International Environmental Modelling and Software Society (iEMSs)
8th International Congress on Environmental Modelling and Software
Toulouse, France, Sabine Sauvage, Jos-Miguel Snchez-Prez, Andrea Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2016-proceedings

Profiling residential water users' routines by eigenbehavior modelling

A. Cominola[a], **A. Moro** [a], **L. Riva** [a], **M. Giuliani** [a], **A. Castelletti** [a]

[a] *Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Piazza L. da Vinci, 32, I-20133 Milano, Italy (andrea.cominola@polimi.it, matteo.giuliani@polimi.it, andrea.castelletti@polimi.it)*

Abstract: Developing effective demand-side management strategies is essential to meet future residential water demands, pursue water conservation, and reduce the costs for water utilities. Yet, the effectiveness of water demand management strategies relies on our understanding of water consumers' behavior and their consumption habits and routines, which can be monitored through the deployment of smart metering technologies and the adoption of data analytics and machine learning techniques. This work contributes a novel modeling procedure, based on a combination of clustering and principal component analysis, which allows performing water users' segmentation on the basis of their eigenbehaviors (i.e., recurrent water consumption behaviors) automatically identified from smart metered consumption data. The approach is tested against a dataset of smart metered water consumption data from 175 households in the municipality of Tegna (CH). Numerical results demonstrate the potential of the method for identifying typical profiles of water consumption, which constitutes essential information to support residential water demand management.

Keywords: user profiling; user segmentation; principal component analysis; water demand management; machine learning

1 INTRODUCTION

Urban population growth, combined with changing climate and society, is expected to boost residential water demand in urban contexts in the next decades. Projections show that the number of mega-cities, namely cities with more than 10 million inhabitants, is expected to grow over 40 by 2030 [UNDESA, 2014]. In such a context, developing suitable demand-side management strategies through the promotion of cost-effective water-efficient technologies, revised economic policies, appropriate national and local regulations, education, and social marketing is essential to meet future water demands, pursue water conservation, and reduce costs for water utilities [Gleick et al., 2003]. Yet, the effectiveness of water demand management strategies (WDMS) relies on our understanding of water users' behavior, their consumption habits, routines, and the drivers of their demand [Jorgensen et al., 2009].
While low spatial and temporal resolution water consumption data, as traditionally gathered for billing purposes through conventional water meters, hardly support this understanding, the advent of high-resolution, smart metering technologies allowed for quasi real-time monitoring water consumption at the single household level, providing instantaneous information to water utilities on the network status and continuously informing about users' consumption and savings behaviors [Boyle et al., 2013]. Smart metered data provide essential information for accurately modeling individual users' behaviors [for a review see Cominola et al., 2015a, and references therein], especially through the

application of data analytics and machine learning techniques [e.g., Cardell-Oliver, 2013; Cominola et al., 2015b]. Two distinctive approaches exist for modeling water users' consumption behaviors: *descriptive models*, which aim at performing users' segmentation through the analysis of observed water consumption patterns and historical trends [e.g., Beal et al., 2011; Beal and Stewart, 2014], and *predictive models*, which instead provide estimates of the expected water demands, possibly conditioned upon natural and socio-psychographic factors, or in response to alternative water demand management strategies [e.g., Maggioni, 2015; Makki et al., 2015].

In this paper, we contribute a novel descriptive modeling procedure for performing users' segmentation on the basis of smart metered water consumption data. Our procedure is based on a combination of clustering and principal components extracted from water demand data, extending the idea originally proposed by Eagle and Pentland [2009] for the identification of routines in the temporal location of 100 individuals from MIT, monitored using 100 Nokia 6600 smart phones. The extraction of principal components from behavioral data defines a set of vectors spanning the "behavioral space" of monitored individuals, characterizing their behavioral variation in time. These components, called *eigenbehaviors*, are computed as the eigenvectors of the covariance matrix of behavior data, where the vectors associated to high weights represent a type of recurrent behavior, i.e., a *routine*. In this work, this idea is extended for the identification of typical water consumption behaviors from a dataset of smart metered water consumption readings. The proposed approach is tested on a dataset of hourly-sampled water consumption records from 175 households in the municipality of Tegna (CH), which have been equipped with smart meters by Società Elettrica Sopracenerina as part of the SmartH2O Project.[1]

