Mieux tenir compte des données non quantifiées : une application de la méthode du maximum de vraisemblance pour étudier les tendances spatio-temporelles du glyphosate et de l'AMPA dans la Seine.

Working with left-censored data: an illustration of the maximum likelihood estimation method to study the spatio-temporal trends of glyphosate and AMPA in the Seine River.

Lisa Teyssier[1], Jérémie Bonneau[1], Boram Kim[1], Damien Tedoldi[1]

[1]INSA Lyon, DEEP, UR 7429, 69621 Villeurbanne, France

RÉSUMÉ
Le travail présenté s'intéresse aux tendances spatio-temporelles des concentrations de glyphosate et de l'un de ses principaux produits de dégradation, l'acide aminométhylphosphonique (AMPA), en plusieurs points de surveillance de deux grandes rivières de l'agglomération parisienne (la Marne et la Seine, en amont et en aval de Paris). Pour traiter les données non quantifiées (jusqu'à ~50% des valeurs), nous avons appliqué la méthode du maximum de vraisemblance, afin d'illustrer une manière de ne pas dénaturer l'information contenue dans des séries statistiques contenant de telles données. Nous avons modélisé les distributions des concentrations en glyphosate et AMPA par une loi log-normale, et constaté une contribution significative des zones urbaines aux niveaux de contamination dans les rivières. Notamment, les concentrations en AMPA en aval de Paris étaient supérieures d'un ordre de grandeur par rapport aux sections rurales de la Marne. Les analyses des tendances temporelles ont montré un déclin général des concentrations de glyphosate, et d'AMPA dans une moindre mesure, suite à l'entrée en vigueur de la loi Labbé qui interdit l'usage des pesticides en milieu urbain. Bien que nos résultats mettent en évidence des tendances globales, l'étude des processus sous-jacents dépasse le cadre de ce travail : des analyses plus larges et plus systémiques avec des données supplémentaires sont nécessaires pour confirmer et consolider ces résultats.

ABSTRACT
This work investigates the spatial and temporal trends of glyphosate and one of its main degradation products, aminomethylphosphonic acid (AMPA), at several river monitoring points within the greater Paris area (Marne and Seine Rivers, upstream and downstream of the urban areas). To handle datasets containing up to ~50% left-censored data, the maximum likelihood estimation (MLE) method was applied, thus avoiding distorting the information through common practices such as ignoring these data points or substituting them with half the reporting limit. We aimed to illustrate a practical approach to handle left-censored data in environmental studies using glyphosate and AMPA as examples. We modeled the variability of their concentrations with a lognormal distribution, and found a significant contribution of the urban areas to pollution levels in surface waters. Notably, AMPA concentrations downstream of Paris were an order of magnitude higher compared to rural sections of the Marne river. Temporal trend analyses showed a general decline in glyphosate concentrations, as well as AMPA to a lesser extent, following the "Labbé" law prohibiting the use of pesticides in urban areas. While our findings suggest an influence of the Greater Paris urban areas to pollution levels, the study of underlying processes were beyond the scope of this work. Broader, more comprehensive analyses with additional data are necessary to confirm and consolidate these results.

MOTS CLÉS

Contaminants, Limite de quantification, Maximum de vraisemblance, Pesticides, Rejets urbains

Contaminants, Limit of quantitation, Maximum likelihood estimation, Pesticides, Urban discharges

1 INTRODUCTION

Surface waters undergo increasing pressure from both urban and agricultural areas. Among the causes of degradation of these water bodies is chemical contamination, which itself encompasses numerous substances with different emission and transfer processes, and varying degrees of impact. The widespread presence of biocidal substances in surface waters has become a matter of concern in recent decades, and several studies suggested that emphasis on agricultural pesticides may have led to overlooking urban sources of certain compounds (Wittmer et al., 2010; Merel et al., 2018). A pair of substances that typify these interwoven issues is glyphosate and its main degradation product, AminoMethylPhosphonic acid (AMPA), the occurrence of which has been ascertained in various rivers worldwide (Carles et al., 2019; Medalie et al., 2019) as well as in agricultural leachates and urban effluents (Grandcoin et al., 2017).

