# A Bayesian approach to determine the occurrence of feminine hygiene products on the shores of Lake Geneva.

Roger Erismann, Romain Tramoy, Bhavish Patel, Montserrat Filella

***Affiliations***

**Roger Erismann**, - Hammerdirt, Brugstrasse *39,* CH-*2503* Biel, Switzerland

**Romain Tramoy** - Laboratoire Eau Environment et Systèmes Urbains (LEESU), 61 Av. du Général de Gaulle, F-94000 Créteil, France

**Bhavish Patel** - Paul Scherer Institute (PSI), Forschungsstrasse 111, CH-5232 Villigen PSI, Switzerland

**Montserrat Filella** - Department F.-A. Forel, University of Geneva, Boulevard Carl-Vogt 66, CH-1205 Geneva, Switzerland

**Contact:** *roger@hammerdirt.ch*

***Abstract***

*[to be completed]*

**Keywords**: Switzerland, Lake Geneva, Bayes theorem, feminine hygiene products,marine-litter, riverine-input, LCA,

INTRODUCTION

The United Nations published the international guide to collecting beach litter in 2008 {cite}`unepseas`. This publication was followed by another guide developed by OSPAR {cite}`osparguidelines` in 2010 and then in 2013 the EU published Guidance on Monitoring Marine Litter in European Seas {cite}'mlwguidance`. Riverine Litter Monitoring - Options and Recommendations was published in 2016 as evidence was mounting that rivers are major sources of marine litter. {cite}`riverinemonitor`

As a result, thousands of observations have been collected following a very similar protocol {cite}`mlwdata` {cite}`ospardata`. These data are collected by different organizations throughout the continent. Each observation is a categorical list of objects and their respective quantities per meter of shoreline together with the location of the observation {cite}`mlwguidance` {cite}`osparguidelines`. The same protocol has been in place in Switzerland since November 2015, targeting regional lakes and rivers. {cite}`iqaasl`

The data collected by volunteers were considered fit for the purpose of establishing beach-litter threshold values by the Marine Litter Technical Group of the EU. The lack of quantitative research on the harmful effects of beach litter, specifically the dose-effect relationship between beach-litter and ecological harm, precludes the establishment of threshold values based on this metric. Therefore, threshold values and baselines were adopted according to the precautionary principle and determined by using the upper limit of the 95% confidence interval of the 10th percentile from the combined data set of 2015-2016 beach litter surveys within the EU. {cite}`threshholdeu`

The potential applications for these data go beyond threshold values, in particular with respect to Life Cycle Assessments and the sustainable economy. The Life Cycle Initiative (LCI) is a public private partnership that includes France, Switzerland, Germany and the EU with a stated goal of advancing the understanding of life cycle thinking by private and public decision makers. In partnership with Plastics Europe, LCI has been attempting "to integrate potential environmental impacts of marine litter, especially plastic, in Life Cycle Assessment (LCA) results". {cite}`marilca` {cite}`lci`.

The consequences of plastic leaked into the environment are not accounted for in the current practice of LCA {cite}`woods2021107918`. This could be in part for the same reasons that the EU adopted the precautionary principle as opposed to a one based on dose-effect when developing threshold values. When social-economic consequences are taken into consideration, it is difficult to find quantitative data that accurately describes the effects of beach-litter, thus excluding this type of information from a systematic method to assess the consequences (threshold values or LCA). {cite}`threshholdeu`

Nevertheless, recent attempts have been made to develop a Marine Litter Indicator (MLI) for LCAs. Based on the commonsense hypothesis that there is a relationship between the number of objects produced, the likelihood that they will be littered and the time it takes to degrade them, it was concluded that returnable PET bottles or returnable glass bottles would have the lowest MLI, all things considered {cite}`plasticorglass`. A very similar method was used when proposing an MLI for plastic carrier bags. {cite}` civancikuslu2019621`

What is not included in the LCA, or in the threshold values is a coherent method to determine progress towards reduction goals or the accuracy of the LCA. In the case of threshold values, the sampling period is six years with a minimum of 40 samples, intermediate assessments are considered voluntary and no further guidance is given {cite}`threshholdeu`. Beach litter survey results are not included in the calculation of the proposed MLI. {cite}`plasticorglass` {cite}`civancikuslu2019621`.