2 Methodology

For each smart metered household, we have a time series of water consumption readings sampled with hourly resolution, which can be organized in a $[D \times 24]$ individual water consumption matrix C^k (with $k = 1, \dots, U$, being U the total number of users/households), where each row corresponds to one day and each column to one hour of the day. We then transformed this matrix into a binary matrix Γ^k, where the observed values of hourly water consumption are classified into N mutually exclusive classes Λ (e.g., low, medium, high consumption) based on hourly consumption thresholds. Each row of Γ^k hence contains $24 * N$ elements, where the binary values in the i-th row of Γ^k associate the 24 consumption readings of the i-th day in C^k to one of the N consumption classes Λ. The generic element x of matrix Γ^k is hence defined as:

$$x(i, j + (n-1) \times 24) = \begin{cases} 1, & C^k(i,j) \in \Lambda_n \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $i = 1, \dots, D$; $j = 1, \dots, 24$; $n = 1, \dots, N$. For each user, we can identify its average individual behavior Ψ^k as well as the daily deviation $\Phi^k(i)$ from this average behavior, defined as

$$\begin{aligned} \Psi^k &= \frac{1}{D} \sum_{i=1}^{D} \Gamma^k(i) \\ \Phi^k(i) &= \Gamma^k(i) - \Psi^k \end{aligned} \tag{2}$$

[1] SmartH2O official website: www.smarth2o-fp7.eu

Finally, we can extract the eigenbehaviors by performing a Principal Component Analysis [Joliffe, 2002] on the resulting matrices Φ^k. PCA is a dimensionality reduction technique, which searches for linear combinations of the original variables such that the coefficients of the output combinations (the principal vectors) form a low-dimensional sub-space defined by directions explaining maximal variance in the original data. Few principal components explain a high percentage of the variance of the original variables, ensuring dimension reduction. In addition, the representation of the original data in the projected space defined by principal components is uncorrelated, thus providing a useful tool for physical and statistical interpretations.
PCA is performed via an eigenvalue decomposition of the covariance matrix R of Φ^k (which has D rows corresponding to D days and $Q = 24 * N$ columns corresponding to the binary labels classifying the hourly water consumption), i.e.

$$R = \frac{1}{D} \sum_{i=1}^{D} \Phi^k(i)^T \cdot \Phi^k(i) \tag{3}$$

where the resulting eigenvectors $\mathbf{w}_q^k$ (with $q = 1, \dots, Q$) are the eigenbehaviors of the k-th user and allow mapping the original matrix Φ^k into its principal components, i.e.

$$\varrho_q^k = \Phi^k \cdot \mathbf{w}_q^k \quad \text{with } q = 1, \dots, Q \tag{4}$$

After performing the dimensionality reduction step via PCA, we classify the metered users into different consumption profiles on the basis of their first eigenvector $\mathbf{w}_1^k$ (with $k = 1, \dots, U$). The first eigenvector accounts for as much as possible of the behavioral data variance of the considered user's consumption, which is quantified by the associated highest eigenvalue. Our classification was run using the K-means clustering method [MacQueen et al., 1967], a widely adopted technique that allows grouping multidimensional points in clusters by minimizing the average squared distance between points and centroid of each cluster. Formally, K-means algorithm partitions the U-dimensional set $\mathcal{W} = \{\mathbf{w}_1^1, \dots, \mathbf{w}_1^U\}$, containing the first eigenbehavior of each user, into $M < U$ profiles $\mathcal{P} = \{P_1, \dots, P_M\}$ by solving the following minimization problem:

$$\mathcal{P} = \arg\min_{\mathcal{P}} \sum_{m=1}^{M} \sum_{\mathbf{w}_1^k \in P_m} ||\mathbf{w}_1^k - \mu_m||^2 \tag{5}$$

where μ_m is the mean of $\mathbf{w}_1^k \in P_m$, with the resulting clusters satisfying the following conditions: (i) the union of all clusters contains all the original point, i.e. $\cup_{m=1}^{M} P_m = \mathcal{W}$; (ii) each point belongs to a single cluster, i.e. $P_i \cap P_j = \emptyset$; (iii) the clusters cannot be empty and a single cluster cannot include all the points, i.e. $\emptyset \subset P_m \subset \mathcal{W} \quad \forall m$.

