Assessment of predictive models for estimation of water consumption in public preschool buildings

NebojšaM. Jurišević*, DušanR. Gordić, Vladimir Vukašinović and ArsoM. Vukicevic

University of Kragujevac, Faculty of Engineering, Sestre Janjića 6, 34000 Kragujevac, Serbia.
* Corresponding Author: jurisevic@kg.ac.rs

ABSTRACT

Preschool buildings are among the biggest water consumers in the public buildings sector, in which efficient management of water consumption could make considerable savings in city budgets. The aim of this study was twofold: 1) to assess prognostic performances of 21 parameters that influence the water consumption and 2) to assess performances of two different approaches (statistical and machine learning-based) with 6 various predictive models for the estimation of water consumption by using the observed parameters. The considered data set was collected from the total share of public preschool buildings in the city of Kragujevac, Serbia, over a three-year period. Top-performing statistical-based model was multiple linear regression, while the best machine learning method was random forest. Particularly, random forest gained the best overall performances while the multiple linear regression showed the same precision as the random forest when dealing with buildings that consume more than $200 m^3$/month. It is found that both methods provide satisfying estimates, leaving for potential users to choose between better performances (random forest) or usability (multiple linear regression).

Keywords: Water consumption; Public buildings; Consumption indicators; Predictive models.

INTRODUCTION

Rapid urbanization and population growth create additional requirements for water demand in cities, especially in developing countries (Oduro-Kwarteng et al., 2009). Consequently, the instability of water resources is increasing, not only in areas that have typically experienced water stress (Aleisa and Alshayji, 2019), but also in other areas because of the uneven distribution of rainfall caused by climate changes (Otaki, Ueda, and Sakura, 2017). Assuming these trends, it may be found that even small improvements in the management and consumption of resources in the service sector could make considerable savings in the future. The present study is focused on water consumption in public kindergartens, in which inefficient management increases buildings' operation costs.

In general, increasing price of water represents a step towards decreasing water consumption in residential buildings. Unfortunately, such an approach is less effective in the public building sector, where operational costs do not fall on the direct burden of building occupants. This may be justified with the fact that the water demand in the public sector depends on various determinants, including climate, culture, eating habits, working environment,

socio-economic factors, and physiology (Ahuja, 2016). Referring to this statement, educational buildings have a special status due to the following two reasons. First, educational buildings are among the largest consumers of water per occupant among public buildings. For example, educational buildings in the city of Bologna (Italy) consume 20% of the total share of water consumption, whereby educational institutions of lower educational levels have the highest specific water consumption (Farina et al., 2011). Therefore, these buildings should be prioritized for the implementation of strategies related to optimization and control of water consumption. Second – educational institutions play an important role in creating awareness of sustainable development among the youth population. Besides youth, teaching staff and parents are indirectly included in the process. Thus, the approach of early education could result in multiple benefits in the long term (Davis et al., 2008).

There are only a few studies that investigated water consumption in the public educational buildings. In general, energy audits commonly neglect water consumption, assuming that it does not affect the energy consumption within the buildings (Gallego Sánchez-Torija, Larrumbide Gómez-Rubiera, and Bedoya Frutos, 2017). On the other side, authors that accounted water consumption into their energy audits had water consumption per floor area as an indicator for ranking the buildings (Vatin, Gamayunova, and Nemova 2014), (Vatin Nikolai, Gamayunova Olga, and Nemova Darya, 2014). The drawback of these indicators is related to the fact that it does not explain the causes of water consumption. Consequently, a reliable prediction of water consumption cannot be made solely on this indicator. Another group of authors (Farina et al., 2011) studied the influence of building occupants to daily water consumption. Despite observed correlation, this variable can be considered as insufficient for the development of precise predictive models. In order to enable estimation of water consumption in nurseries and elementary public schools, the group of authors developed the multiple linear regression model (MLR) using eight significant predictive variables (Ilha, Gonçalves, and Amorim, 2005). Briefly, these authors examined the influence of more than 15 potentially significant variables, which were considered to be relevant for the cultural and geographical locality of the analyzed buildings. These parameters were not examined and assessed for other climate and cultural localities, which may have different organizational structures described by other consumption parameters. To sum up, the correlation between consumption and examined variables may not be the same for different localities. In a recent study, the MLR was also successfully employed with the intention to model water consumption in residential buildings in southern Brazil (Flores, Kalbusch, and Henning, 2018). The obtained findings indicated that MLR represents a reliable method for estimating water consumption.

