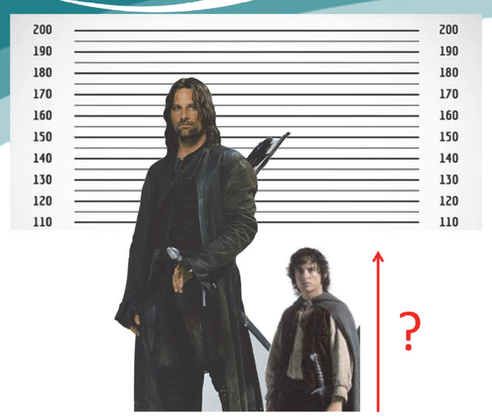
I.S. INTEGRATIVE SCIENCES Rivers RECHERCHES ET ACTIONS



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.

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.

Introduction

Surface waters undergo increasing pressure. Among the causes of degradation of these water bodies is chemical contamination, which itself encompasses numerous substances with different emission and transfer processes. 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[1,2]. 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[3,4] as well as in agricultural runoff and urban effluents[5].

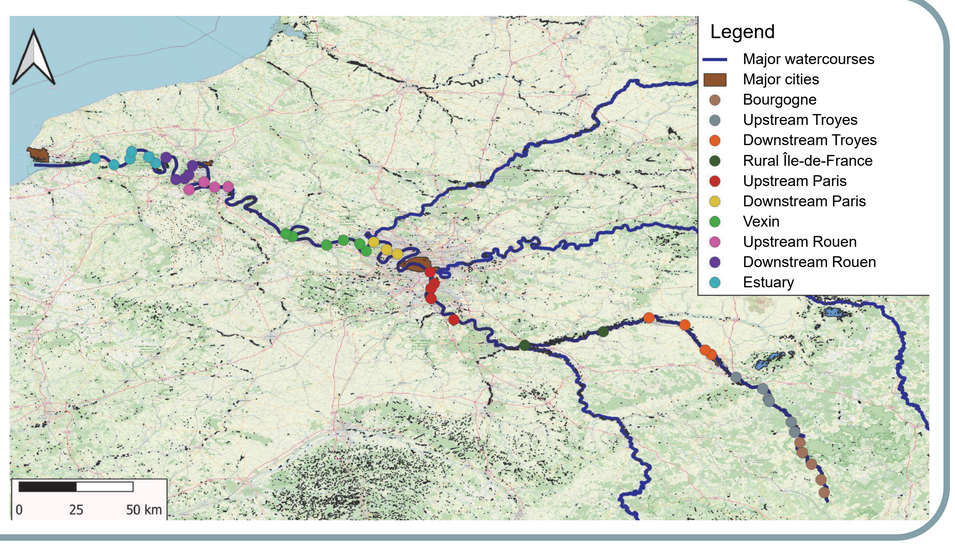
When handling and interpreting water quality data, a recurring difficulty lies in the presence of **data below the reporting limit** (either a limit of detection or limit of quantitation). Additionally, the fact that reporting limits are likely to change in time and space – together with the changes of chemical analysis service providers and/or analytical techniques – complicates the calculation of relevant statistical measures and the identification of possible spatio-temporal trends. While statistical methods do exist for taking appropriate account of left-censored data, studies addressing surface water quality rarely take full advantage of them.

Hence, the objective is to explore 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.

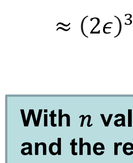
[1] Wittmer *et al.*, 2010, *Water Res.* [2] Merel *et al.*, 2018, *Environ. Pollut.* [3] Carles *et al.*, 2019, *Environ. Int.* [4] Medalie *et al.*, 2019, *Sci. Total Environ.* [5] Grandcoin *et al.*, 2017, *Water Res.*

Study area, data collection and analysis

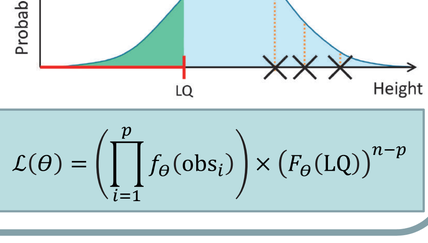
- Data sourced from the French national database on surface water quality NAÏADES
- 50 measurement stations selected along the Seine River, generally with >1 sample per month and station between 2015 and 2024
- Pooling into 4 temporal × 10 spatial groups (stretches of river without major tributary and with homogenous land use), so as to get > 40 values in each group
  - LQ between 50 ng/L and 1 µg/L
  - Quantitation freq. between 0 and 100%
- Modelling with a lognormal distribution, fitted with Maximum Likelihood Estimation

