Drink. Water Eng. Sci. Discuss.,
https://doi.org/10.5194/dwes-2017-19-AC1, 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 design of the problem addressed by the article and the justification of the methodological choices.

This paper deals with an unsupervised classification problem based on water consumption time series. In other words, the problem is to affect a categorical label representing a mode of consumption to each water meter. Water demand forecasting is not the issue in this paper ; nevertheless, the resulting segmentation of water con-

sumption time series can be used for several scientific problems including sequential detection, predictive classification or demand forecasting. 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. As the number of clusters is not known, we used a statistical criterion to help us adjusting the model selection but this BIC criterion deceases continuously due to the large amount of data ; a way around this problem would be to penalize more the log-likelihood but it would not be a BIC criterion anymore. Partitioning the 8 clusters in 4 categories would suggest non negligible variations in residential use as well as commercial use, and extra investigations about the users which should not be underestimated in term of time and cost. Finally, the EM algorithm used to fit the Fourier REgression Mixture (FReMix) model is flexible and can be reformulated in a future work with a semi-supervision (by fixing a set of posterior probabilities) or a partial supervision (e.g. using belief functions).

We chose a functional clustering scheme because this is suitable to deal with the analysis of our consumption curves. Indeed, these real-valued data can be seen as the realizations of a one-dimensional stochastic process, recorded on the same time grid (hourly spaced) of ordered times. In practice, data frames are frequently sent by modules which are physically connected to the meters ; each consumption time series can be re-constructed based on a sequence of the data frames. 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.

However, we wanted to integrate some prior knowledge about day/week seasonality and exceptional public non-working days. A Fourier-based decomposition has the capacity to easily take into account this prior knowledge and this decomposition is consistent with our probabilistic FReMix model definition. A wavelet-based analysis could also be used for decomposition (keeping some local properties along temporal patterns) but integrating such prior knowledge might not be straightforward and the number of parameters in this case should be not be reduced significantly.

In this article, we evaluate a probabilistic method and a geometrical approach. This second method is based on a Kmeans and minimizes the intra-cluster inertia which can be seen as an aggregated distance over the water time series. As our knowledge, the complexity of time series distances (e.g. dynamic time warping) can be prohibitive with a large time series dataset.

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