10.1016/j.chemosphere.2020.125818

Burns, E.E., Carter, L.J., Kolpin, D.W., Thomas-Oates, J., Boxall, A.B.A., 2018. Temporal and spatial variation in pharmaceutical concentrations in an urban river system. Water Res. 137, 72–85.

https://doi.org/10.1016/j.watres.2018.02.066

Céspedes, R., Lacorte, S., Ginebreda, A., Barceló, D., 2006. Chemical monitoring and occurrence of alkylphenols, alkylphenol ethoxylates, alcohol ethoxylates, phthalates and benzothiazoles in sewage treatment plants and receiving waters along the ter River basin (Catalonia, N. E. Spain), in: Analytical and Bioanalytical Chemistry. pp. 992–1000. https://doi.org/10.1007/s00216-006-0448-8

Challis, J.K., Cuscito, L.D., Joudan, S., Luong, K.H., Knapp, C.W., Hanson, M.L., Wong, C.S., 2018. Inputs, source apportionment, and transboundary transport of pesticides and other polar organic contaminants along the lower Red River, Manitoba, Canada. Sci. Total Environ. 635, 803–816. https://doi.org/10.1016/j.scitotenv.2018.04.128

Chi, E.C., Allen, G.I., Baraniuk, R.G., 2017. Convex biclustering. Biometrics 73, 10–19. https://doi.org/10.1111/biom.12540

Cui, Y., Wang, Y., Pan, C., Li, R., Xue, R., Guo, J., Zhang, R., 2019. Spatiotemporal distributions, source apportionment and potential risks of 15 pharmaceuticals and personal care products (PPCPs) in Qinzhou Bay, South China. Mar. Pollut. Bull. 141, 104–111. https://doi.org/10.1016/j.marpolbul.2019.02.012

Daughton, C.G., 2003. Cradle-to-cradle stewardship of drugs for minimizing their environmental disposition while promoting human health. I. Rational for and avenues toward a green pharmacy. Environ. Health Perspect. 111, 757–774. https://doi.org/10.1289/ehp.5947

Daughton, C.G., 2001. Emerging pollutants, and communicating the science of environmental chemistry and mass spectrometry: Pharmaceuticals in the environment. J. Am. Soc. Mass Spectrom. 12, 1067–1076. https://doi.org/10.1016/S1044-0305(01)00287-2

de Jesus Gaffney, V., Almeida, C.M.M., Rodrigues, A., Ferreira, E., Benoliel, M.J., Cardoso, V.V., 2015. Occurrence of pharmaceuticals in a water supply system and related human health risk assessment. Water Res. 72, 199–208. https://doi.org/10.1016/j.watres.2014.10.027

De Voogt, P., Janex-Habibi, M.L., Sacher, F., Puijker, L., Mons, M., 2009. Development of a common priority list of pharmaceuticals relevant for the water cycle. Water Sci. Technol. 59, 39–46. https://doi.org/10.2166/wst.2009.764

Dhillon, I.S., 2001. Co-clustering documents and words using Bipartite Spectral Graph Partitioning.

Donnachie, R.L., Johnson, A.C., Sumpter, J.P., 2016. A rational approach to selecting and ranking some pharmaceuticals of concern for the aquatic environment and their relative importance compared with other chemicals. Environ. Toxicol. Chem. 35, 1021–1027. https://doi.org/10.1002/etc.3165

Fekadu, S., Alemayehu, E., Dewil, R., Van der Bruggen, B., 2019. Pharmaceuticals in freshwater aquatic environments: A comparison of the African and European challenge. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2018.11.072

Fent, K., Weston, A.A., Caminada, D., 2006. Ecotoxicology of human pharmaceuticals. Aquat. Toxicol. https://doi.org/10.1016/j.aquatox.2005.09.009

Fick, J., Lindberg, R.H., Tysklind, M., Larsson, D.G.J., 2010. Predicted critical environmental concentrations for 500 pharmaceuticals. Regul. Toxicol. Pharmacol. 58, 516–523. https://doi.org/10.1016/j.yrtph.2010.08.025

Fuentes, A., Pineda, M., Venkata, K., 2018. Comprehension of Top 200 Prescribed Drugs in the US as a Resource for Pharmacy Teaching, Training and Practice. Pharmacy 6, 43. https://doi.org/10.3390/pharmacy6020043

Govaert, G., n.d. Simultaneous clustering of rows and columns CIRCE-SIGFRIED View project PLAINE View project.

Grill, G., Khan, U., Lehner, B., Nicell, J., Ariwi, J., 2016. Risk assessment of down-the-drain chemicals at large spatial scales: Model development and application to contaminants originating from urban areas in the Saint

Lawrence River Basin. Sci. Total Environ. 541, 825–838. https://doi.org/10.1016/j.scitotenv.2015.09.100

Gros, M., Petrović, M., Barceló, D., 2006. Multi-residue analytical methods using LC-tandem MS for the determination of pharmaceuticals in environmental and wastewater samples: A review. Anal. Bioanal. Chem. https://doi.org/10.1007/s00216-006-0586-z

Grossberger, A., Hadar, Y., Borch, T., Chefetz, B., 2014. Biodegradability of pharmaceutical compounds in agricultural soils irrigated with treated wastewater. Environ. Pollut. 185, 168–177. https://doi.org/10.1016/j.envpol.2013.10.038

Hartigan, J.A., 1972. Direct clustering of a data matrix. J. Am. Stat. Assoc. 67, 123–129. https://doi.org/10.1080/01621459.1972.10481214