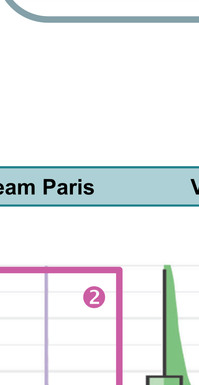


Maximum Likelihood Estimation



You've just arrived in Middle Earth, but you don't know much about that land. You have read that it is populated by elves, hobbits and humans [*sorry for the oversimplification*], whose size distribution you've been informed about. You observe four individuals from the same population, measure them (black crosses), and try to guess in whose territory you are…





If that observation were an elf, then:

$$\mathbb{P}\big(\text{Height} \in \mathcal{V}(\text{obs})\big) = \mathbb{P}(\text{Height} \in [\text{obs}-\epsilon\,;\,\text{obs}+\epsilon]) = \int_{\text{obs}-\epsilon}^{\text{obs}+\epsilon} f_{\text{elf}}(x)\,dx \approx 2\epsilon\, f_{\text{elf}}(\text{obs})$$

So if that were a group of elves, then:

$$\mathbb{P}\left(\begin{matrix}\text{Height}_1 \in \mathcal{V}(\text{obs}_1) \cap \cdots \\ \cap\, \text{Height}_4 \in \mathcal{V}(\text{obs}_4)\end{matrix}\right) \approx (2\epsilon)^4 \prod_{i=1}^{4} f_{\text{elf}}(\text{obs}_i)$$

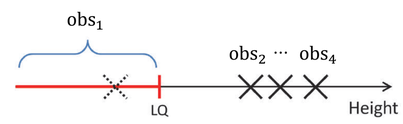
$\Longrightarrow$ very ***unlikely*** as $f_{\text{elf}}(\text{obs}_1) \approx 0$!

The population from which the observations are *most likely* to originate is the one that maximizes the likelihood function:

$$\mathcal{L}(\text{Population}) = \prod_{i=1}^{n} f_{\text{Population}}(\text{obs}_i)$$

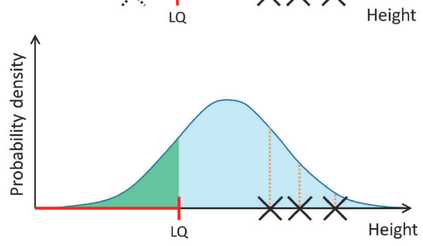
In the presence of left-censored data

- The available observations are replaced by:



- The probability of interest becomes:

$$\mathbb{P}\left(\begin{matrix}\text{Height}_2 \in \mathcal{V}(\text{obs}_2) \cap \cdots \cap \text{Height}_4 \in \mathcal{V}(\text{obs}_4) \\ \cap\, \text{Height}_1 < \text{LQ}\end{matrix}\right) \approx (2\epsilon)^3 \times \prod_{i=2}^{4} f(\text{obs}_i) \times \int_0^{\text{LQ}} f(x)\,dx$$



With n values among which p are quantified and the remainder are <LQ, the likelihood is replaced by the following expression:

$$\mathcal{L}(\theta) = \left(\prod_{i=1}^{p} f_\theta(\text{obs}_i)\right) \times \big(F_\theta(\text{LQ})\big)^{n-p}$$

Results

▲ Evolution of the concentrations of glyphosate and AMPA along the Seine river, for different time periods. For each spatio-temporal group, the boxplot (left) is drawn from the data, after censoring at the highest reporting limit (red line). The modelled lognormal distribution (right) is obtained with the MLE method.

Some keys to interpretation of this figure

❶ Ability to estimate the parent distribution, even when the highest quantitation limit is comparatively high, and hiders the use of "conventional" descriptive statistics such as censored boxplots.

❷ Evidence of an impact of urban areas (Troyes, Paris conurbation) on the contamination of the Seine River – which for Paris cannot be explained by the inputs of the Marne River (*data not shown*).

❸ Decline of glyphosate concentrations in surface waters following the implementation of the "Labbé law" prohibiting the use of pesticides in urban areas, but less impact on AMPA $\Longrightarrow$ additional urban sources.

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON deep

Lisa Teyssier, Jérémie Bonneau, Boram Kim, Damien Tedoldi

INSA Lyon, DEEP, UR 7429, France