The current EU threshold for beach-litter greater than 2.5 cm is 15 objects per 100 m of shoreline (15 p/100 m) {cite}`threshholdeu`. Considering that the median value was 133 p/100 m, regional administrations will need to allocate resources to meet the threshold value if they want to achieve good environmental status under the Marine Strategy Framework Directive {cite}`goodenvironmentalstanding`. Determining the most effective solutions will require that stakeholders conduct intermediate assessments to evaluate progress and compare probable outcomes based on that progress.

Furthermore, if the adoption of an MLI is intended to reduce the consequences of marine litter, it is not clear from either proposed method how the LCA would change litter-survey outcomes or how that change will be quantified going forward. The topic of threshold values and LCA came to our attention when we wanted to determine the probability of finding a feminine hygiene product on the shores of Lake Geneva, Switzerland based on previous beach litter survey data.

XXXX HERE INTRODUCE BAYES APPROACH (just say that we will use it) APPLIED TO A PARTICULAR CASE: what follows. Mention why St Sulpice, why Mediterranean beaches.

End period plastics is a campaign in the UK to eliminate plastics in feminine hygiene products (FHP). The campaign includes meetings throughout Europe with manufactures, suppliers and vendors of plastic FHP. In preparation for a meeting with a producer of FHP in Geneva, the team from End Period Plastics had questions about the incidence of these objects in Lake Geneva {cite}`endperiodplastics`.

Tampon applicators and tampons are part of a group of specific items (code G96 & G144) that are found on the beach and that most likely originate from toilet flushing or a wastewater treatment facility {cite}`obriain2020116021` {cite}`padbackingstrips` {cite}`increasingplastics`

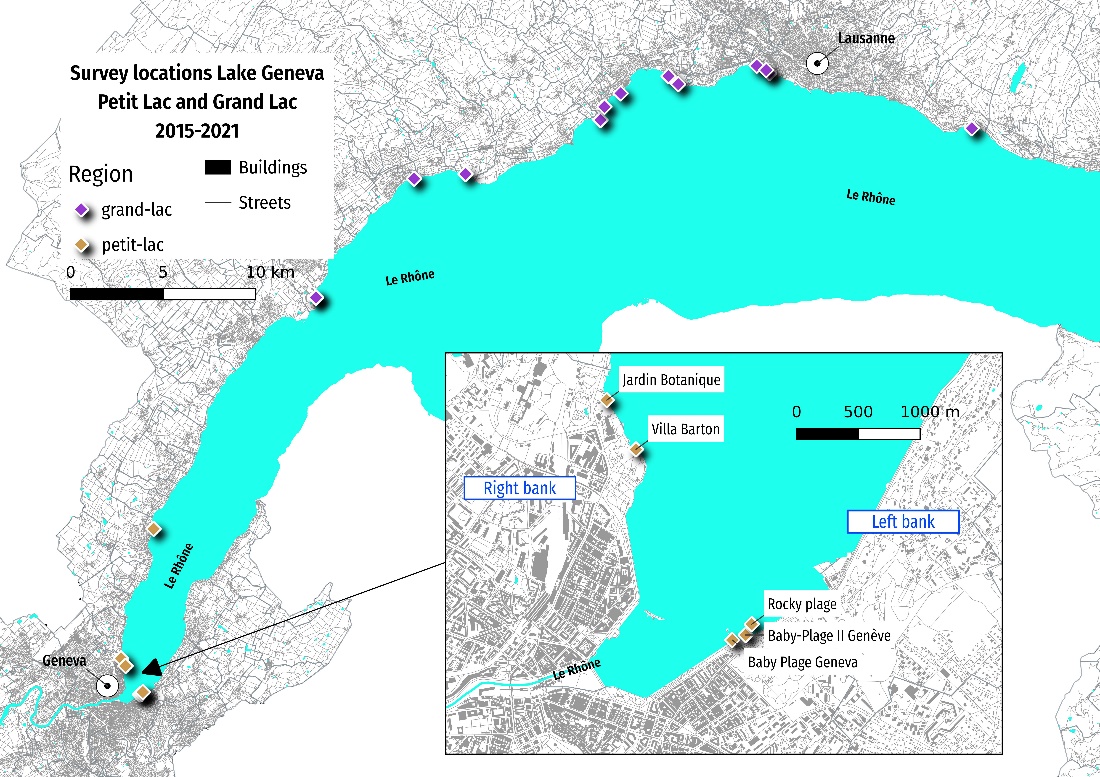
To illustrate the abundance of FHP on the lake, a litter survey was conducted with End Period Plastics staff and hammerdirt surveyors. Because of time and transportation constraints, this litter survey had to be conducted on a site with the following criteria: (I) access with public transport, (ii) completed in less than three hours, and (iii) preferably a location where there was a chance to find this specific item quickly.