Hill, M.O., Harrower, C.A., Preston, C.D., 2013. Spherical k-means clustering is good for interpreting multivariate species occurrence data. Methods Ecol. Evol. 4, 542–551. https://doi.org/10.1111/2041-210X.12038

Hoff, P.D., Raftery, A.E., Handcock, M.S., 2002. Latent space approaches to social network analysis. J. Am. Stat. Assoc. 97, 1090–1098. https://doi.org/10.1198/016214502388618906

Huggett, D.B., Cook, J.C., Ericson, J.F., Williams, R.T., 2003. A theoretical model for utilizing mammalian pharmacology and safety data to prioritize potential impacts of human pharmaceuticals to fish. Hum. Ecol. Risk Assess. https://doi.org/10.1080/714044797

Hughes, S.R., Kay, P., Brown, L.E., 2013. Global Synthesis and Critical Evaluation of Pharmaceutical Data Sets Collected from River Systems. Environ. Sci. Technol. 47, 661–677. https://doi.org/10.1021/es3030148

Isobe, K.O., Zakaria, M.P., Chiem, N.H., Minh, L.Y., Prudente, M., Boonyatumanond, R., Saha, M., Sarkar, S., Takada, H., 2004. Distribution of linear alkylbenzenes (LABs) in riverine and coastal environments in South and Southeast Asia. Water Res. 38, 2449–2459. https://doi.org/10.1016/j.watres.2004.02.009

Jiang, J.Q., Zhou, Z., Sharma, V.K., 2013. Occurrence, transportation, monitoring and treatment of emerging micro-pollutants in waste water - A review from global views. Microchem. J. https://doi.org/10.1016/j.microc.2013.04.014

Johnson, A.C., Dumont, E., Williams, R.J., Oldenkamp, R., Cisowska, I., Sumpter, J.P., 2013. Do concentrations of ethinylestradiol, estradiol, and diclofenac in European rivers exceed proposed EU environmental quality standards? Environ. Sci. Technol. 47, 12297–12304. https://doi.org/10.1021/es4030035

Jones-Lepp, T.L., Sanchez, C., Alvarez, D.A., Wilson, D.C., Taniguchi-Fu, R.L., 2012. Point sources of emerging contaminants along the Colorado River Basin: Source water for the arid Southwestern United States. Sci. Total Environ. 430, 237–245. https://doi.org/10.1016/j.scitotenv.2012.04.053

Jones, O.A.H., Voulvoulis, N., Lester, J.N., 2002. Aquatic environmental assessment of the top 25 English prescription pharmaceuticals. Water Res. 36, 5013–5022. https://doi.org/10.1016/S0043-1354(02)00227-0

Jones, O.A.H., Voulvoulis, N., Lester, J.N., 2001. Human pharmaceuticals in the aquatic environment a review. Environ. Technol. (United Kingdom) 22, 1383–1394. https://doi.org/10.1080/09593330.2001.11090873

Kaplan, S., 2013. Review: Pharmacological pollution in water. Crit. Rev. Environ. Sci. Technol. https://doi.org/10.1080/10934529.2011.627036

Kaushal, S.S., Belt, K.T., 2012. The urban watershed continuum: Evolving spatial and temporal dimensions. Urban Ecosyst. 15, 409–435. https://doi.org/10.1007/s11252-012-0226-7

Kay, P., Hughes, S.R., Ault, J.R., Ashcroft, A.E., Brown, L.E., 2017. Widespread, routine occurrence of pharmaceuticals in sewage effluent, combined sewer overflows and receiving waters. Environ. Pollut. 220, 1447–1455. https://doi.org/10.1016/j.envpol.2016.10.087

Kehrein, N., Berlekamp, J., Klasmeier, J., 2015. Modeling the fate of down-the-drain chemicals in whole watersheds: New version of the GREAT-ER software. Environ. Model. Softw. 64, 1–8. https://doi.org/10.1016/j.envsoft.2014.10.018

Keller, V.D.J., Williams, R.J., Lofthouse, C., Johnson, A.C., 2014. Worldwide estimation of river concentrations of any chemical originating from sewage-treatment plants using dilution factors. Environ. Toxicol. Chem. 33, 447–452. https://doi.org/10.1002/etc.2441

Khamis, M., Karaman, R., Ayyash, F., Qtait, A., Deeb, O., Manssra, A., 2011. Efficiency of Advanced Membrane Wastewater Treatment Plant towards Removal of Aspirin, Salicylic Acid, Paracetamol and p-Aminophenol, Journal of Environmental Science and Engineering.

Klečka, G.M., Staples, C.A., Clark, K.E., Van Der Hoeven, N., Thomas, D.E., Hentges, S.G., 2009. Exposure analysis of bisphenol A in surface water systems in North America and Europe. Environ. Sci. Technol. 43, 6145–6150. https://doi.org/10.1021/es900598e

Kolpin, D.W., Skopec, M., Meyer, M.T., Furlong, E.T., Zaugg, S.D., 2004. Urban contribution of pharmaceuticals and other organic wastewater contaminants to streams during differing flow conditions. Sci. Total Environ. 328, 119–130. https://doi.org/10.1016/j.scitotenv.2004.01.015

Komori, K., Suzuki, Y., Minamiyama, M., Harada, A., 2013. Occurrence of selected pharmaceuticals in river water in Japan and assessment of their environmental risk. Environ. Monit. Assess. 185, 4529–4536. https://doi.org/10.1007/s10661-012-2886-4

Kostich, M.S., Lazorchak, J.M., 2008. Risks to aquatic organisms posed by human pharmaceutical use. Sci. Total Environ. 389, 329–339. https://doi.org/10.1016/j.scitotenv.2007.09.008