With the ongoing progress of Industry 4.0[1] and the concept of smart cities, eased acquisition and availability of data have imposed the data science as a solution for more efficient predicting and management of the water consumption in public buildings. Following the growing demands for research on the topic, the aim of this study was twofold. First is to determine the most significant factors that influence water consumption in kindergartens. And second is to assess various predictive models (traditional and contemporary models) for the purpose of estimating water consumption in kindergartens.

MATERIALS AND METHODS

Water Consumption Data of the Analyzed Preschool Buildings

Data presented in this study describe water consumption of 13 public preschool buildings in the city of Kragujevac, Serbia, that were observed over approximately three-year period (Table 1). In these buildings, 437 employees provide care and services for approximately 3800 children of different ages. Children aged from 1 to 3

---

[1] Industry 4.0refers to the intelligent networking of machines and processes for industry with the help of information and communication technology(Federal Ministry for Economic Affairs andf Energy (Federal Republic of Germany) 2021).

years belong to the nursery group, while those aged from 3 to 7 years belong to the kindergarten group. The average ratio of children enrolled in kindergartens to children in nurseries is 3:1. Moreover, 9 of the total 13 observed preschool buildings are open from 5:30 to 16:30h, during the working days. The other four buildings (K3, K4, K5, and K12), provide the 4-hour preschool preparation program for children aged from 6 to 7 years. Therefore, their working hours during the period from September to June are extended to 20:30 h. The cooked meals for all users of kindergarten and nursery services are prepared in K6 and then thereby distributed to mess rooms in other buildings. Every building has a mess room and a handy kitchen for the dishwashing. The 4-hour preschool preparation program does not include meals for their attendants. Although almost all buildings have one or two washing machines (with capacities from 6-7 kg per wash) for their everyday necessities in kitchen and rooms, the main laundry rooms are placed in K5 and K9 with industrial washing machines (with capacities of 30 and 40 kg per wash, respectively).

Table 1. Characteristics of the analyzed public preschool building in the city of Kragujevac.

Preschool building	Built year	Shifts	Year of major reconstruction of water installations	Surface [m²]	Occupancy capacity	Kitchen [yes • / no -]	Londery room [yes • / no -]	Number of toilets	Number of toilet tanks	Number of ball faucets	Number of cartridge faucets	Max. water flow on faucets in handy kitchen
K1	1969	1	-	106	36	-	-	1	1	2	1	6.5
K2	1974	1	-	2002	427	-	-	8	16	32	2	14
K3	1980	2	-	1294	281	-	-	12	22	5	42	16
K4	1974	2	-	1306	288	-	-	5	15	22	5	20
K5	1973	2	-	942	236	-	•	5	9	23	1	29
K6	1974	1	-	1908	504	•	-	14	26	34	5	23
K7	1948	1	-	400	290	-	-	4	7	11	2	15
K8	1947	1	-	635	220	-	-	3	3	10	0	16
K9	1974	1	-	903	228	-	•	5	13	16	6	18
K10	2008	1	-	420	157	-	-	5	11	0	14	45
K11	1974	1	-	1486	406	-	-	15	27	33	6	18.3
K12	1982	2	-	2769	705	-	-	23	33	58	10	20
K13	2010	1	-	500	237	-	-	5	9	0	20	27

All data referring to water consumption were provided by the city utility company for water distribution. The rest of the information related to attendance registers, buildings' characteristics, equipment, and organizational structure were collected in coordination with representatives of the considered preschool buildings. Data referring to the equipment usage and process organization were collected through the process of surveying employees. The average period of observation was 34 months, although it varied for individual buildings depending on the availability of their data. The obstacle that has limited the retrospective analysis was the access to class registers,