Prior to selecting a meeting place, the location that had the highest likelihood of containing at least one FHP had to be determined.

DATA AND METODS

Study site

Lake Geneva is a perialpine lake at an altitude of 372 m above mean sea level located between France and Switzerland. It has an elongated shape, a surface area of 580 km2, and maximum length and width of 72.3 and 14 km, respectively. The average water residence time is 11.3 years. The lake is subdivided into two sub-basins: the *Grand Lac* (large lake) (309 m deep) and the *Petit Lac* (small lake) (medium-sized basin with a volume of 3 km3 and a maximum depth of about 70 m. Lake Geneva is fed by a large number of rivers and streams but most of the water enters through the Rhône River. The permanent population (2011) in its watershed is: 1,083,431 (France: 142,229, Switzerland: 941,202) but also hosts a large tourism population. There are 171 wastewater treatment plants (population equivalent: 3,009,830). {cite}`cipel2019`



Sampling

The first sample was recorded on Lake Geneva on November 23, 2015. Between November 2015 and June 2021 there were 247 beach-litter surveys at 38 different locations. In total there were 78,104 objects removed and identified, of which 358 objects were either tampons or tampon applicators (0.45%). {cite}`iqaasl`

There are three separate sampling periods, each sampling period represents an initiative or campaign to communicate the quantities of litter on the shoreline. The sampling periods are not of the same length nor is the sampling frequency fixed, except for a few specific locations in period two and three. There were no samples collected in Geneva during sampling period two.

1. Project one: 2015-11-15 to 2017-03-01; the first project on Lac Léman (MCBP)
2. Project two: 2017-04-01 to 2018-05-01; the Swiss Litter Report (SLR)
3. Project three: 2020-03-01 to 2021-11-01; the start of IQAASL up to two weeks before the survey with Ella

FHP were present in 102 samples (41%). However, FHP are not found at the same rate at all locations. There were 10 locations with only one sample (all periods included), of those ten at least one FHP was identified in 4 of the 10 samples/locations.

Of the other 28 locations, FHP were found at all locations except for three. Those three locations, in different regions of the lake had two samples each. All locations with at least three surveys had at least one FHP identified in a survey. For all locations with at least three samples, the minimum ratio of samples with FHP was 0.07 at baye-de-clarens and there were two locations that have a ratio of 1 (all samples have FHP products), la-pecherie at Allaman and parc-des-pierrettes in Saint Sulpice, see Annex table 1a.

Table 1. Number of locations, samples and positive samples per sampling period Lake Geneva. Table 2. Number of locations, samples and positive samples per period for all locations within the city limits of Geneva.



Statistical methods

The location with the greatest probability, theta, of finding an FHP will be determined by considering the results from each location of interest given the data from the lake up to the day prior to sampling. The statistical conclusions about the parameter theta are made in terms of probability statements. {cite}`bayesgelman`

Before any statement can be made about the relationship between the data and theta, they must be combined into a joint probability distribution. Bayes' rule does that by setting the conditional probability of theta to the observed data:

1. p( theta | data) = p(theta)p(data | theta)/p(data)

The expression to the left of the equal sign reads "the probability of theta given the data*", which is different than p( data | theta ). In 1 theta is being conditioned by the observations, in the latter the data is being conditioned by* theta. {cite}`bayesdowney`

