# Initial data analysis: Diclofenac Measurements Rhine Koblenz

## Content

# 1 Discussion of data

Figure 1 shows the outcomes of the weekly non-targeted LC-MS measurements for diclofenac (DCF) plotted against time. Outcomes of both, the positive and negative ionization mode were plotted. The plot shows a strong correlation between amounts detected in positive and negative mode. Given this strong correlation results from the positive and negative ionization mode might be used for a mutual plausibility check. Based on such a plausibility check, the unusual high concentration observed in the negative mode measurement in September 2020 was excluded as an outlier from the further analysis.

With few exceptions, the amount of DCF detected in the positive mode is higher compared to what has been found in the negative mode, possibly due to differences in the ionization efficiency of diclofenac in the positive and negative mode, respectively. Moreover, data gaps are less abundant in the positive mode data. When looking at the locations of data gaps, it gets evident that they seem to be more likely during spring and summer months and that they are typically preceded by a period with a negative trend in DCF concentrations and followed by a period with a positive trend in DCF concentrations. Based on this it seems reasonable to assume that data gaps are due to DCF concentrations dropping below the detection limit of the respective analysis and that the actual concentration of DCF in the sample lies somewhere between 0 and the detection limit. Depending on the further statistical analysis such data gaps are of concern. For example, not having equidistant time intervals in a time series analysis may lead to a dilution of seasonal effects and/or trends. Data gaps in the positive ionization mode measurements were therefore filled with a random number between 0 and the detection limit, with the detection limit being estimated by the mean of all data points preceding or following a data gap.

Having the data gaps in the positive ionization mode data filled and plotted against an equidistant time axis, a clear seasonal pattern with lower DCF concentrations in spring and summer months becomes apparent (see Figure 2). Decreased concentrations during these months, on the one hand might be due to a decreased input of DCF in the aquatic ecosystem of the Rhine during these months and on the other hand due to an increase in biotic and abiotic processes leading to a degradation of DCF. In fact, DCF is not stable in aquatic environments for a long time and known to undergo rapid photolysis, giving rise to its transformation products (Bartels & von Tümpling, 2007). Seasonal plots (Figure 2 and Figure 3) also show clusters of variance in the data hierarchy, with the variance in measurements typically being higher in autumn and winter weeks. The same gets visible in a box plot of the raw data grouped by month (see Figure 4), with an increase in outliers during autumn and winter months and interquartile ranges typically being larger compared to summer and spring months. In the boxplot with the raw data being grouped by year a decrease of the median DCF concentrations and the variances of DCF concentrations over the years is indicated (see Figure 6).

# 2 Results

To assess whether the decrease in average DCF concentrations is statistically significant an ANOVA analysis followed by a multiple comparison post hoc-test was used. In its most simple version a one-way ANOVA that models Peak area as a function of the year was used. Results of such a one-way ANOVA for positive ionization mode measurements with gaps filled as described above are shown in Figure 7. The p value of the year value is low ($p < 0.001$), and the null hypothesis of average yearly concentrations being equal can be clearly rejected. It appears that mean concentrations indeed change over the years. Results of a TukeyHSD post-hoc multiple comparison test (see Figure 8) further indicate that the mean concentrations in year 2017 is statistically significantly ($p < 0.05$) lower compared to the mean concentration in year 2023. In fact, significant groupwise differences are indicated in Figure 8 anywhere the 95% confidence interval doesn't include zero.

Seasonal plots and box plots (Figure 2, Figure 3, Figure 4, Figure 6) show cluster of variance in the data hierarchy, with variance typically being higher in autumn and winter weeks. This elevated variance in the data that affects certain months/calender weeks (Figure 2, Figure 3, Figure 4) and high residual variances in linear and multiple regression models indicates the introduction of further covariates to control for variation in weather conditions, thereby reducing unexplained variance in models further and allow for a more in-depth analysis regarding the cause of the decrease in DCF concentrations over the years. Furthermore, the elevated variance in the data that affects certain months/calender weeks indicates a more in-depth analysis with a mixed effect model to further reduce unexplained variance in the data. A forecast via time series analysis (see Figure 5) can be leveraged as an early warning system for unexpected high DCF concentrations.