Kumar, A., Xagoraraki, I., 2010. Pharmaceuticals, personal care products and endocrine-disrupting chemicals in U.S. surface and finished drinking waters: A proposed ranking system. Sci. Total Environ. 408, 5972–5989. https://doi.org/10.1016/j.scitotenv.2010.08.048

Kümmerer, K., 2009a. Antibiotics in the aquatic environment - A review - Part II. Chemosphere. https://doi.org/10.1016/j.chemosphere.2008.12.006

Kümmerer, K., 2009b. Antibiotics in the aquatic environment - A review - Part I. Chemosphere. https://doi.org/10.1016/j.chemosphere.2008.11.086

Lehner, B., Verdin, K., Jarvis, A., 2008. New Global Hydrography Derived From Spaceborne Elevation Data. Eos, Trans. Am. Geophys. Union 89, 93. https://doi.org/10.1029/2008EO100001

Letsinger, S., Kay, P., 2019. Comparison of Prioritisation Schemes for Human Pharmaceuticals in the Aquatic Environment. Environ. Sci. Pollut. Res. 26, 3479–3491. https://doi.org/10.1007/s11356-018-3834-9

Loos, R., Locoro, G., Comero, S., Contini, S., Schwesig, D., Werres, F., Balsaa, P., Gans, O., Weiss, S., Blaha, L., Bolchi, M., Gawlik, B.M., 2010. Pan-European survey on the occurrence of selected polar organic persistent pollutants in ground water. Water Res. 44, 4115–4126. https://doi.org/10.1016/j.watres.2010.05.032

Marsik, P., Rezek, J., Židková, M., Kramulová, B., Tauchen, J., Vaněk, T., 2017. Non-steroidal anti-inflammatory drugs in the watercourses of Elbe basin in Czech Republic. Chemosphere 171, 97–105. https://doi.org/10.1016/j.chemosphere.2016.12.055

Meyer, W., Reich, M., Beier, S., Behrendt, J., Gulyas, H., Otterpohl, R., n.d. Measured and predicted environmental concentrations of carbamazepine, diclofenac, and metoprolol in small and medium rivers in northern Germany. https://doi.org/10.1007/s10661-016-5481-2

Murata, A., Takada, H., Mutoh, K., Hosoda, H., Harada, A., Nakada, N., 2011. Nationwide monitoring of selected antibiotics: Distribution and sources of sulfonamides, trimethoprim, and macrolides in Japanese rivers. Sci. Total Environ. 409, 5305–5312. https://doi.org/10.1016/j.scitotenv.2011.09.014

Oaks, J.L., Gilbert, M., Virani, M.Z., Watson, R.T., Meteyer, C.U., Rideout, B.A., Shivaprasad, H.L., Ahmed, S., Chaudhry, M.J.I., Arshad, M., Mahmood, S., Ali, A., Khan, A.A., 2004. Diclofenac residues as the cause of vulture population decline in Pakistan. Nature 427, 630–633. https://doi.org/10.1038/nature02317

Ort, C., Lawrence, M.G., Rieckermann, J., Joss, A., 2010. Sampling for Pharmaceuticals and Personal Care Products (PPCPs) and Illicit Drugs in Wastewater Systems: Are Your Conclusions Valid? A Critical Review. Environ. Sci. Technol. 44, 6024–6035. https://doi.org/10.1021/es100779n

Osorio, V., Larrañaga, A., Aceña, J., Pérez, S., Barceló, D., 2016. Concentration and risk of pharmaceuticals in freshwater systems are related to the population density and the livestock units in Iberian Rivers. Sci. Total Environ. 540, 267–277. https://doi.org/10.1016/j.scitotenv.2015.06.143

Osorio, V., Marcé, R., Pérez, S., Ginebreda, A., Cortina, J.L., Barceló, D., 2012a. Occurrence and modeling of pharmaceuticals on a sewage-impacted Mediterranean river and their dynamics under different hydrological conditions. Sci. Total Environ. 440, 3–13. https://doi.org/10.1016/j.scitotenv.2012.08.040

Osorio, V., Pérez, S., Ginebreda, A., Barceló, D., 2012b. Pharmaceuticals on a sewage impacted section of a Mediterranean River (Llobregat River, NE Spain) and their relationship with hydrological conditions. Environ. Sci. Pollut. Res. 19, 1013–1025. https://doi.org/10.1007/s11356-011-0603-4

Peake, B.M., Braund, R., Tong, A.Y.C., Tremblay, L.A., 2015. The Life-Cycle of Pharmaceuticals in the Environment, The Life-Cycle of Pharmaceuticals in the Environment. Elsevier Inc. https://doi.org/10.1016/C2013-0-18158-5

Peng, X., Yu, Y., Tang, C., Tan, J., Huang, Q., Wang, Z., 2008. Occurrence of steroid estrogens, endocrine-disrupting phenols, and acid pharmaceutical residues in urban riverine water of the Pearl River Delta, South China. Sci. Total Environ. 397, 158–166. https://doi.org/10.1016/j.scitotenv.2008.02.059

Petrovic, M., Solé, M., López De Alda, M.J., Barceló, D., 2002. Endocrine disruptors in sewage treatment plants, receiving river waters, and sediments: Integration of chemical analysis and biological effects on feral carp. Environ. Toxicol. Chem. 21, 2146–2156. https://doi.org/10.1002/etc.5620211018