which were available only for the period of the last 2–4 years. Moreover, it was noted that the public utility company has not measured water consumption every month in some buildings, since the bills for some months were very low or unavailable at all, while the bills for the following months were almost twice as high. Because of that, every two successive months with large oscillations in water consumption were excluded from the analysis. Most of the water consumption indicators used in literature are expressed with respect to these units, that is, m3/(occupant) or m3/m2. Although both indicators are equally represented in literature, this study was focused on the number of occupants since it is variable (and more convenient for estimating the consumption). Moreover, Farina et al., 2011 concluded that the presence of occupants in buildings is highly correlated with water consumption and essential for estimation of water consumption. Also, the recent study (Gamarra et al., 2018) concluded that results calculated per student are much easier to interpret by the school community, which is important for engaging them in the process of efficient water management. The values of these indicators for preschool buildings in the city of Kragujevac (Serbia), as well as some other European cities, are presented in Figure 1.
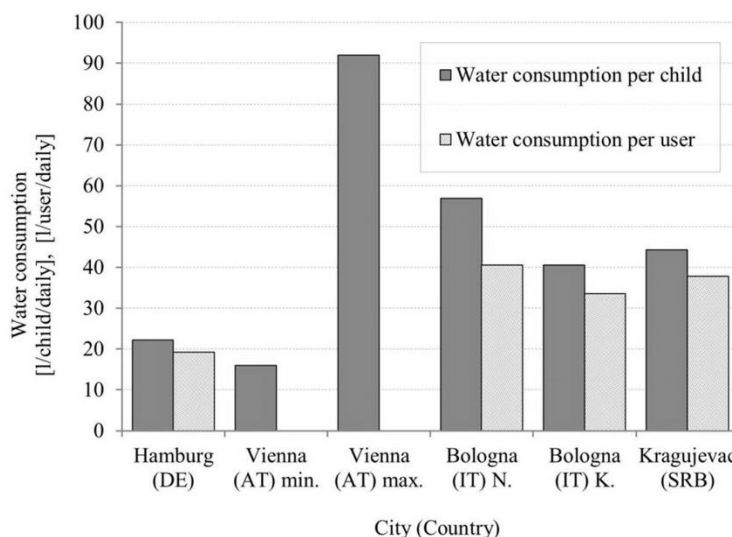

Figure 1. Average water consumption per child (red) and water consumption per user (blue) in different regions.
Note: min: minimum, max: maximum, N: nursery, K: kindergarten
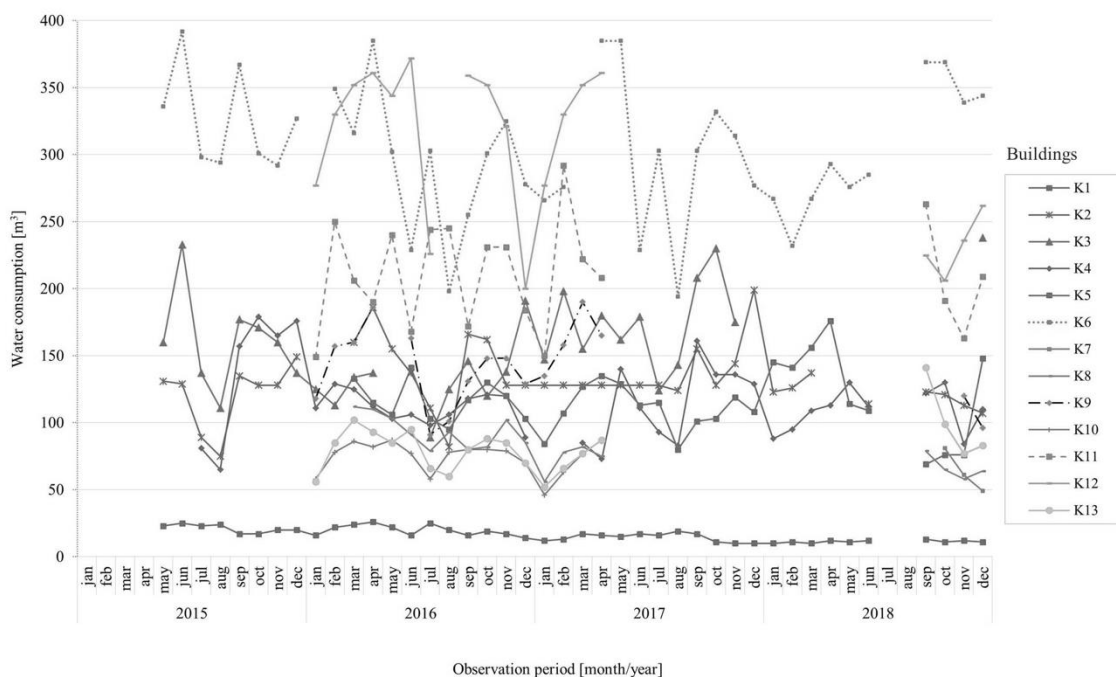DE: Germany, AT: Austria, IT: Italy, and SRB: Serbia.

