Calculation of annual 90th percentile on time-series.

Case of nitrate concentrations in surface water

Mariangela Donati1, Chantal de Fouquet2, Roberto Bruno1, Gaëlle Deronzier3 & Katell Petit4

1 *School of Engineering and Architecture of the University of Bologne, viale Risorgimento 2, 40136 Bologne, Italy.*

*mariangela.donati@studio.unibo.it, roberto.bruno@unibo.it*

2 *École des Mines de Paris- ARMINES, centre de géosciences - géostatistique, 35 rue Saint-Honoré - F77305 Fontainebleau Cedex, France.*

*chantal.de\_fouquet@mines-paristech.fr*

3 *Office National de l’Eau et des Milieux Aquatiques – Onema, 5 square Félix Nadar - 94300 Vincennes, France.*

*gaelle.deronzier@onema.fr*

4 *Office International de l’Eau – OIEau, 15 rue Edouard Chamberland – 87065 Limoges Cedex, France.*

*k.petit@oieau.fr*

**Extended abstract**

After the degradation of water quality due to nitrates detected at the end of the 20th century, the European Commission adopted the directive 91/676/EEC (*Nitrates Directive*). This Directive requires Member States to deliver to the European Commission a report, at least every four years, containing among other things the results of monitoring campaigns of nitrate concentrations in surface water.

In France the report contains the values of minimum and maximum annual concentrations, annual mean, winter mean and 90th percentile of concentrations, calculated at different scales (per station, per hydrographic basin and at national level).

In this work we focused on the annual 90th percentile. The principle is to select the value below which 90 percent of the measurement time-series may be found. This method is chosen in order to avoid taking into account exceptional situations values. How can the calculation of the 90th percentile be improved? How to compare two percentiles in time if the sampling dates vary between years?

For the study are used time-series of nitrate concentrations registered in 121 monitoring stations located in five different hydrographic basins during about 40 years.

The frequency of measurements varies between years and stations, but in most cases the annual number of sampling is equal to twelve. The concentrations present generally a seasonal variability and temporal correlation.

The 90th percentile is estimated by the usual statistical way, using empirical calculation, according to the method initially recommended at a national-level by the “SEQ-Eau” (System of evaluation of the river quality), and then in the *Technical assessment guide of the water state* (2012) meeting the criteria of the *Water Framework Directive*.

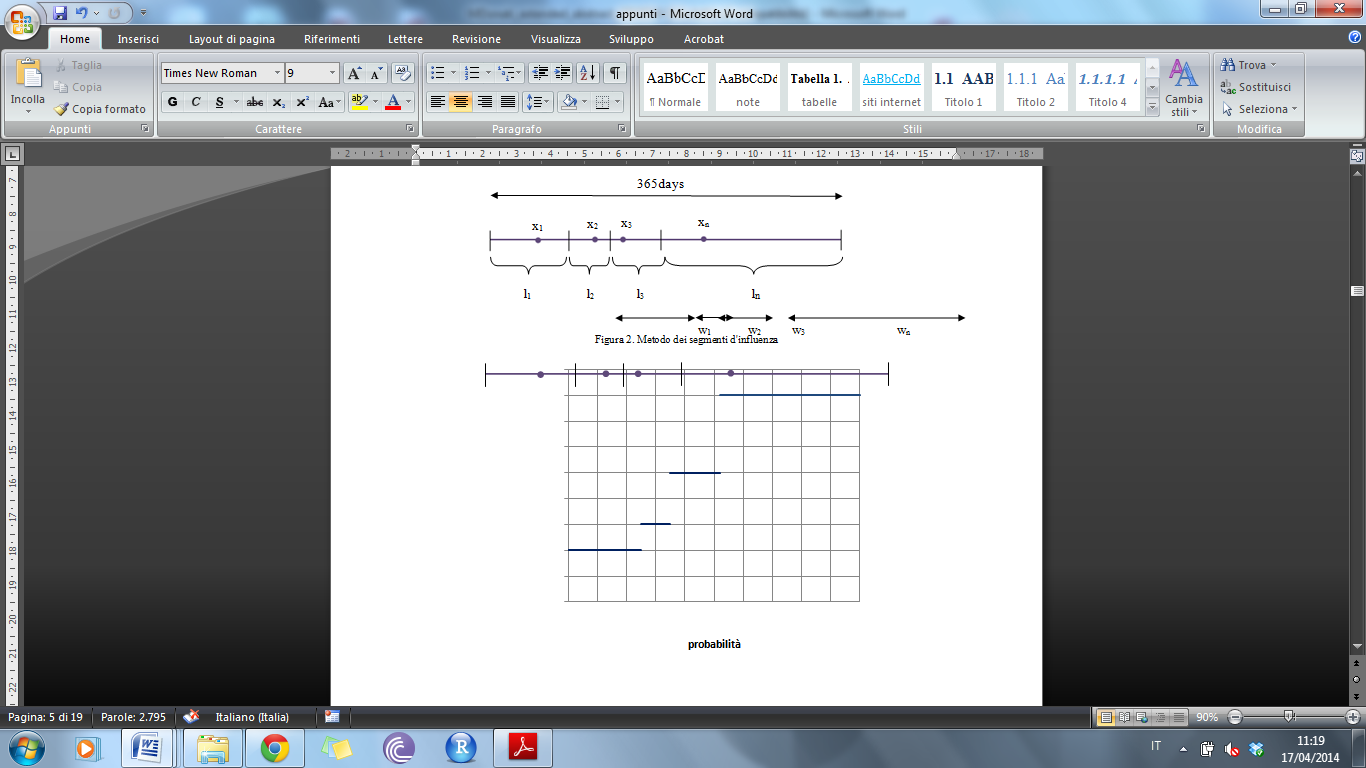
This method is only based on measured concentrations, but for nitrate concentrations, the necessary hypotheses of independence and stationarity of data are not verified.