Rehman, M.S.U., Rashid, N., Ashfaq, M., Saif, A., Ahmad, N., Han, J.I., 2015. Global risk of pharmaceutical contamination from highly populated developing countries. Chemosphere. https://doi.org/10.1016/j.chemosphere.2013.02.036

Roos, V., Gunnarsson, L., Fick, J., Larsson, D.G.J., Rudén, C., 2012. Prioritising pharmaceuticals for environmental risk assessment: Towards adequate and feasible first-tier selection. Sci. Total Environ. 421–422, 102–110. https://doi.org/10.1016/j.scitotenv.2012.01.039

Santos, J.L., Aparicio, I., Alonso, E., 2007. Occurrence and risk assessment of pharmaceutically active compounds in wastewater treatment plants. A case study: Seville city (Spain). Environ. Int. 33, 596–601. https://doi.org/10.1016/j.envint.2006.09.014

Schwarzenbach, R.P., Escher, B.I., Fenner, K., Hofstetter, T.B., Johnson, C.A., Von Gunten, U., Wehrli, B., 2006. The challenge of micropollutants in aquatic systems. Science (80-.). https://doi.org/10.1126/science.1127291

Shimizu, A., Takada, H., Koike, T., Takeshita, A., Saha, M., Rinawati, Nakada, N., Murata, A., Suzuki, T., Suzuki, S., Chiem, N.H., Tuyen, B.C., Viet, P.H., Siringan, M.A., Kwan, C., Zakaria, M.P., Reungsang, A., 2013. Ubiquitous occurrence of sulfonamides in tropical Asian waters. Sci. Total Environ. 452–453, 108–115. https://doi.org/10.1016/j.scitotenv.2013.02.027

Shrestha, S., Pandey, V.P., 2016. Groundwater as an Environmental Issue in Asian Cities, in: Groundwater Environment in Asian Cities: Concepts, Methods and Case Studies. Elsevier Inc., pp. 1–13. https://doi.org/10.1016/B978-0-12-803166-7.00001-5

Sui, Q., Wang, B., Zhao, W., Huang, J., Yu, G., Deng, S., Qiu, Z., Lu, S., 2012. Identification of priority pharmaceuticals in the water environment of China. Chemosphere 89, 280–286.

https://doi.org/10.1016/j.chemosphere.2012.04.037

Tanay, A., Sharan Ý Ron Shamir, R., 2004. Biclustering Algorithms: A Survey.

Thomes, M.W., Vaezzadeh, V., Zakaria, M.P., Bong, C.W., 2019. Use of sterols and linear alkylbenzenes as molecular markers of sewage pollution in Southeast Asia. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-019-05936-y

Van Boeckel, T.P., Gandra, S., Ashok, A., Caudron, Q., Grenfell, B.T., Levin, S.A., Laxminarayan, R., 2014. Global antibiotic consumption 2000 to 2010: An analysis of national pharmaceutical sales data. Lancet Infect. Dis. 14, 742–750. https://doi.org/10.1016/S1473-3099(14)70780-7

Verlicchi, P., Al Aukidy, M., Zambello, E., 2012. Occurrence of pharmaceutical compounds in urban wastewater: Removal, mass load and environmental risk after a secondary treatment-A review. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2012.04.028

Walters, E., McClellan, K., Halden, R.U., 2010. Occurrence and loss over three years of 72 pharmaceuticals and personal care products from biosolids-soil mixtures in outdoor mesocosms. Water Res. 44, 6011–6020. https://doi.org/10.1016/j.watres.2010.07.051

Webb, S., Ternes, T., Gibert, M., Olejniczak, K., 2003. Indirect human exposure to pharmaceuticals via drinking water. Toxicol. Lett. 142, 157–167. https://doi.org/10.1016/S0378-4274(03)00071-7

(a)

	Phrama 1	Phrama 2	Phrama 3	Phrama 4	Phrama 5	Phrama 6	Phrama 7	Phrama 8	Phrama 9	Phrama 10	Phrama 11	Phrama 12	Phrama 13	Phrama 14
River 1	0.18	0.45	0.5	0.25	0.75	0.8	0.48		0.2	0.65	0.75			
River2	0.25	0.4		0.22	0.6	0.55	0.33	0.9	0.3	0.48			0.35	0.85
River 3	0		0.1	0.3	0.6		0.15	0.75	0	0.53		0.2		0.85
River 4	0.3				0.5	0.45	0.59	0.88	0.28	0.8	0.6	0.25	0.49	0.95
River 5		0.33	0.58	0.26		0.4	0.67	0.8	0.21	0.75	0.45	0.32	0.48	0.8
River 6	0	0.45	0.3		0.35	0.55		0.8	0	0.31		0.22		1
River 7	0.32	0.29	0.32	0.23		0.45	0.52	0.85			0.54	0.21		
River 8		0.1	0.05			0.6	0		0.15	0.55	0.5	0.2	0	0.8
River 9	0	0		0.1	0.47	0.4	0.1	0.75	0	0.5		0.2	0.18	
River 10		0.37	0.38	0.55	0.65	0.58			0.48	0.8		0.27	0.48	1

(b)