3 Case Study

Our testing case study consists of real-world dataset of anonymized smart metered water consumption provided by Società Elettrica Sopracenerina (SES) as part of the SmartH2O project [see Rizzoli et al., 2014, for an overview about the project]. SES, a multi-utility based in Locarno (CH) installed 400 smart meters in the Locarno district during the first two years (2014-15) of the SmartH2O project.
For this study, we consider a dataset that, after some pre-processing needed for removing missing readings, measuring errors, etc., comprises 175 households monitored for around 7 months, from March, 30th to November, 7th, 2015. Each user is therefore associated to 5,352 hourly readings (223 days), with the full dataset including 936,600 data.

4 RESULTS

The first step of our methodology required to transform the hourly water consumption readings matrix of each user C^k (with $k = 1, \dots, 175$) into binary matrices Γ^k, where the consumption data are classified into mutually exclusive consumption classes (see Section 2). We initially observed that 46% of the data in our dataset are equal to 0. This can be easily explained by the fact that generally there is no consumption during night or when people are not at home (e.g., during working hours). Considering the distribution of the nonzero values (see Figure 1), we decided to partition these latter in three classes with respect to a threshold of 12 L/h, representative of low consumption levels (e.g., faucet, toilet), and a threshold of 100 L/h, which allows distinguishing medium consumption events (e.g., a 10-minute shower, efficient clothes washer programs) from high consumption ones (e.g., outdoor uses, inefficient devices). In summary, we classified the available hourly water consumption readings into the following four consumption classes, whose sizes are illustrated in Figure 2:

- **L0**: hourly consumption $x = 0$ L/h.
- **L1**: hourly consumption $x \in (0, 12]$ L/h.
- **L2**: hourly consumption $x \in (12, 100]$ L/h.
- **L3**: hourly consumption $x > 100$ L/h.

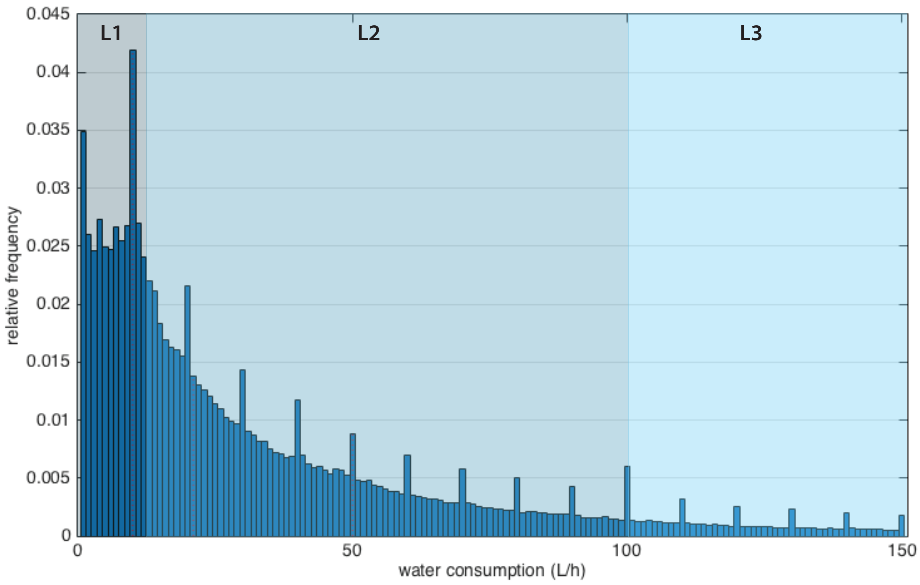

Figure 1. Empirical distribution of nonzero water consumption readings, with the shading representing the three selected classes L1, L2, L3.