# 3 Plots & Statistical analyses

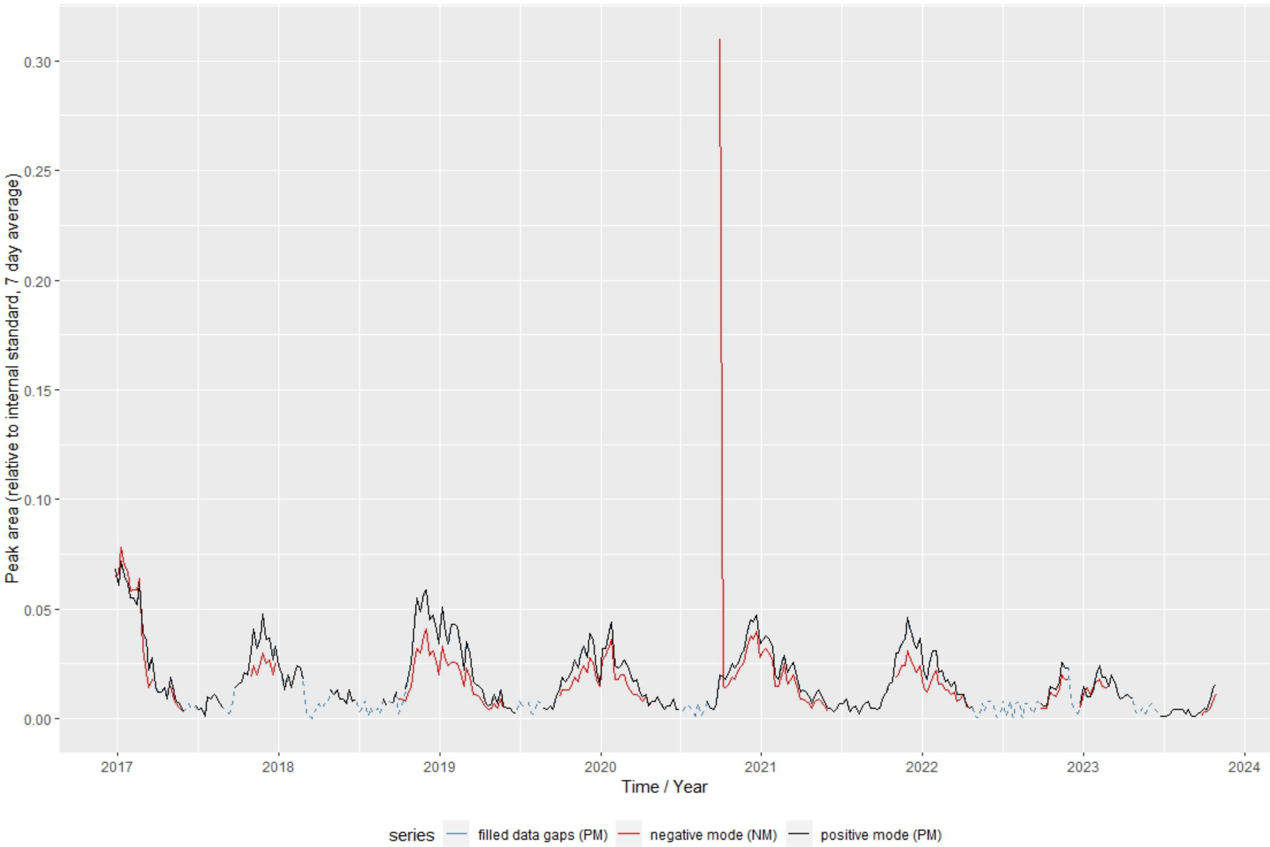


**Figure 1: Plot of weekly relative DCF peak areas observed in positive and negative ionization mode with daily composite samples taken from the left bank of the river Rhine in Koblenz**