		PHARMACEUTICAL CLUSTER A				PHARMACEUTICAL CLUSTER B				PHARMACEUTICAL CLUSTER C				PHARMACEUTICAL CLUSTER D	
		Phrama 1	Phrama 9	Phrama 4	Phrama 12	Phrama 7	Phrama 3	Phrama 13	Phrama 2	Phrama 10	Phrama 5	Phrama 11	Phrama 6	Phrama 8	Phrama 14
RIVER GROUP 3	River 1	0.18	0.2	0.25		0.48	0.5		0.45	0.65	0.75	0.75	0.8		
	River 10		0.48	0.55	0.27		0.38	0.48	0.37	0.8	0.65		0.55		1
	River 7	0.32		0.23	0.21	0.52	0.32		0.29			0.54		0.85	
	River 4	0.3	0.28		0.25	0.59		0.49		0.8	0.5	0.6	0.45	0.88	0.95
	River 5		0.21	0.26	0.32	0.67	0.58	0.48	0.33	0.75		0.45	0.4	0.8	0.8
RIVER GROUP 2	River 2	0.25	0.3	0.22		0.33		0.35	0.4	0.48	0.6		0.55	0.9	0.85
	River 6	0	0		0.22		0.3		0.45	0.31	0.35		0.45	0.8	1
RIVER GROUP 1	River 3	0	0	0.3	0.2	0.15	0.1			0.53	0.6		0.6	0.75	0.85
	River 8		0.15		0.2	0	0.05	0	0.1	0.55		0.5	0.4		0.8
	River 9	0	0	0.1	0.2	0.1		0.18	0	0.5	0.47		0.58	0.75	

(c)

	PHARMACEUTICAL CLUSTER A	PHARMACEUTICAL CLUSTER B	PHARMACEUTICAL CLUSTER C	PHARMACEUTICAL CLUSTER D
RIVER GROUP 3	0.29	0.46	0.63	0.9
RIVER GROUP 2	0.17	0.34	0.45	0.88
RIVER GROUP 1	0.11	0.07	0.52	0.78

Figure 1. Schematic representing simultaneous clustering of 10 (hypothetical) rivers and 14 (hypothetical) pharmaceuticals studied on those rivers. (a): Detection rate (how often a pharmaceutical was positively detected when analyzed for) of the 14 pharmaceuticals (columns) measured across 10 rivers (rows) arranged in alphabetical order. Pharmaceuticals that are not studied in a river are shown as blank. (b) Rearranged blocks of pharmaceuticals and rivers that exhibit high degree of similarity. The SBM divides the 14 pharmaceuticals in 4 clusters (A to D, separated by blue vertical lines). The algorithm also divides the 10 rivers in three groups (1 to 3, separated by magenta horizontal lines). Each color represents a river-pharmaceutical block. As an example, "pharmaceutical cluster A – river group 1" reveals that the detection rates of pharmaceuticals in cluster A have the lowest detection rates for river group 1 and "pharmaceutical cluster D – river group 3" reveals that the detection rates of pharmaceuticals in cluster D have the highest detection rates for river group 3. (c) The probability of positively detecting an unstudied pharmaceutical (for example, pharma 8 at river 1) is 0.9 (as they belong to "pharmaceutical cluster D – river group 3" block).

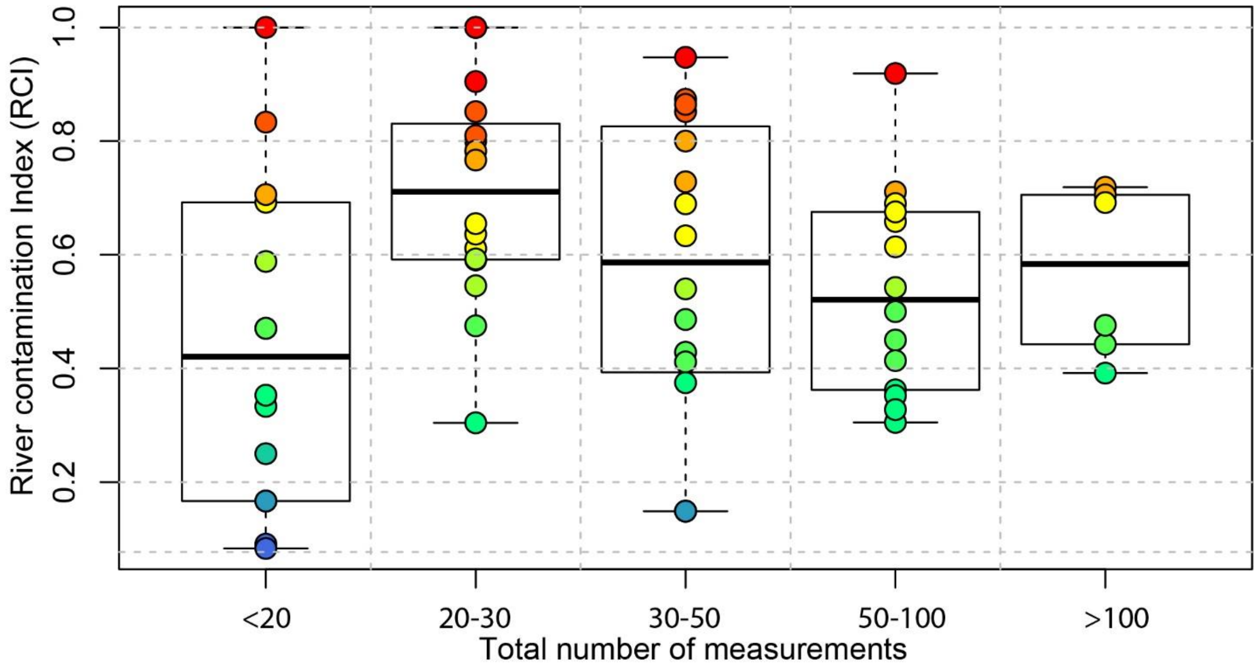

Figure 2. RCI of the rivers grouped by the total number of measurements on the river. The color palette represents lower to higher RCI (blue to red).