When handling and interpreting water quality data, a recurring difficulty lies in the presence of data below the reporting limit (either a limit of detection or a limit of quantitation). What makes this aspect even more complex is the fact that reporting limits are likely to change in time and space – together with the changes of chemical analysis service providers and/or analytical techniques – which complicates the calculation of relevant statistical measures and the identification of possible spatio-temporal trends. Although statistical methods do exist for taking appropriate account of left-censored data (Helsel, 2012), studies addressing surface water quality rarely take full advantage of them. Substitution methods (*e.g.*, replacing left-censored data with half the reporting limit) remain relatively common, and are even required by the French regulations for the calculation of mean values prior to assessing the chemical status of a watercourse (French Ministry of Ecology, 2023).

Hence, the objective of this communication is to analyze the pluriannual dynamics of glyphosate and AMPA in the Seine River, and to illustrate the refinements brought by Maximum Likelihood Estimation to preserve the integrity of datasets including left-censored data. In so doing, the main practical outcome is to assess the influence of urban areas on surface water contamination. In this abstract, we present preliminary data with a focus on the Seine and Marne Rivers upstream and downstream of the Paris conurbation.

2 MATERIALS & METHODS

2.1 Study area and data collection

Data were retrieved from the "Naïades" database (https://naiades.eaufrance.fr/) which gathers records of surface water quality observations on a national scale in France. Monitoring (*i.e.*, water sampling and analysis) is carried out at fixed "stations" on the various watercourses, at a frequency that globally varies from a few samples per year to a few samples per month. Samples from environmental compartments other than water (e.g., sediments) were not considered. Data are subject to a validation procedure before being made public in the "Naïades" database: only data qualified as "correct" have been retained for this analysis.

Homogeneous groups of stations were formed by pooling locations from a stretch of river (i) within an area with relatively uniform land use, and (ii) without major tributary (see Figure 1). Five groups of stations will be presented here: Marne River upstream of Paris, distinguishing between its rural and urban stretches; Seine River upstream of Paris (before its confluence with the Marne River), between Paris and the discharge point of the major wastewater treatment plant, and downstream as far as Gargenville. The earliest date with common records of glyphosate and AMPA at all stations was 2015. Stations that have been monitored on an occasional basis (a few months over the last two decades) were not included in this analysis.

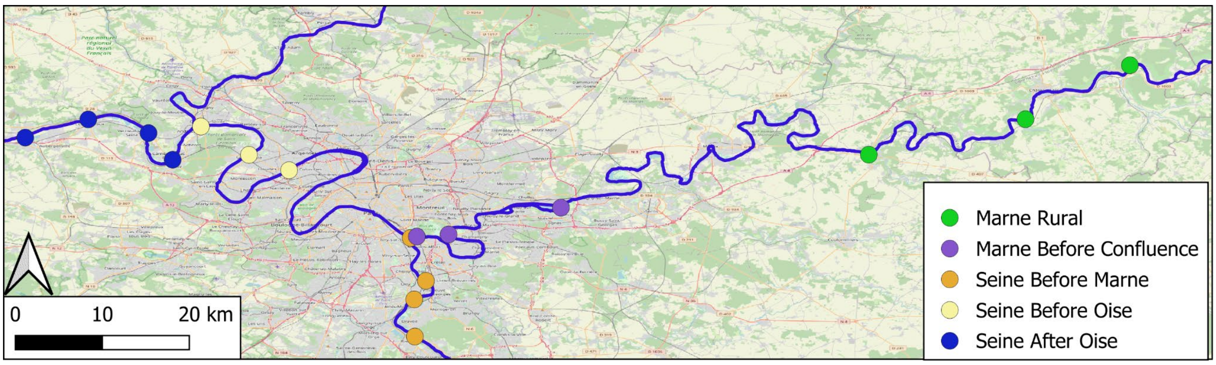

Figure 1 – Map of the study area (Paris region, France), with the five groups of stations along the Marne and Seine Rivers.

The available data between 2015 and 2024 were also grouped by time periods in order to assess temporal trends – if any – in glyphosate and AMPA concentrations. The delimitations of these periods were chosen to gather at least 40 values in each sub-group, and to distinguish notable regulatory changes regarding pesticide use in urban areas – especially the "Labbé" law prohibiting the use of pesticides from 2017 for public structures, and from 2019 for city-dwellers.