These statements conform to a common-sense interpretation about any conclusions that may arise from the observed survey results. The uncertainty of those conclusions is defined by the range of the 94% Highest Density **Interval of probability** (HDI) as opposed to a confidence interval. The interval of probability contains the likely values of theta given the data and our **prior beliefs**. {cite}`bayesgelman` {cite}`bayesdowney` {cite}`bayespillon`

The equality can be defined in more general terms:

2. Posterior = (prior \* likelihood)/normalizing constant

The prior is the estimate or belief about theta prior to seeing the data. The likelihood is the chance of observing the data given theta. The posterior is the probability distribution of theta given the data. The normalizing constant is the total probability of the data and ensures that when theta is integrated on [0,1] it integrates to 1. {cite}`bayesrules`

To construct the model the prior, posterior and likelihood functions need to be defined. The probability of finding an FHP anywhere on the lake ranges from [0,1] and is not constant. Thus, theta can take on different values depending on the quantity at the location of interest and the surveyor's ability to find and recognize the object as an FHP.

The results for each survey were transformed to Boolean values. If an FHP was recorded at a survey FHP = 1 and 0 otherwise. This reduces each survey to one independent Bernoulli trial. Therefore, the probability of finding an FHP at any location in n samples can be described by the probability mass function of the Binomial distribution {cite}`bindistwiki`:

3. f(FHP,n,theta) = (n! / FHP!(n-FHP)!) \* thetak \* (1-theta)n-FHP

In this configuration (3) theta only takes on one value. To account for the different possible values of theta, its' value can be set to the Beta distribution resulting in a Binomial distribution with parameter theta which is defined by a Beta distribution with parameters alpha and beta.

4. FHP | theta ~Bin(n, theta)

5. theta ~ Beta(alpha, beta)

The Beta distribution, defined on [0,1], takes two parameters, the number of times that an FHP was found = alpha, and the number of times that an FHP was not found = beta (n-FHP). *A conjugate prior* to the Binomial distribution, this model has many applications to any situation where the parameter of interest is in the range [0,1]. {cite}`bayesgelman` {cite}bayesjefferys {cite}`bayesrules`

The probability density function of the Beta distribution is {cite}`betawiki`:

6. f(theta; alpha, beta) = (1/Beta(alpha, beta))\*thetaalpha-1\*(1-theta)beta-1

When sampling started in 2015 there were no reference values for the region. Count surveys of litter data in the maritime environment had produced volumes of data but under significantly different conditions. Without reference values we had no prior assumptions on the probability of finding an FHP and assumed the probability was the same for all locations on the lake.

The assumed distribution of theta prior to November 2015 is therefore a standard uniform distribution defined by Beta(1,1). This reflects the experience and expectations at the time. {cite}`bayesjefferys` {cite}`catalogofpriors` {cite}`noninformativepriors`

The general forms in 1 and 2 can now be defined using elements of 3 and 6, according to the model defined in 4 and 5. Note that the second and third expressions on the right side of the equations in 3 and 6 are the same for the Binomial and Beta distributions.

7. Prior = (1/Beta(alpha, beta))\*thetaalpha-1\*(1-theta)beta-1

8. Likelihood = (n! / FHP!(n-FHP)!) \* thetaFHP \* (1-theta)n-FHP

9. Posterior = P( theta | data )

Putting that together by dropping the normalizing constants that do not depend on theta (10) and combining like values results in the posterior distribution of theta (11).

10. P( theta | data ) <prop to>thetaalpha-1\*(1-theta)beta-1 \* thetaFHP \* (1-theta)n-FHP

11. P( theta | data ) <prop to> theta(alpha + FHP) – 1\*(1-theta)(beta + n – FHP)-1

The posterior distribution (11) is an unnormalized Beta distribution with parameters alpha and beta. This means that the **prior** uninformed estimate in November 2015 can be updated with the data from each sampling period in sequence. {cite}`bayesrules` {cite}`bayesdowney`

Using this method, the results from sampling period one (which incorporate the initial estimate) become the prior distribution for the results of sampling period two and this process is repeated until the last sampling period. {cite}`bayesdowney`