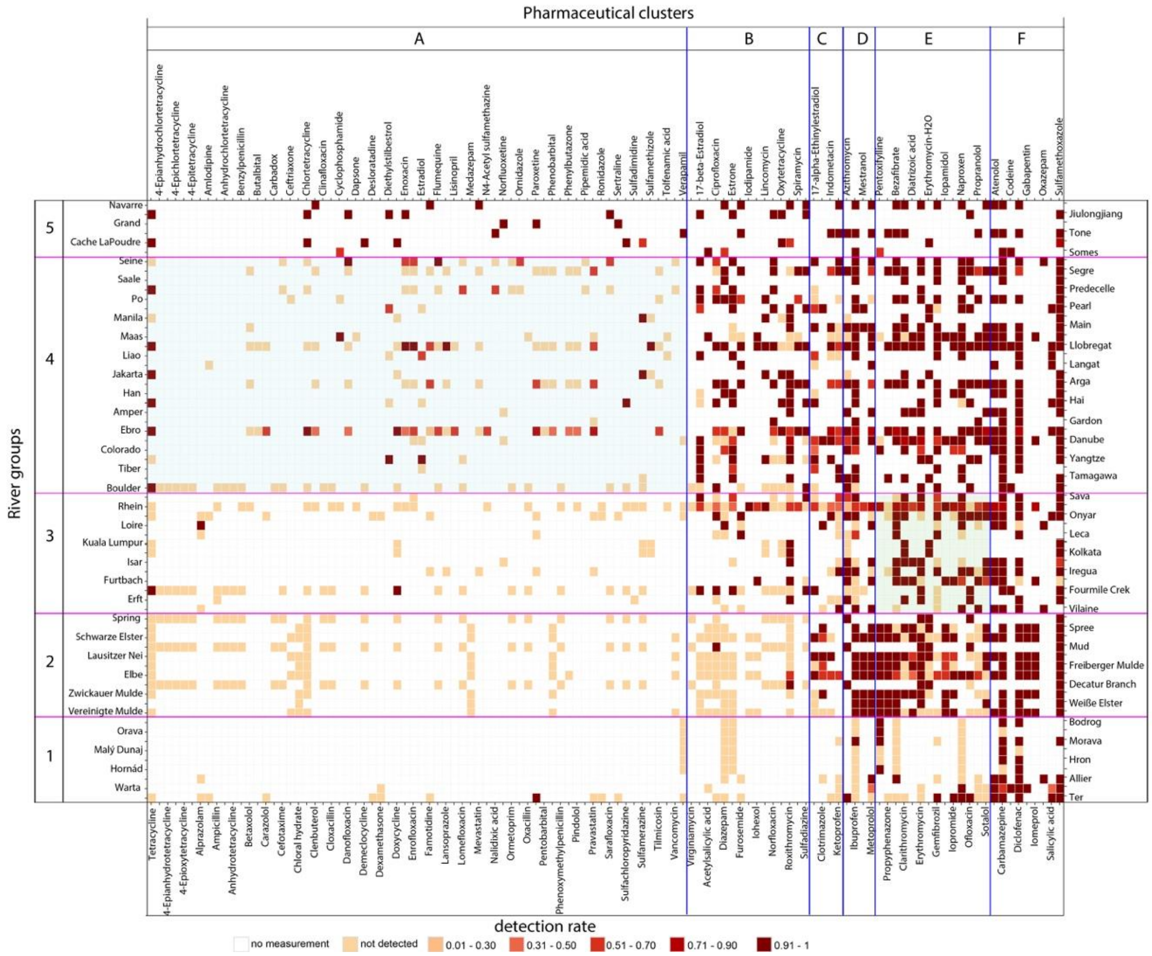

Figure 3. Detection rate of the 112 pharmaceuticals (columns) studied across the 64 rivers (rows). White square represents pharmaceuticals that were not studied at that river. The matrix has been ordered according to the river-pharmaceutical block. Rows 1-9, 10-21, 22-35, 36-60 and 61-64 represents river groups 1,2,3,4, and 5 (partitioned by magenta lines). Columns 1-67, 68-81, 82-85, 86-91, 92-103 and 104-112 represents pharmaceutical clusters A, B, C, D, E and F (partitioned by blue lines). The pharmaceuticals clusters are arranged according to the detection rate. Each rectangle enclosed by the magenta and blue lines is a pharmaceutical-river block. Blocks "A–4" and "E–3" are highlighted (lightly shaded) for illustrative purposes. Mean detection rate (and the 95% credible interval) for each river-pharmaceutical block is shown in figure 4. The name of the rivers in Asia are highlighted in (blue), Western Europe and North America (purple) and Eastern Europe (orange).