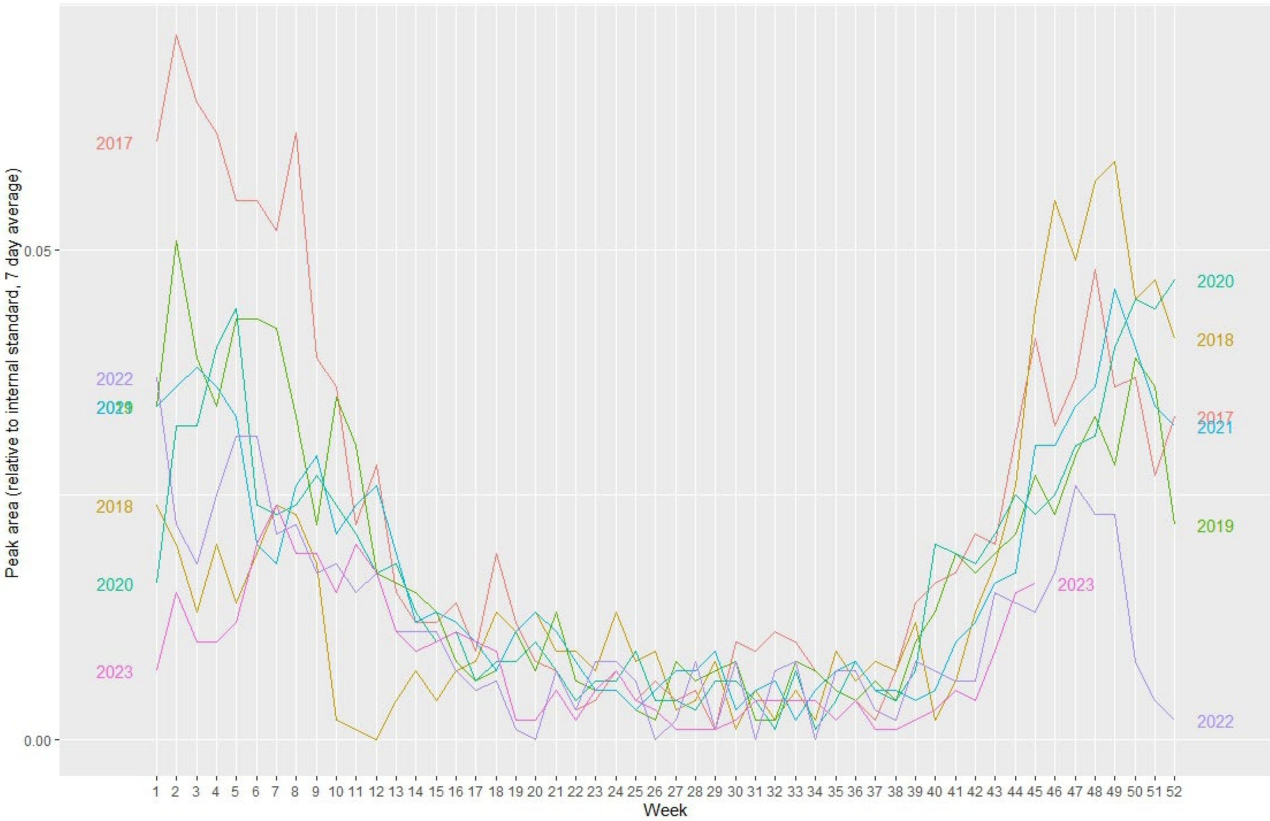


**Figure 2: Seasonal plot of diclofenac measurements taken in the positive ionization mode (Rhine Koblenz). Calendar weeks without data were filled with numbers generated via a stochastic process that chose a random number between 0 and the estimated detection limit.**