According to the thresholds set for the above classification, the hourly water consumption matrices C^k are transformed into the corresponding binary matrices Γ^k, from which the eigenbehaviors of each user can be extracted via PCA. Results show that considering solely the first principal component for representing the behaviors of the 175 considered

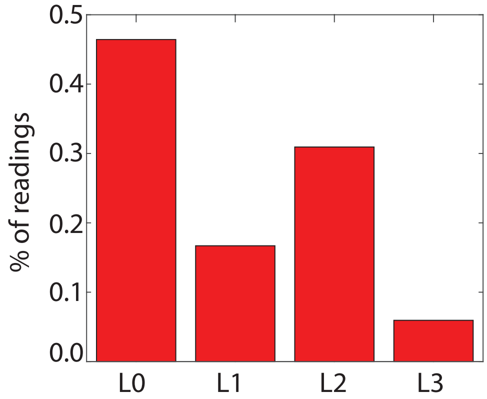

Figure 2. Partitions of hourly consumption reading data across the four consumption classes L0, L1, L2, L3.

users allows explaining, on average, 50% of the variance in the original data. The second principal component contributes an additional 8%, while 10 principal components would be necessary to explain 75% of the variance, with a saturation to around 100% obtained with 60 principal components. The large share of variance explained by the first eigenbehavior, along with the significant difference between the contribution of the first and second component, confirms that working on the first eigenbehavior is sufficient for capturing a large share of the variability in users' behaviors, thus supporting effective users' segmentation.

As a last step of our methodology, we applied K-means clustering on the set of first eigenbehavior of each user, where high values identify recurrent and relevant behaviors (e.g., frequent concentration of consumption events of a given class in specific hours, including periods with no consumption). K-means clustering would allow classifying users with similar consumption routines, represented by similar values of $\mathbf{w}_1^k$, in the same cluster, supporting effective users' segmentation. To demonstrate the validity of our procedure, we report the results obtained setting the number of clusters equal to three, thus simply distinguishing low, medium, and high consumption profiles (additional experiments with higher number of clusters are reported in Moro and Riva [2016]). Results are illustrated in Figure 3, where each column represents the first eigenbehavior of a user and each row corresponds to the hour of the day for each consumption class. The users' classified within profile P1 are characterized by high values in the first eigenbehavior only for L0, meaning they are usually not consuming: this may be the case of vacation houses or, when the eigenbehavior coefficient is positive in classes L1 or L2, they are houses occupied by just one inhabitant, with low consumption. Also profile P2 shows high values for L0, but mostly during the night. In addition, this profile exhibits a bimodal behavior for L3 and non-null coefficients for L2. This identifies a typical situation of average consumers that do not consume in the night, but have two peaks of medium water consumption in the morning and in the evening and low consumption (e.g., toilet) equally weighted during day hours. Finally, profile P3 is characterized by high values in the first eigenbehavior mainly associated to L2 and L3, again with a double peak in the morning and in the evening. This behavior might simply suggest houses in this cluster have larger size (in terms of number of occupants) that those in cluster P2.
The main characteristics of each profile are illustrated in Figure 4 in terms of relative frequency of average hourly consumption for each profile in each consumption class. This analysis confirms that profile P1 is associated to very low consumers, with high frequency

