Drink. Water Eng. Sci. Discuss.,
https://doi.org/10.5194/dwes-2017-19-AC3, 2017

Open Access
Drinking Water
Engineering and Science
Discussions

Interactive comment on "Modeling and Clustering Water Demand Patterns from Real-World Smart Meter Data" *by* Nicolas Cheifetz et al.

Nicolas Cheifetz et al.

nicolas.cheifetz@veolia.com

Received and published: 2 June 2017

Dear Referee,

First, we would like to thank you for taking time to review this paper and providing us constructive comments and suggestions. The major comments concern the method usability and clustering interpretation.

Identifying the major usage profiles from water consumption is an interesting topic to water utilities. Indeed, the resulting segmentation helps the water companies to gain a better knowledge about users consuming the distributed water. The user is having a better experience with the tools developed by their water utility. For instance, users at

Veolia Eau d'Ile de France (Paris area in France) can already monitor their water index / consumption on a dedicated website for free. Based on our clustering results, people could compare with similar patterns and adapt their consumptions according to their needs. In addition, customer services might alert the user if a leakage is detected. Concerning the grid management, each prototype can be used to represent the water behavior of users belonging to the same cluster. An erratic water pattern (like in cluster 2) can be the sign of a leakage and might initiate a corrective action. Sampling a large amount of water meters is useful for several topics (e.g. tracking the meter metrology, estimating the global consumption modes based on a limited number of meters) ; such sampling analysis is straightforward using our meter segmentation.

About the clustering interpretation and validation, we assume no supervision in our setting due to a partial and uncertain knowledge of the usage labels ; users do not inform systematically their water utilities when businesses change or people come in / leave a home. This explains why there is no quantitative accuracy about clustering ; on the other hand, a qualitative validation is exposed using water profiles for each cluster. Each log-consumption time series is standardized before clustering which leads to a discrimination in term of seasonal patterns and not based on water volume. This explains why we entitled the cluster 1 as "'office and industrial usage". Of course, industrial usage might produce erratic water patterns which would be classified in cluster 2. We would like to thank the referee for pointing out some interesting literature about the analysis of specific demand patterns (residential or non-residential) ; this might be used in a future work if the water time series share a similar granularity. Nevertheless, we believe the proposed approach is relatively robust: similar results were obtained by running our methodology on sub-samples of thousand meters. The risks might be the non representativity of the clusters (incorrect model selection with an inappropriate K) and the loss of information by decomposing the water demand time series into a single weekly pattern (e.g. it may mask dynamical changes for a meter use).

Water demand forecasting is not the issue in this paper ; nevertheless, the resulting segmentation of water consumption time series can be used for several scientific problems including sequential detection, predictive classification or demand forecasting. Exogenous variables (e.g. weather inputs or meter localization) are not considered in this work due to a non-significant improvement in the results but might be used in a future work. The "climate change" is only mentioned to emphasize the systemic changes inherent in any smart city. In this article, a double seasonality is taken into account: daily and weekly. This prior knowledge is included both in the modeling of the Fourier-based decomposition and the prototype functions of the Fourier Regression mixture (FReMix) model. The time series x_i have a length of 168 due to the trigonometric modeling of the chosen Fourier basis decomposition. The Fourier coefficients are identified performing a multiple linear regression on the global time series y_i detrended (by subtracting a moving average of order 168) which limits the effect of vacation periods or a long seasonality.

Interactive comment on Drink. Water Eng. Sci. Discuss., https://doi.org/10.5194/dwes-2017-19, 2017.