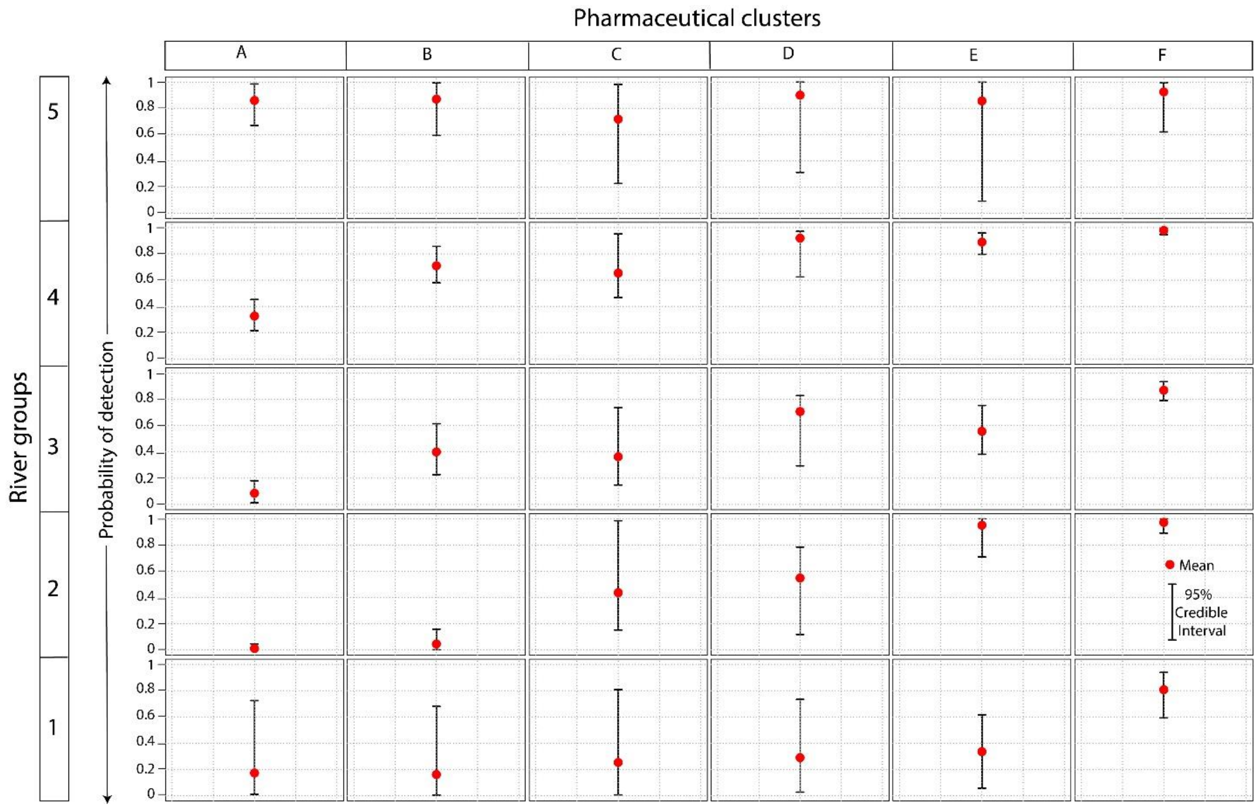

Figure 4. Mean probability (shown by red circle) and 95% credible interval (Shown as error bar) of positively detecting unstudied pharmaceuticals in each pharmaceutical cluster-river group.