It may be noted that indicators of the specific water consumption of kindergartens in Hamburg (Germany) are 5.8 m^3/child/annually, that is, 5 m^3/user/annually (Bernhard et al. 2007), or 22.2 l/child/daily and 16 l/user/daily, respectively.[2] A five-year research conducted on a sample of educational buildings in the city of Bologna (Italy) showed that the average consumption of water is 56 l/child/daily, That is, 43.5 l/user/daily in city nurseries, and 40.6 l/child/daily, that is, 33.6 l/user/daily in the city kindergartens (Farina et al., 2011). Mixed preschools buildings in the city of Kragujevac (Serbia), which account for kindergartens and nurseries, consume 44.3 l/child/daily and 37.9 l/user/daily. Although average values of indicators are interesting to compare with different groups of buildings, comparison of indicators among individual buildings should be considered more carefully. Particularly, relatively high values of indicators do not necessarily indicate irrational consumption, while relatively low values do not always indicate the absence of comfort. Organization of daily activities in the preschool buildings, condition of technical equipment in the building, and cultural and social factors determine water consumption in buildings. Therefore, the consumption of same-purpose buildings in the same city could be different. In Vienna (Austria), water consumption in preschool buildings ranges from 16 l/child/daily up to 92 l/child/daily (Bernhard et al., 2007).

---

[2]Values were normalized with respect to the number of working days in the Federal Republic of Germany

Additionally, the values of monthly consumptions of the same building could significantly vary from month to month during the year.

Variations of the monthly water consumption in kindergartens in the city of Kragujevac (Serbia) are presented in Figure 2. Since the consumption varies among the analyzed buildings, it may be concluded that the precise estimations could be made only by including more than one influential factor, which should provide a deeper understanding of buildings' behavior over time.

Figure 2. Time chart of the water consumption per user in kindergartens in the city of Kragujevac, Serbia.

Proposed Consumption Parameters and Assessment of Their Statistical Significance

The proposed indoor and outdoor parameters (Table 2) may be categorized into three groups.

1) First is the group of variable parameters, which could be related to a) weather conditions during the month (average temperature, sum of daily maximum temperatures, sum of daily minimum temperatures, and average air humidity) and b) presence of occupants during the month (number of kindergarten visits, number of nursery visits, number of adults visits, total number of all occupants' visits, and number of preschool group visits).
2) Second is the group of constant parameters, which could be related to characteristics of buildings (floor area, built year, presence of a kitchen and laundry room, number of toilets, toilet flushers, and faucets).
3) Relatively constant values, which changes during the month, are low or rare (maximal water flow in kitchen and bathroom faucets, as well as the number of laundry washings per month).

Linear correlation between various outdoor/indoor parameters and water consumption was estimated by computing the Spearman correlation coefficient which indicates a positive or negative association between the independent variable (water consumption) and the dependent variables (proposed parameters) (Table 2). Parameters that had the largest correlation with water consumption and the lowest mutual correlation were selected for further development and assessment of MLR model (Table 2, bold text).

It is important to emphasize that the approach used in this study differs from the literature (Ilha et al., 2005), where the authors for their MLR model used so-called transformed variables. Briefly, they used consumption indicators that were obtained by processing a couple of real-measured variables. In this study, the focus was on using untransformed-measurable parameters, which directly influence water consumption and ease the application of the presented findings in practice. Thus, the proposed approach enables practitioners to gain a better understanding of the correlation between individual parameters and water consumption. Following the findings given in Table 2, five independent variables (which describe characteristics and activities in buildings) were selected as the most significant and sufficient for the determination of the water demand in the observed public buildings. Particularly, the selected parameters were

- nursery monthly visits,
- total monthly presence,
- the presence of the kitchen,
- the number of toilet flushers and
- the maximal water flow on the toilet faucets.

The selection of proper influential parameters for other, nonlinear, predictive models was done by the application of the wrapper attribute evaluator algorithm to the corresponding decision table or decision tree classifier in the Weka Software.

Table 2. Spearman correlation between various outdoor and indoor parameters and water consumption.