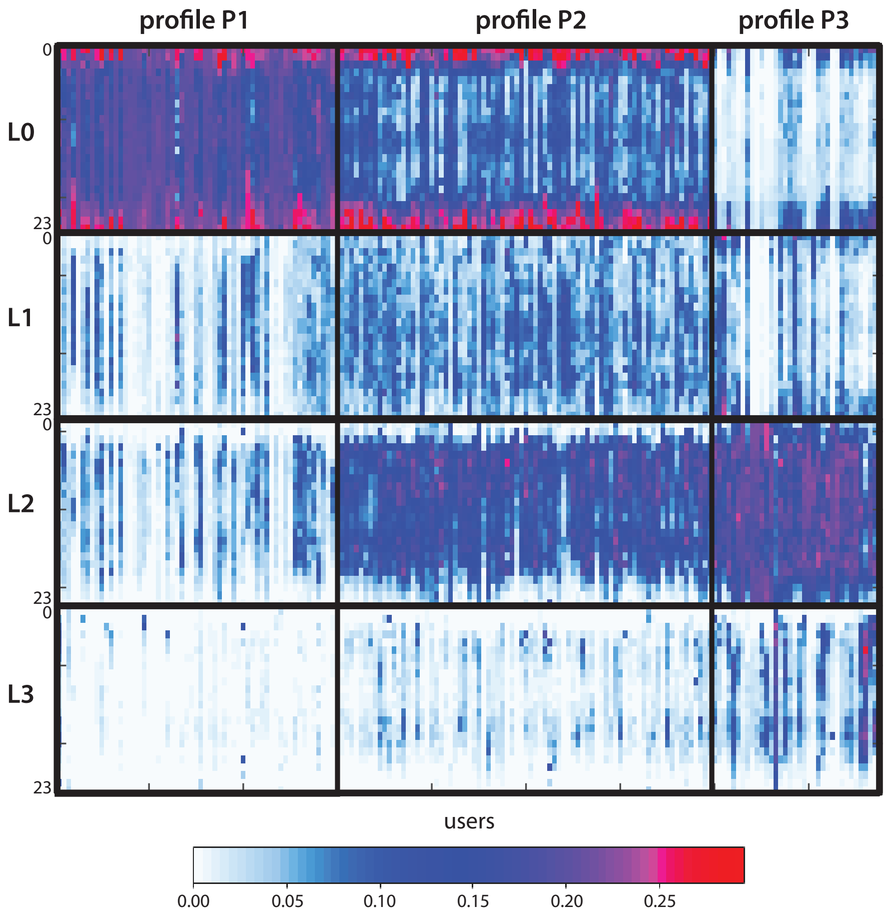

Figure 3. Results of users' segmentation with respect to the values of the first eigenbehaviors. Each column is the first eigenbehavior of a user and each row corresponds to the hour of the day for each consumption class L0, L1, L2, L3. Color scale represents the value of eigenbehaviors coefficients.

of no-consumption events. Profile P2 captures middle consumers that concentrate their water consumption in the morning and evening (see the double peak in the frequency of L2), with almost no consumption during night and few events in class L3. Finally, profile P3 identifies high consumers, characterized by a double peak of high consumption (class L3) and frequent consumption events of class L2 during the entire day.

5 Conclusions

This paper proposes a novel procedure for performing residential water users' segmentation from smart metered consumption data. Our procedure is based on a combination of clustering and principal component analysis and allows the automatic identification of recurrent water consumption behaviors in the form of eigenbehaviors. The approach is tested on a dataset of smart metered water consumption data from 175 households in the municipality of Tegna (CH). Numerical results show that the procedure successfully extract the main routines characterizing the metered users. Indeed, the three profiles identified represent typical consumption patterns [Cardell-Oliver et al., 2016] reflecting different behavioral habits of the users. This segmentation seems promising for inform-