**Assumptions**

1. The samples are independent and identically distributed
2. Theta is approximately equal for all locations which is the expected value for the lake
3. The expected result for the lake or the region is the best estimate for locations without samples
4. exchangeability of data {cite}`bayesgelman`

**Computational methods**

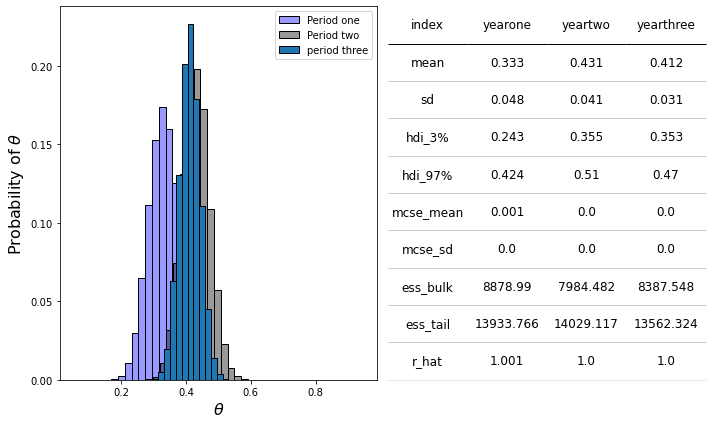
Markov Chain Monte Carlo (MCMC) is a general method used to simulate probability models. A Markov chain is a model that describes a sequence of possible events where the probability of each event depends only on the results of the previous event. {cite}`bayeskruschke` {cite}bayesdowney` {cite}`bayespilon`

The implementation of MCMC methods is done with PyMC3 v3.1 {cite}`pymc` and the results are analyzed with ArviZ v0.11.4 {cite}`arviz`, SciPy v1.7 {cite}`scipy` and pandas v1.34 {cite}`reback2020pandas` all running in a Python v3.7 {cite}`python` environment.

The data and methods are available in the repository: https://github.com/hammerdirt-analyst/finding-one-object

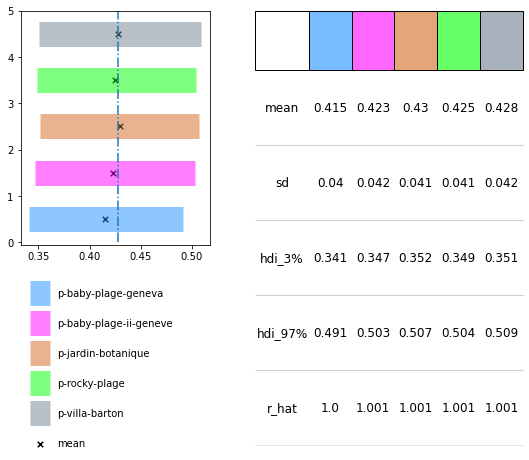
## RESULTS

Before estimating the probable values of theta for a specific location or region, the third assumption needs to be calculated for each sampling period. As shown in Figure 2, at the end of year one it was apparent that the probability of finding FHPs was not the same at every survey. The expected probability is highest in year two, but the smallest 96% HDI is in year three: 12%.

Figure 2

Recall that there were no samples in period two in Geneva and that not all locations are sampled in each sampling period (Table 2). The assumption is that without other evidence, theta at any location is approximately equal to theta for the lake. This concept is reciprocal, if theta lake is under consideration all locations that have valid samples on the lake are included in the estimation (Figure 3). Evidence is defined as sample results.

Figure 3



*Figure 3. The probable values of theta for the survey locations in Geneva given all the survey data for the lake.* ***Left:*** *the difference in means is less than 0.01 for any of the locations.* ***Right:*** *villa-barton has the highest min and max HDI values as well as the highest mean.*

Since Geneva is in the Petit Lac, rather than using all the data from the lake, the probability of each location can also be estimated by restricting prior data to surveys in the sub-basin (figure 4).