2.2 Statistical modelling and analysis

The lognormal distribution has been widely reported as suitable for describing the variability of contaminant concentrations in different environmental matrices, including water and air (Andersson, 2021). As a reminder, this distribution is characterized by two parameters, μ and σ, which respectively correspond to the mean and standard deviation of the natural logarithm of the variable under study. Maximum Likelihood Estimation (MLE) was applied to each homogeneous data group. So as to account for the presence of left-censored data, the likelihood function was defined as:

$$\mathcal{L}(\mu,\sigma) = \prod_{i=1}^{N_{quant}} f_{(\mu,\sigma)}(C_i) \times \prod_{j=1}^{N_{cen}} F_{(\mu,\sigma)}(RL_j) \quad (1)$$

where N_{quant} and N_{cen} refer to the number of quantified and left-censored concentrations, respectively, C_i is the value of the i^{th} quantified concentration, RL_j is the reporting limit for the j^{th} left-censored concentration, $f_{(\mu,\sigma)}$ is the probability density function of the lognormal distribution, and $F_{(\mu,\sigma)}(x) = \int_0^x f_{(\mu,\sigma)}(u)du$ is the cumulative distribution function. Maximization of $\mathcal{L}(\mu,\sigma)$ was carried out in R software, using the '*cenmle*' function available in the 'NADA' package (Lee, 2020). Subsequently, the estimation of μ and σ enabled theoretical properties of the distribution to be calculated.

Additionally, spatial and temporal variability in glyphosate and AMPA concentrations (*i.e.*, the existence of differences between the five stretches of river and time groups, as described in Section 2.1), was explored using the Peto-Peto test.

3 RESULTS & DISCUSSION

Across all groups, the proportion of values below the reporting limit ranged from 2 to 46% for glyphosate, and from 0 to 22% for AMPA, showing the ubiquitous nature of both substances in the Seine and the Marne Rivers. However, the "quantification frequency" is difficult to interpret in itself, because reporting limits are not constant – between 20 and 100 ng/L for both substances. A more robust approach, therefore, consists of comparing the statistical distributions of concentrations, as presented in the subsequent developments.

The results first illustrated the spatial evolution of glyphosate and AMPA concentrations along the rivers (Figure 2). Of interest is the fact that the upstream stretch of the Marne River (*Marne Rural*) showed significantly lower concentrations of AMPA than the following urban section (*Marne Before Confluence*), whatever the period considered – as confirmed by a Peto-Peto test (p-value $\ll 0.01$). Among the four time groups, the median AMPA concentration remained between 100 and 140 ng/L in the *Marne Rural* section, whereas it rose to between 180 and 430 ng/L in the *Marne Before Confluence* section. Conversely, no significant difference was observed for glyphosate between these two stretches of river, where the distributions remained very close, with median values consistently around 50 ng/L. These observations highlight the contribution of urban areas to the contamination of surface waters by AMPA, with some sources being independent of glyphosate. This is in line with the conclusions of Grandcoin *et al*. (2017), who mentioned the existence of sources of AMPA linked to phosphonates in wastewater, in addition to the degradation of glyphosate.

Understanding the increase in AMPA concentrations between the upstream part of the Seine River (*Seine Before Marne*) and the downstream section (*Seine Before Oise*) – here again with significant differences according to the Peto-Peto test – is less straightforward, as it can be related to the inputs from the Marne river and to the contribution of urban effluents from the inner city of Paris as well as the close suburbs.

As regards temporal trends, a clear decline in glyphosate concentrations was observed in all stretches of stations following the implementation of the "Labbé" law, the main steps of which took place between 2017 and 2019. For example, in the downstream section of the Seine River (*Seine After Oise*), the median value decreased from 110 to 50 ng/L between 2017 and the present period (similar to the above, a Peto-Peto test confirmed the significant nature of this difference – p-values < 0.02 for each couple of successive periods between 2017 and 2024).