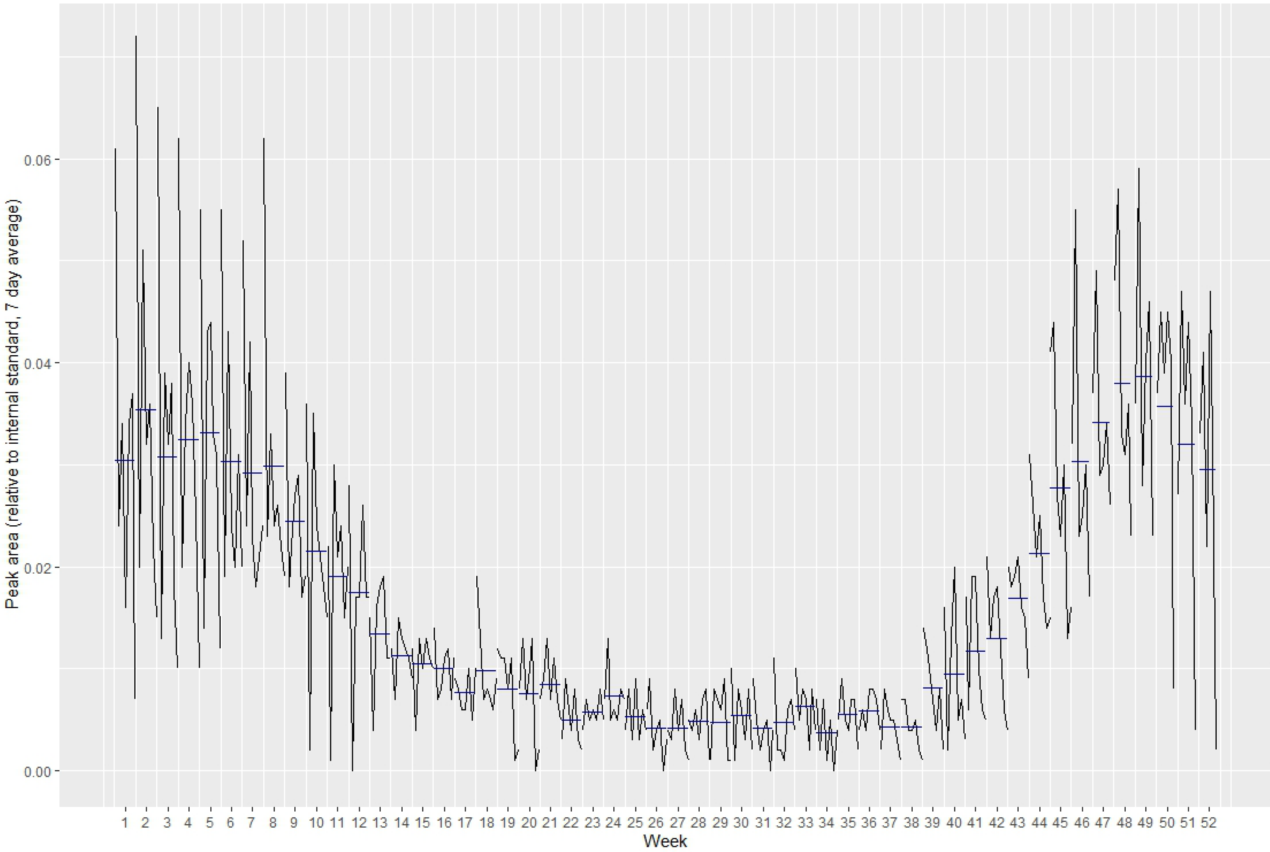


**Figure 3: Seasonal plot of diclofenac measurements taken in the positive ionization mode (Rhine Koblenz) independent of their year of observation. Calendar weeks without data were filled with numbers generated via a stochastic process that chose a random number between 0 and the estimated detection limit.**

