

Directional outlyingness for the detection of functional outliers in water quality data

X. Rigueira ⁽¹⁾, J. Martínez ⁽²⁾, M. Araújo ⁽¹⁾, E. Giráldez ⁽¹⁾

⁽¹⁾ CINTECX, GESSMin group, Department of Natural Resources and Environmental Engineering, University of Vigo, 36310, Vigo, Spain

Principal Author's Phone Number and e-mail: +34 698135727, xurxo.rigueira@uvigo.es

⁽²⁾ CINTECX, GESSMin group, Department of Applied Mathematics, University of Vigo, 36310 Vigo, Spain

1. Introduction – The importance of water for human health and development is undeniable. Consequently, national and local governments are passing stricter regulations to ensure the quality and availability of this natural resource. In this scenario, the live feed of data from several water quality physical and chemical parameters is involved, and the detection of anomalous values —outliers— in these time series becomes of major importance in order to respect the limits of those regulations. Usually, the detection of these values has been carried out with classical statistical methods that study the data as discrete points and fail to evaluate its trend leading to poor results. This study, financed by the Spanish Ministry of Science under grant PID2020-116013RB-I00, proposes a new method for the detection of outliers in water data. Under a mathematical scope, functional data analysis [1] has the advantage of taking into account the time dependency of water quality data because it considers this information as a set of continuous curves, instead of discrete points. More specifically, this research introduces the concept of directional outlyingness proposed by Dai *et al.* [2], which allows the study of the shape and the shift of the curves compared to the rest of the set. This mathematical methodology, alongside with a new outlier detector were applied to a water quality database from the Ebro Basin in Spain. The dataset contains values of ammonium, conductivity, nitrates, dissolved oxygen, pH, temperature, and turbidity, taken every 15 minutes from January 1st, 2019 to December 31st, 2021.

2. Experimental – The water quality data is analysed from a functional point of view, which involves the transformation of the discrete points into a set of functions over a continuum. Moreover, directional outlyingness is implemented with the modified band depth [3], and the time unit selected was weeks as it increases the number of functions compared to months and years. The data processing is performed with Python 3.7 and the library for functional data analysis `scikit-fda`.

3. Results and Discussion – The method presented is able to perform a precise detection of outliers. It does not only consider as abnormal values those functions that present a certain deviation from the rest, but also includes those curves which display an atypical behaviour despite being within the average range of values. The detection of outliers in each of the water quality variables studied enables the analysis of the hidden relations among them, and the trace back of these anomalous events, whether their origin is anthropic or natural.

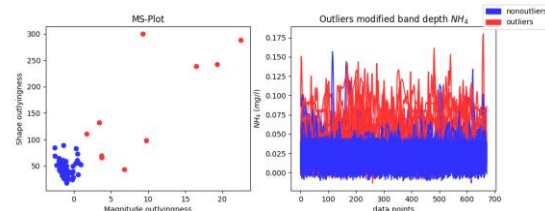


Image 1. Results of the functional analysis for the detection of ammonium outliers.

4. Conclusions – The challenge presented in the detection of outliers in water quality data is addressed from a functional approach. The method implemented and the new outlier detector achieve great precision in the identification of those time frames that contain anomalous values. This allows the automatic detection of outliers in this type of data and facilitates the process of selecting corrective measures to remove those issues causing the atypical records.

5. References

- [1] Ramsay, J. O., & Silverman, B. W. (2002). *Functional Data Analysis* (1st ed.). Springer International Publishing.
- [2] Dai, W., & Genton, M. G. (2018). Multivariate Functional Data Visualization and Outlier Detection. *Journal of Computational and Graphical Statistics*, 27(4), 923–934. <https://doi.org/10.1080/10618600.2018.1473781>
- [3] Lopez-Pintado, S., & Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486), 718–734. <https://doi.org/10.1198/jasa.2009.0108>