Following the pioneering work of Caroline Bernard-Michel (2006), we first confront the empirical calculation with different methods, in order to facilitate comparisons between different campaigns:

* weighting of data to take into account the irregularity of dates;
* linearization of percentile function to mitigate the discontinuities of the empirical percentile function and.

The data are weighted using a simple geometric method, the segment of influence, (*weighted percentile*). It consists in assigning to every measure *xi* an "influence zone" equal to the segment bounded by the bisectors of the segment between two measures subsequent in time scale, *li*.

The weight of each data is the ratio between segment of influence and length of the "reference period" (one, two or three years).



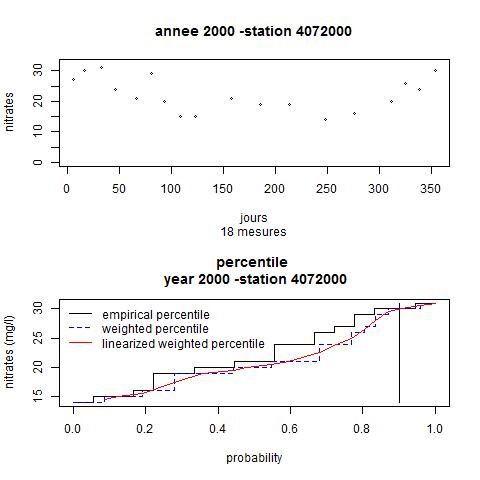
**Figure 1. Method of segment of influence.**

Subsequently we replace the weighted percentile function by a piecewise linear function *(linearized weighted percentile*), passing through the center of each step, following the definition of Dacunha-Castelle and Duflo (1994).

Moreover it is suggested to avoid calculation on time-series with gaps, in order to consider the seasonality of nitrate concentrations.

The results of the three methods (empirical percentile, weighted percentile and linearized weighted percentile) are compared. Weighting and linearization can significantly modify the results of the estimation, but the entity and the sign of the changes are not systematic.

The same is observed for the inter-annual variations.



**Figure 2. Example of the three estimation methods of the 90th percentile for the station 4072000 in 2000.**

To increase the number of measurements for the estimation, at the same frequency of measurement, we increase the "reference period" (from one to two or three years). The "reference period" is the same for the data, for the estimation of the percentile and for his variations.

It is observed that the results are more influenced by the "reference period" considered rather than the estimation method. In fact the increase of "reference period" greatly improves the conditions of estimation. The results can't be related to the single year anymore, but throughout the whole period considered.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Calculation method p90** | **Min** | **Q25** | **Median** | **Mean** | **Q75** | **Max** | **Std. Dev.** |
| **2 y (121 st)**  **[2010, 2011]** | empirical | 3.00 | 11.00 | 17.00 | 20.12 | 27.80 | 59.90 | 11.57 |
| linearized | 2.88 | 11.59 | 18.89 | 20.53 | 28.31 | 59.70 | 11.52 |
| **1 y (118 st) 2010** | empirical | 3.10 | 11.55 | 17.75 | 21.11 | 28.67 | 59.90 | 12.08 |
| linearized | 2.98 | 12.47 | 19.72 | 21.65 | 30.13 | 58.46 | 12.04 |

**Tab 1. 90th percentile of nitrate concentrations expressed in mg/l (empirical or linearized weighted), calculated with data of the year N or of the years N and N+1. y=years, st=station. Statistics of the results relative to all the stations which have at least 7 measures in the "reference period".**

Finally we validate the methods by a comparison with reality, carried out on complete time-series of nitrates. The 90th empirical percentile and the 90th weighted and linearized percentile are calculated with all the data of one or two years. The percentiles with many different selections of dates of measurements are estimated with different density of sampling (12 or 24 measures per year).

Weighting and linearization substantially improve the results of the annual percentile in case of 12 measures per year. With 24 measures (for a "reference period" of one or two years) the improvement is not so systematic.

In the cases examined, 24 measures seem to be adequate for a sufficiently precise estimation of the percentile for nitrate concentrations.

**Keywords**

Environment, water quality, nitrate concentration, annual percentile, irregular sampling, annual variations.

**References**

[1] MEDDE. 2012. Guide technique Évaluation de l’état des eaux de surface continentales (cours d’eau, canaux, plans d’eau).

[2] C. Bernard-Michel. 2006. Indicateurs géostatistiques de la pollution dans les cours d’eau. Doctorate in Geostatistics, École Nationale Supérieure des Mines de Paris.

[3] Dacunha-Castelle, Duflo. 1994. Probabilités et Statistiques. Tome 1. Problèmes à temps fixe. Éditions Masson.