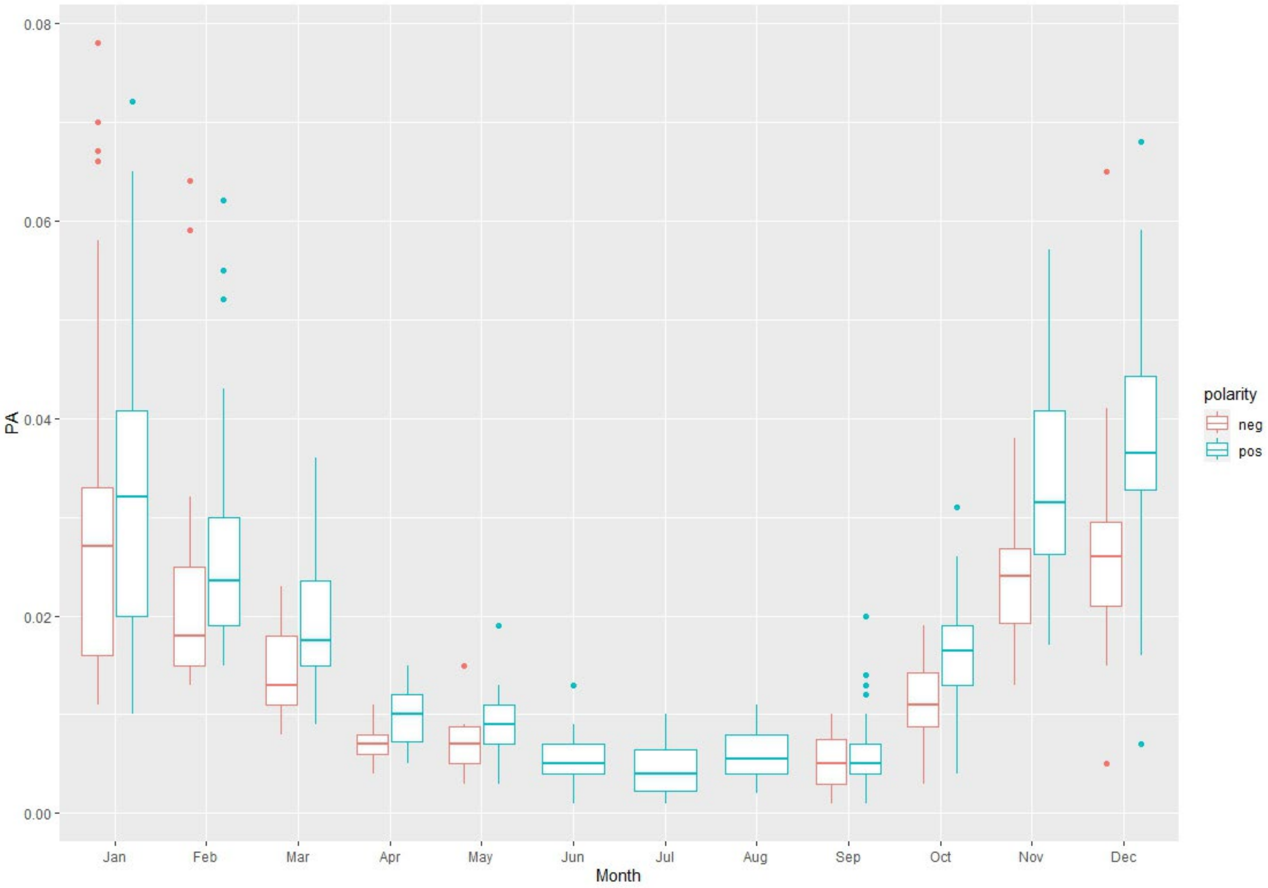


**Figure 4: Boxplot of the relative DCF peak area observed in positive and negative ionization mode (with the outlier in the negative mode from Sep 2020 excluded) grouped by month (before fill of data gaps)**