Figure 4

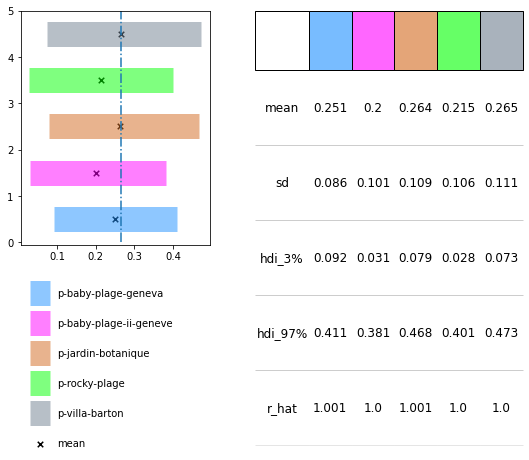


Figure 4. The probable values of theta given the data from the Petit Lac. **Left:** the locations on the right bank still maintain the advantage, but the expected value of theta for the Petit Lac is much lower than the rest of the lake. **Right:** the 94% HDI of each location in Geneva has a greater range when just the data from the Petit Lac is considered.

The shores of Saint Sulpice have been monitored yearly since 2016 by students from the EPFL. Contrary to locations in Geneva there is only one sample per location per year, apart from tiger-duck beach, which had two samples in year five (figure 5).

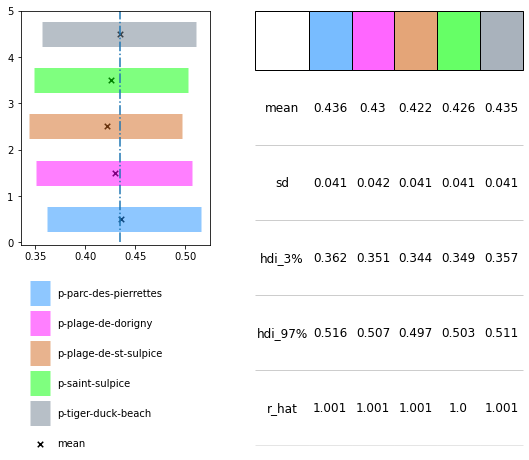


Figure 5

Saint Sulpice is in the Grand Lac. This subbasin has a higher expected value of theta (figure 6). The expected value of theta is approximately the same for all locations and higher than the lake or Geneva.

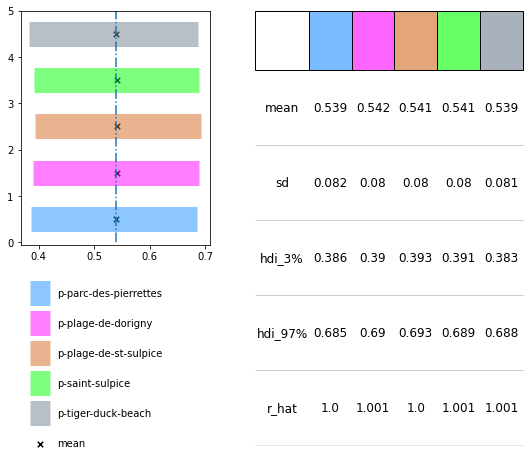


Figure 6

DISCUSSION

It was decided that the best chances of finding an FHP in Geneva was at villa-barton. The proximity to the train station and the time it takes to complete a survey as well as the likelihood of finding an FHP were considered.

The locations on the left bank of the lake, (baby-plage, rocky-plage and baby-plage-ii) had only 3/14 positive samples in the most recent sampling period, well below the mean for the lake.

FHP had been found at villa-barton and at jardin-botanique in the previous sampling periods (4/15) and those locations had only been sampled twice in the third period, with a success rate of ½. Suggesting that the results from figure 2 confirm the assumption that the probability of finding an FHP is not equal for all locations and there is a slightly better chance on the right bank.

Cotton buds or cotton swabs are found in 81% of all surveys on Lake Geneva {cite}`iqaasl` and mostly originate from the same source as FHP: toilets and wastewater treatment facilities. When cotton swabs are considered the difference between villa-barton and the other locations is more evident. The expected value of theta is highest and the 94% HDI is smallest (support 1a). Supporting the decision to survey the right bank as opposed to the left bank because the occurrence of objects with the same source is also elevated.