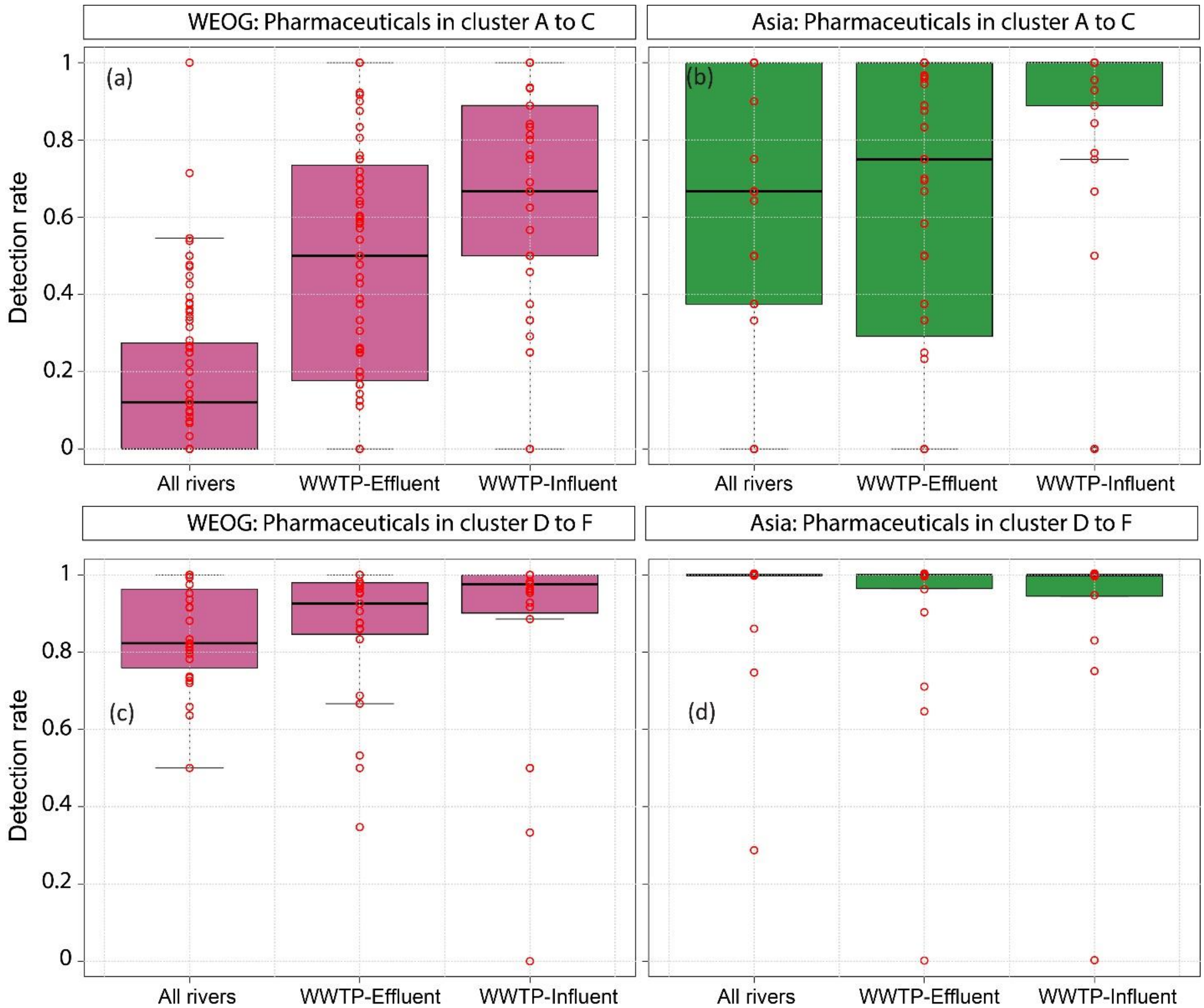

Figure 5. Detection rate of pharmaceuticals in rivers, WWTP-effluents and WWTP-influents. (a): pharmaceuticals in clusters A to C in WEOG, (b): pharmaceuticals in clusters A to C in Asia, (c): pharmaceuticals in clusters D to F in WENA and (d): pharmaceuticals in clusters D to F in Asia.

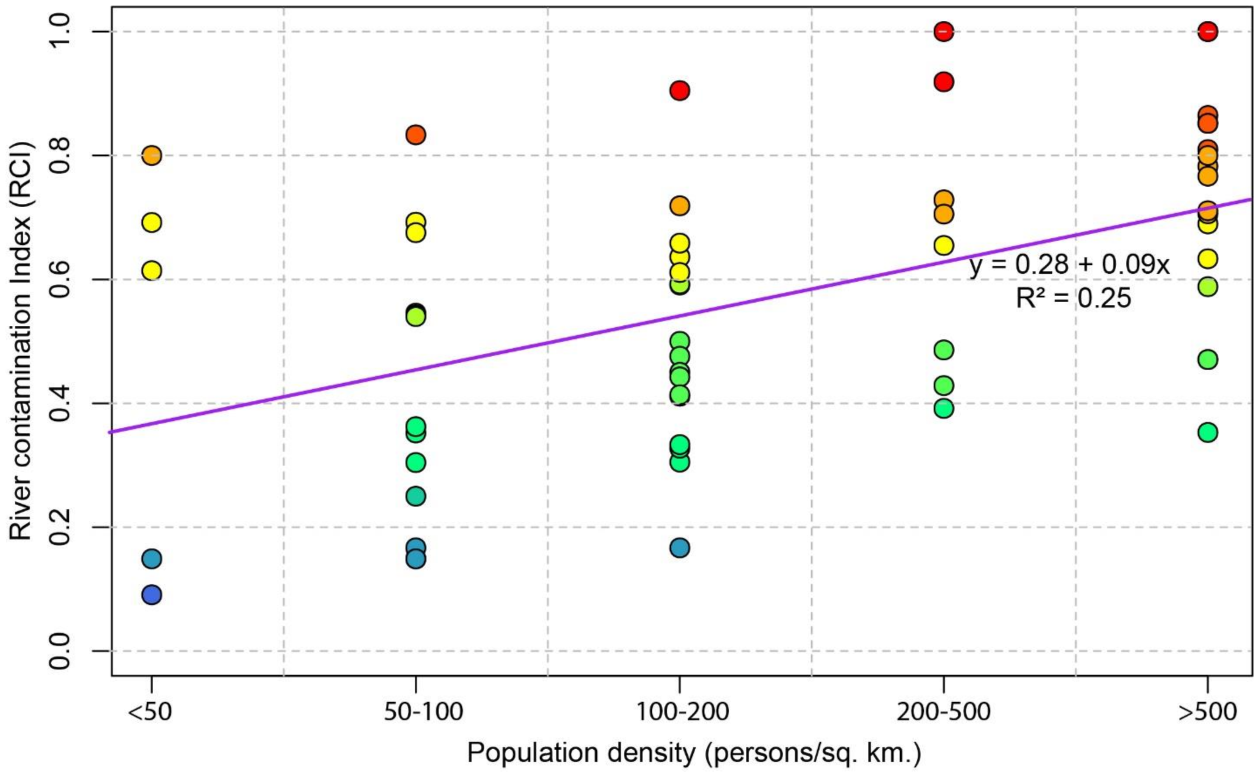

Figure 6. Relationship between river contamination index (RCI) and population density for the rivers analyzed in this study. Population density has been divided into 5 sub-classes (<50, 50-100, 100-200, 200-500 and >500 persons/square kilometer). Correlation between population density and RCI are statistically significant (p <0.05). The color palette represents lower to higher RCI (blue to red).

Table 1: Mean probability and 95% credible interval (values in bracket) of the detection rate of selected pharmaceuticals for River Colorado, Rhine and Elbe.

River	Pharmaceutical	Probability of detection
Colorado	Estradiol	35% (20-50%)
Colorado	Ciprofloxacin	65% (60-85%)
Colorado	Erythromycin	90% (80-100%)
Colorado	Diclofenac	98% (95-100%)
Rhine	Estradiol	10% (0-20%)
Rhine	Ciprofloxacin	40% (20-60%)
Rhine	Erythromycin	60% (40-80%)
Rhine	Diclofenac	90% (80-95%)
Elbe	Estradiol	0% (0-5%)
Elbe	Ciprofloxacin	3% (0-10%)
Elbe	Erythromycin	95% (65-100%)
Elbe	Diclofenac	97% (95-100%)