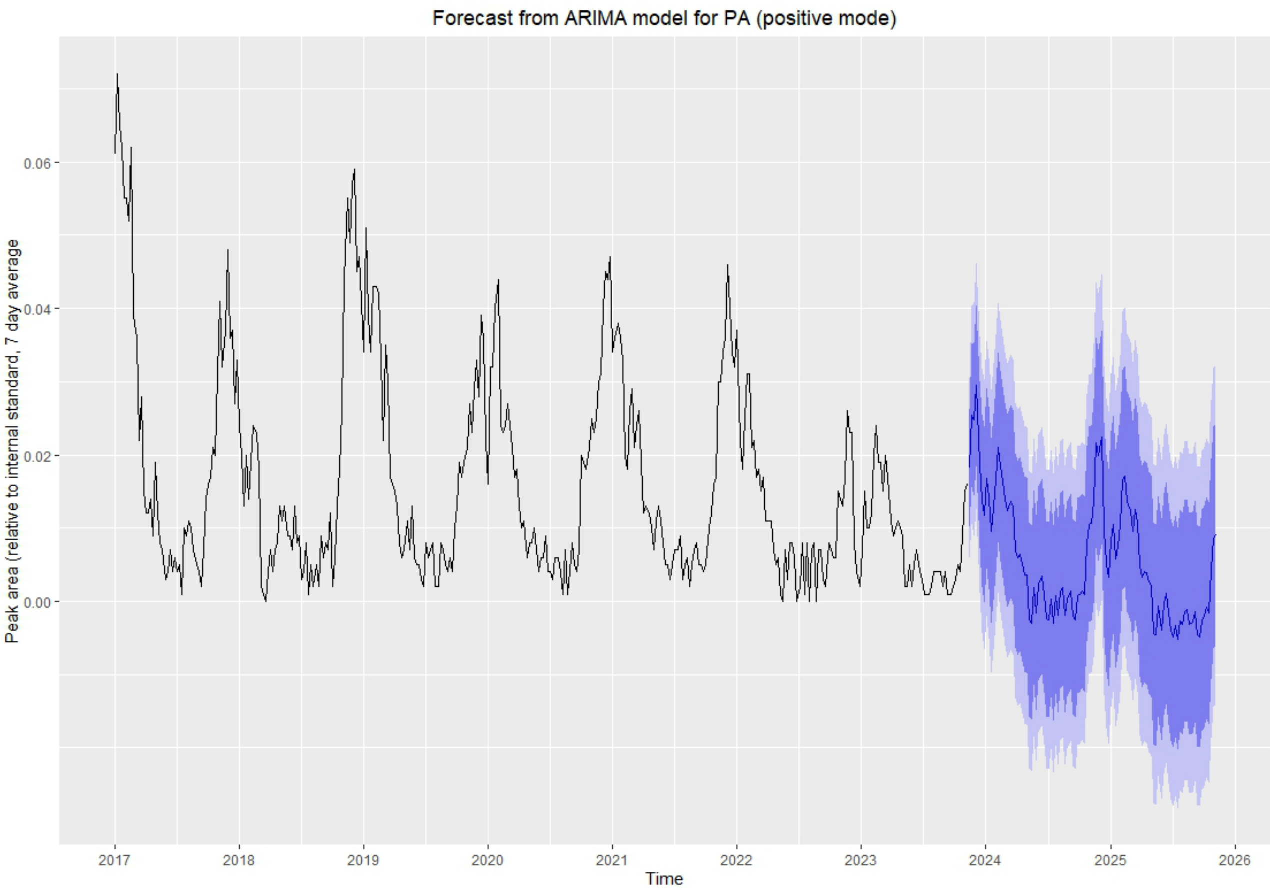


**Figure 5: Two year forecast from ARIMA model for relative peak area for measurements made in the positive mode (blue line) with the 80%- and 95%-confidence interval, being shown in dark and bright blue, respectively.**