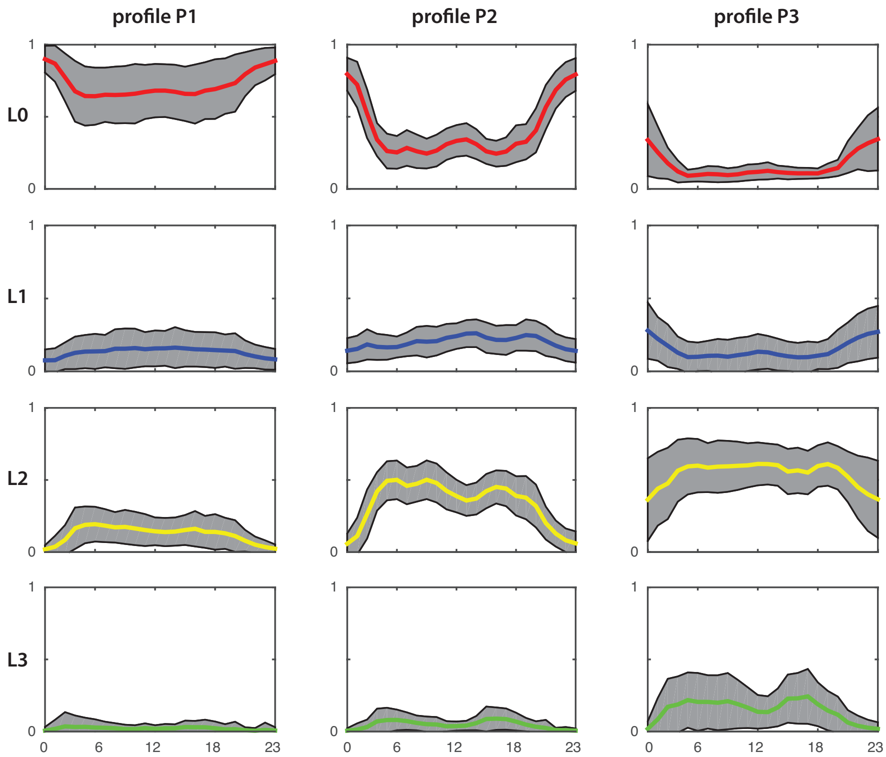

Figure 4. Relative frequency of average hourly consumption for each profile in each consumption class. The gray area represents standard deviation.

ing the design of customized water demand management strategies. Further research will focus on the comparison of the proposed methodology with other user profiling techniques and on the application of the procedure on larger datasets, possibly involving hundreds or thousands water users from different contexts.

ACKNOWLEDGMENTS

This work was partially supported by the *SmartH2O: an ICT Platform to leverage on Social Computing for the efficient management of Water Consumption* research project funded by the EU Seventh Framework Programme under grant agreement no. 619172.

REFERENCES

Beal, C. and Stewart, R. (2014). Identifying Residential Water End Uses Underpinning Peak Day and Peak Hour Demand. *Journal of Water Resources Planning and Management*, 140(7).

Beal, C., Stewart, R. A., Spinks, A., and Fielding, K. (2011). Using smart meters to identify social and technological impacts on residential water consumption. *Water Science and Technology-Water Supply*, 11(5):527–533.

Boyle, T., Giurco, D., Mukheibir, P., Liu, A., Moy, C., White, S., and Stewart, R. (2013). Intelligent metering for urban water: A review. *Water*, 5(3):1052–1081.

Cardell-Oliver, R. (2013). Discovering water use activities for smart metering. In *2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pages 171–176. IEEE.

Cardell-Oliver, R., Wang, J., and Gigney, H. (2016). Smart meter analytics to pinpoint opportunities for reducing household water use. *Journal of Water Resources Planning and Management*, page 04016007.

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., and Rizzoli, A. (2015a). Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environmental Modelling & Software*, 72:198–214.

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., Rizzoli, A., and Anda, M. (2015b). Modelling residential water consumers' behaviors by feature selection and feature weighting. In *Proceedings of the 36th IAHR World Congress*, The Hague (the Netherlands).

Eagle, N. and Pentland, A. S. (2009). Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066.

Gleick, P. H., Haasz, D., Henges-Jeck, C., Srinivas, V., Wolff, G., Cushing, K. K., and Mann, A. (2003). *Waste not, want not: The potential for urban water conservation in California*. Pacific Institute for Studies in Development, Environment, and Security Oakland, CA.

Joliffe, I. (2002). *Principal Component Analysis*. Springer, New York, N.Y.

Jorgensen, B., Graymore, M., and O'Toole, K. (2009). Household water use behavior: An integrated model. *Journal of Environmental Management*, 91(1):227 – 236.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.

Maggioni, E. (2015). Water demand management in times of drought: What matters for water conservation. *Water Resources Research*.

Makki, A. A., Stewart, R. A., Beal, C. D., and Panuwatwanich, K. (2015). Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption. *Resources, Conservation and Recycling*, 95:15–37.

Moro, A. and Riva, L. (2016). Inferring and clustering residential water consumers' routines by eigenbehavior modeling. Master's thesis, Politecnico di Milano, Italy.

Rizzoli, A., Castelletti, A., Cominola, A., Fraternali, P., dos Santos, A., Storni, B., Wissmann-Alvese, R., Bertocchi, M., Novak, J., and Micheel, I. (2014). The SmartH2O project and the role of social computing in promoting efficient residential water use: a first analysis. In *Proceedings of the 7th International Congress on Environmental Modelling and Software*.

UNDESA (2014). *World urbanization prospects: The 2014 Revision. Highlights (ST/ESA/SER.A/352)*. United Nations Department of Economic and Social Affairs Population Division.