The choice of prior data has an influence on the posterior distribution {cite}`catalogofpriors` {cite}`noninformativepriors` {cite}`bayesjefferys`. When the lake results are considered both city's aggregate near the expected value for the lake, the differences between the locations is minimized because of the overwhelming number of surveys on the lake when compared to any single location (figure 4 and figure 2).

When the prior data was limited to observations in the Petit Lac, figure 3, the expected value of theta is 15-20 points lower than the expected value of the rest of lake. The locations on the right bank have the highest expected values of theta and the 94% HDI extends to 41%, supporting the conclusion that the expected value of theta is greater on the right bank.

When just the results from the sub-basins are considered the range of the 94% HDI increases both in Saint Sulpice and Geneva, reflecting the increase in uncertainty because there is less data being used to assess the probable values of theta. The number of samples per location is very similar in Saint Sulpice as well as the survey results, thus the probability of finding an FHP is approximately equal for all locations (figure 5). In Geneva the locations that have the fewest surveys (rocky-plage, baby-plage-ii) also had negative survey results for FHP in the most recent sampling period. Consequently, the expected value of theta is lowest at these locations but very close to baby-plage and the expected value for the Petit Lac (figure 3).

The probability of FHP in Saint Sulpice is greatest when just the results from the Grand Lac are considered. In Geneva the expected value is less than the lake when just the results from Petit Lac are considered. This confirms that the probability of finding an FHP is lower in Geneva than the lake and most likely lowest on the left bank (figure 3). *Making the right bank of the Petit Lac in Geneva the best choice for finding an FHP with a mean value of 43% and most likely between 35% and 50% (figure 2).*

Lake Geneva is connected to the Mediterranean Sea by the Rhône River. The median beach litter survey value on Mediterranean Beaches was 294 p/100m {cite}`threshholdeu`. The city of Port Saint Louis is at the mouth of the Rhône where it discharges into the sea. Port Saint Louis benefits from regular monitoring using the same protocol and coding system as Switzerland {cite}`merter`. The MLE of thetafhp for the 14 surveys on plage-napolean since January 2019 is 9/14 or .64, well within the 94% HDI of the Grand Lac and Saint Sulpice. Suggesting that the occurrence of FHP on a beach in the Grand Lac is just as likely as a beach on the Mediterranean in the region of Port Saint Louis.

It can be inferred that there is a real chance of finding an FHP anywhere along the Rhône between Geneva and Port Saint Louis. Furthermore, the rate at which these objects occur is not equal. This presents real challenges for stakeholders with respect to allocation of resources and differentiating between initiatives that produce the most cost-effective solution with respect to the expected outcome and those that do not.

ENVIRONMENTAL POLICY IMPLICATIONS

The challenge for producers is how to account for this in the LCA model. The beach is obviously one possible outcome amongst many. However, changes in operations or communications require resources and, in this case, it is unclear how any investment by a producer can prevent the consumer from flushing the object down the toilette. Changing the material of tampon applicators to paper does not remove them from the environment nor does it prevent them being flushed along with the tampon.

Fortunately, the process of collecting samples is an excellent method for stakeholders to build consensus. Beach litter surveys and the protocol have been in service for over 12 years, data collection is often assured by the end user. If a beach-litter survey is considered as a detailed customer feedback form, then we have presented a method to quantify that feedback with respect to the experience of the consumer. Many companies have used this formula to improve product quality and customer satisfaction.

That these results correspond with the experience of the surveyors follows from the math. The derivation of the Beta-Binomial conjugate model is more complex than its proposition: *each survey within a region adds to the cumulative knowledge of that region and the locations that were surveyed*. This defines the benefit both of collecting data in the field and reporting the results as a likelihood or expectation of what a person may find. This often-overlooked advantage imparts three critical pieces of information to the decision maker:

1. What the status was
2. What it would most likely be today
3. The source of the information

If the survey results are considered reliable, stakeholders have a method to anticipate the user-experience and enact policies to improve that experience. This places the assessment of quality and satisfaction into the hands of the end-user. For producers of goods that appear on beach-litter surveys there is now a method to determine how likely a product will end up on a survey. This gives another metric to determine the accuracy of the Lifecycle Assessment andimprove product outcomes with respect to end-of-life.