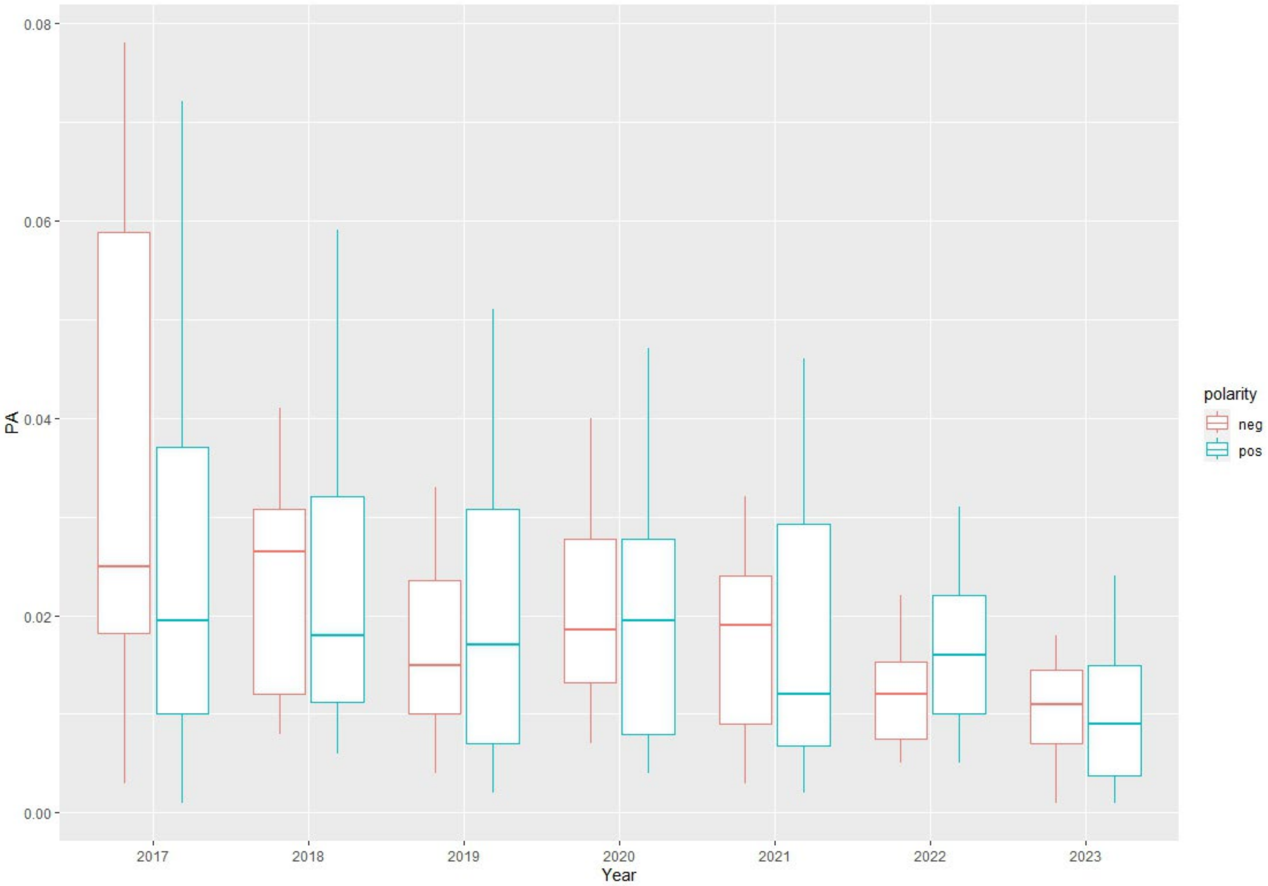


**Figure 6: Boxplot of the relative DCF peak area observed in positive and negative ionization mode (with the outlier in the negative mode from Sep 2020 excluded) grouped by year (before fill of data gaps)**

```
             Df  Sum Sq   Mean Sq F value    Pr(>F)    
Year          6 0.00762 0.0012698   6.798 7.86e-07 ***
Residuals   349 0.06519 0.0001868
```