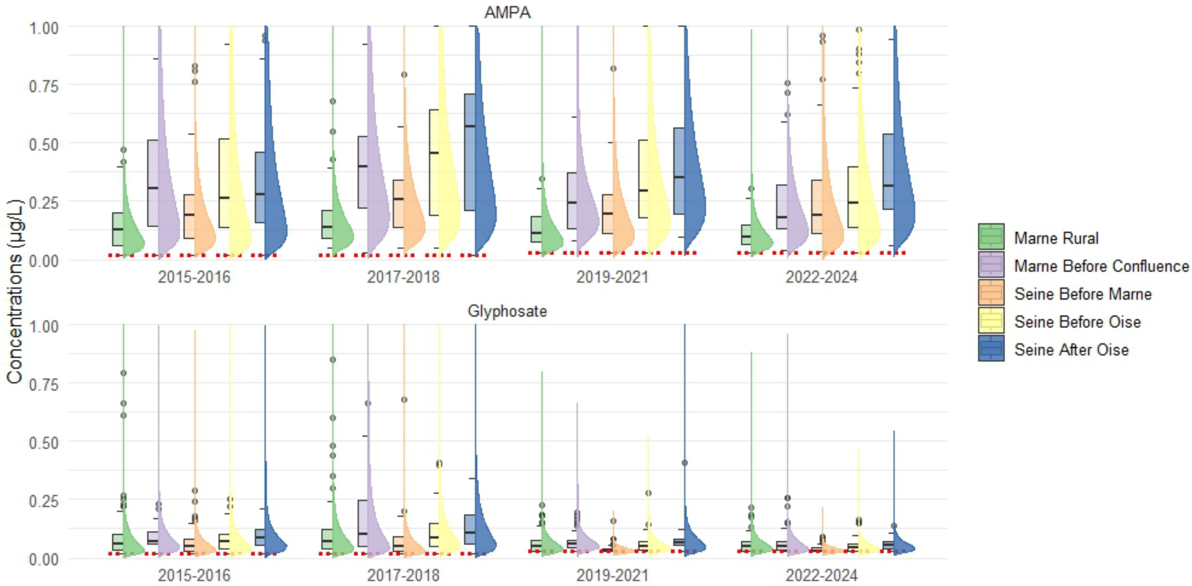

Figure 2 – Spatio-temporal evolution of AMPA (top) and glyphosate (bottom) concentrations in µg/L (zoomed in on the 0-1 µg/L range) in the Marne and the Seine Rivers, between 2015 and 2024 and across the five spatial groups of stations. Boxplots represent the measured data, censored at the highest reporting limit; continuous distributions represent the fitted lognormal distributions with the MLE method.

4 CONCLUSION

Concentrations of glyphosate and its main degradation product AMPA were extracted from the Naïades database, for the period 2015-2024, in both the Seine and the Marne rivers around the Greater Paris conurbation. We applied the maximum likelihood estimation method to handle left-censored data, to model statistical distributions of concentrations within homogeneous groups of stations. The results confirm that urban areas are significant contributors to AMPA pollution in aquatic environments, in line with previous studies. The story for glyphosate was more complex, although concentrations tend to decrease, as one would expect following regulations from the last decade. In conclusion, we illustrate a statistically adequate way to handle left-censored data, with the example of two pollutants of concern, that can be used by practitioners in environmental analysis.

REFERENCES

Andersson, A., 2021. Mechanisms for log normal concentration distributions in the environment. *Sci. Rep.* 11, 18075. https://doi.org/10.1038/s41598-021-97822-2

Carles, L., *et al.*, 2019. Meta-analysis of glyphosate contamination in surface waters and dissipation by biofilms. *Environ. Int.* 124, 284-293. https://doi.org/10.1016/j.envint.2018.12.064

French Ministry of Ecology, 2023. Arrêté du 9 octobre 2023 modifiant l'arrêté du 25 janvier 2010 relatif aux méthodes et critères d'évaluation de l'état écologique, de l'état chimique et du potentiel écologique des eaux de surface pris en application des article R. 212-10, R. 212-11 et R. 212-18 du code de l'environnement.

Grandcoin, A., *et al.*, 2017. AminoMethylPhosphonic acid (AMPA) in natural waters: Its sources, behavior and environmental fate. *Water Res.* 117, 187-197. https://doi.org/10.1016/j.watres.2017.03.055

Helsel, D.R., 2012. Statistics for censored environmental data using Minitab and R. Wiley, Hoboken, New Jersey.

Lee, L., 2020. Package 'NADA': Nondetects and Data Analysis for Environmental Data. R package version 1.6-1.1.

Medalie, L., *et al.*, 2020. Influence of land use and region on glyphosate and aminomethylphosphonic acid in streams in the USA, *Sci. Total Environ.* 707, 136008. https://doi.org/10.1016/j.scitotenv.2019.136008

Merel, S., *et al.*, 2018. Occurrence and overlooked sources of the biocide carbendazim in wastewater and surface water, *Environ. Pollut*. 239, 512-521. https://doi.org/10.1016/j.envpol.2018.04.040

Wittmer, I.K., *et al.*, 2010. Significance of urban and agricultural land use for biocide and pesticide dynamics in surface waters, *Water Res.* 44, 2850-2862. https://doi.org/10.1016/j.watres.2010.01.030