## Acknowledgements

**Data collection:** Association Pour la Sauvegarde du Léman, Precious Plastic Léman, Students of Solid Waste Engineering (École Polytechnique Fédérale), World Wildlife Fund Switzerland, Hackuarium and hammerdirt.

Special thanks to the Municipal Waste section of the Swiss Federal Office of the Environment for funding the IQAASL project {cite}`iqaasl`. Without their support in 2020-2021 we would have completed far fewer surveys.

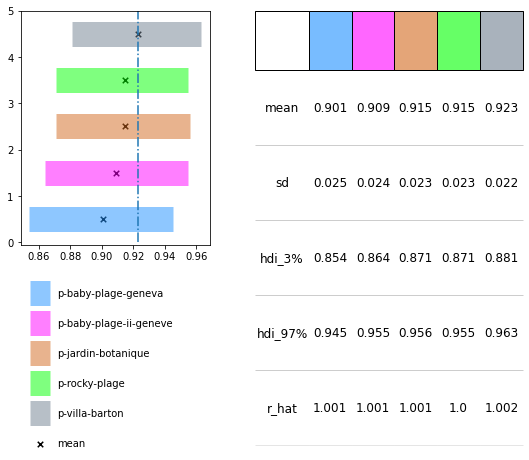
## Supporting Material

***Table 1a.*** *locations, number of samples and number of samples with at least one FHP on Lake Geneva.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| City | Beach | Samples | Positive samples | Rate |
| Genève | baby-plage-geneva | 11 | 3 | 0.27 |
| Genève | jardin-botanique | 3 | 2 | 0.66 |
| Genève | villa-barton | 11 | 1 | 0.09 |
| Gland | lacleman\_gland\_kubela | 11 | 3 | 0.27 |
| Gland | lacleman\_gland\_lecoanets | 11 | 4 | 0.36 |
| La Tour-de-Peilz | bain-des-dames | 4 | 3 | 0.75 |
| La Tour-de-Peilz | maladaire | 15 | 3 | 0.20 |
| La Tour-de-Peilz | oyonne | 3 | 2 | 0.66 |
| Lausanne | lacleman\_vidy\_santie | 12 | 10 | 0.83 |
| Lausanne | vidy-ruines | 7 | 2 | 0.28 |
| Montreux | baye-de-clarens | 13 | 1 | 0.07 |
| Montreux | baye-de-montreux-d | 16 | 5 | 0.31 |
| Montreux | baye-de-montreux-g | 18 | 2 | 0.11 |
| Montreux | le-pierrier | 5 | 3 | 0.60 |
| Préverenges | preverenges | 14 | 9 | 0.64 |
| Saint-Gingolph | grand-clos | 13 | 9 | 0.69 |
| Saint-Sulpice) | parc-des-pierrettes | 4 | 4 | 1.0 |
| Saint-Sulpice | plage-de-st-sulpice | 4 | 2 | 0.5 |
| Saint-Sulpice ) | tiger-duck-beach | 5 | 3 | 0.6 |
| Versoix | versoix | 4 | 1 | 0.25 |
| Vevey | arabie | 5 | 2 | 0.40 |
| Vevey | quai-maria-belgia | 22 | 13 | 0.59 |
| Vevey | veveyse | 17 | 6 | 0.35 |

***Table 2a*** *locations, number of samples and number of samples with at least one FHP in Geneva for each sampling period.*

|  |  |  |  |
| --- | --- | --- | --- |
| Beach | Period | Samples | Positive samples |
| baby-plage-geneva | 3 | 11 | 3 |
| baby-plage-ii-geneve | 3 | 2 | 0 |
| jardin-botanique | 1 | 3 | 2 |
| rocky-plage | 3 | 1 | 0 |
| villa-barton | 1 | 9 | 0 |
| villa-barton | 3 | 2 | 1 |

**Figure 1a.** The probability of finding at least one cotton-swab at a beach-litter survey in Geneva.

***Figure 2a*** *the probability of finding an FHP at any of the previous survey locations on Lake Geneva.*