**Figure 7: Results of one-way ANOVA for positive ionization measurements**

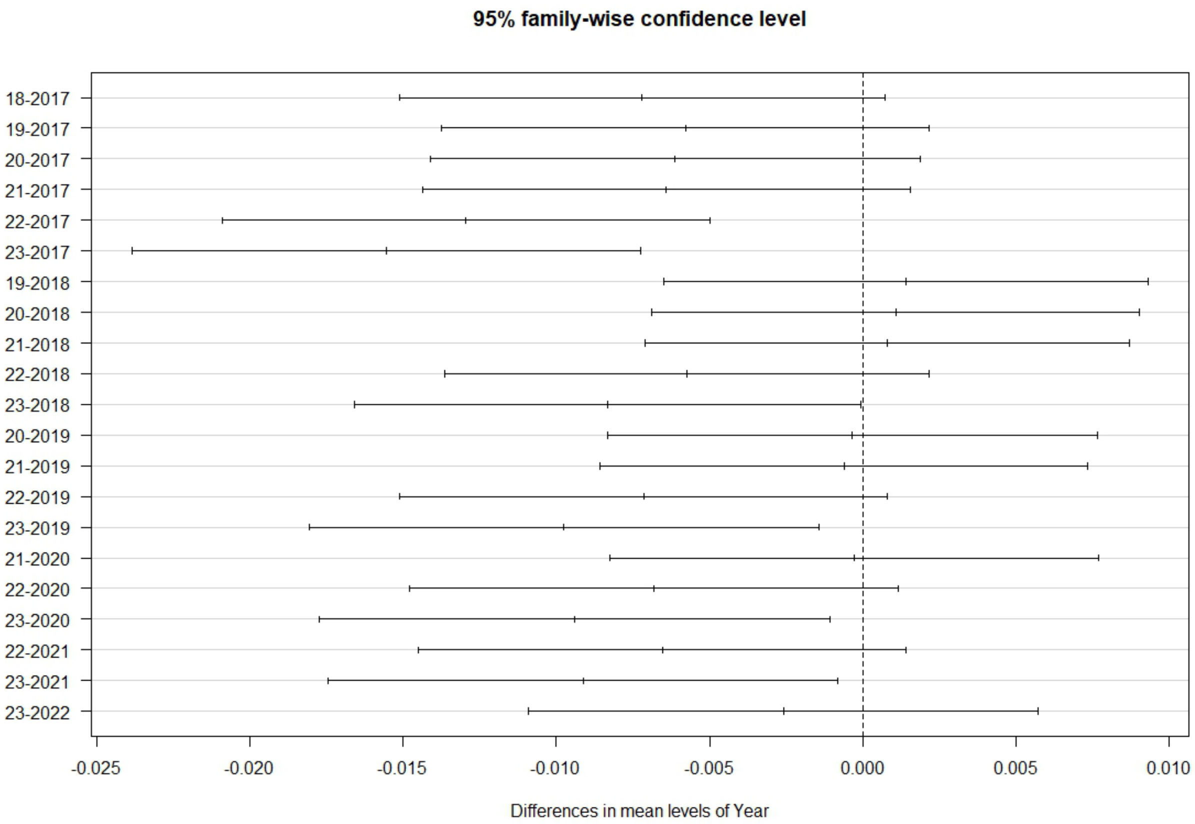


**Figure 8: Outcome of Tukey HSD multiple comparison test**

# 4 Contact

MM Toxicology Consulting & Scientific Support
Oberwilerstrasse 78, 4054 Basel, Switzerland;
Email: info@medicalmatters.eu
Homepage: medicalmatters.eu
Telephone: +41 21 515 45 23

# 5 References

Bartels, P., & von Tümpling, W. (2007). Solar radiation influence on the decomposition process of diclofenac in surface waters. *Science of The Total Environment*, *374*(1), 143–155. https://doi.org/10.1016/J.SCITOTENV.2006.11.039