No of Parameter/attribute		Monthly water consumption [m³]	Kindergarten monthly visits [-]	Nursery monthly visits [-]	Monthly adults visits [-]	Total monthly presence[3] [-]	Building floor area [m²]	Built year	Kitchen no/yes 0/1	Laundry no/yes 0/1	Average monthly temperature [C°]	M.S.[4] of daily max temperatures [C°]	M.S. of daily min temperatures [C°]	Average monthly air humidity [%]	M.K.W.F[5] [m³/min]	M.T.W.F[6] [m³/min]	Number of toilets [-]	Number of toilet flushers [-]	Number of ball faucets [-]	Number of cartridge faucets [-]	Total number of faucets [-]	Laundry washings per month	Presence of preschool group
	Monthly water consumption [m³]	1																					
1	Kindergarten monthly visits [-]	0.73	1																				
2	Nursery monthly visits [-]	**0.87**	0.71	1																			
3	Monthly adults visits [-]	0.86	0.85	0.83	1																		
4	Total monthly presence[3] [-]	**0.84**	0.96	0.84	0.94	1																	
5	Building floor area [m²]	0.80	0.80	0.84	0.88	0.87	1																
6	Built year	-0.41	-0.37	-0.29	-0.27	-0.33	-0.37	1															
7	Kitchen no/yes 0/1	**0.51**	0.34	0.41	0.43	0.39	0.35	-0.03	1														
8	Laundry no/yes 0/1	0.02	-0.28	-0.01	-0.03	-0.18	-015	0.22	-0.15	1													
9	Average monthly temperature [C°]	-0.02	-0.11	0.08	-0.10	-0.06	0.03	0.59	0.05	-0.03	1												
10	M.S.[4] of daily max temperatures [C°]	-0.03	-0.11	0.08	-0.11	-0.06	0.03	0.06	0.04	-0.03	0.99	1											
11	M.S. of daily min temperatures [C°]	0.00	-0.52	0.09	-0.07	-0.01	0.04	0.05	0.05	-0.04	0.95	0.94	1										
12	Average monthly air humidity [%]	0.00	0.06	-0.09	0.06	0.02	-0.02	-0.03	-0.02	0.02	-0.80	-0.80	-0.78	1									
13	M.K.W.F[5] [m³/min]	0.39	0.27	0.20	0.29	0.26	-0.20	-0.44	0.44	0.29	-0.04	-0.04	-0.42	0.02	1								
14	M.T.W.F[6] [m³/min]	**0.47**	0.21	0.28	0.33	0.24	0.27	-0.29	0.56	0.30	0.01	0.01	0.01	0.00	0.65	1							
15	Number of toilets [-]	0.88	0.74	0.84	0.82	0.81	0.85	-0.59	0.42	-0.16	-0.01	-0.01	0.00	0.00	0.36	0.36	1						
16	Number of toilet flushers [-]	**0.88**	0.76	0.84	0.82	0.83	0.87	-0.58	0.41	-0.23	-0.05	-0.01	0.00	0.00	0.30	0.36	0.98	1					
17	Number of ball faucets [-]	0.77	0.71	0.81	0.86	0.81	0.87	-0.00	0.49	0.08	-0.27	0.02	0.03	-0.02	0.24	0.23	0.73	0.73	1				
18	Number of cartridge faucets [-]	0.41	0.29	0.28	0.20	0.27	0.23	-0.91	0.03	-0.23	-0.07	-0.06	-0.06	0.03	0.35	0.27	0.58	0.58	0.63	1			
19	Total number of faucets [-]	0.86	0.72	0.79	0.41	0.79	0.86	-0.55	0.34	-0.15	0.02	0.01	0.03	-0.16	0.26	0.37	0.95	0.95	0.69	0.55	1		
20	Laundry washings per month	0.63	0.37	0.62	0.64	0.51	0.62	0.03	0.50	-0.49	0.07	0.06	0.06	-0.03	0.36	0.48	0.47	0.47	0.76	-0.11	0.49	1	
21	Presence of preschool group[7]	0.21	0.18	0.04	0.26	0.20	0.20	-0.13	-0.23	0.31	-0.06	-0.06	-0.04	0.03	0.27	0.17	0.14	0.14	0.17	0.09	0.33	0.30	1

[3]Total number of occupants' vistis.

[4] M.S.:monthly sum.

[5] Max water flow on kitchen faucets.

[6] Max water flow on toilet faucets.

[7] Presence of children in the second shift (from 16:30 to 20:30).

Development of Predictive Models for The Estimation of Water Consumption in Preschool Buildings

In order to develop and objectively assess various predictive models, data acquired in this study were split into training and independent test set. Data stratification was performed by random splitting of the overall data set (N=346) into the training (~70% samples) and test set (~30% samples), ensuring that both sets kept characteristics of the original set. Statistical characteristics of data sets used in this study are given in Table 3.

Table 3. Characteristics of data sets used for learning and assessment of the considered predictive models.

Characteristics	Overall (N=346)	Training (N=242)	Test (N=104)	p-value	VIF
Water consumption [m^3]	139.66±93.20	142.95±97.81	132±81.4	0.61	-
Nursery monthly visits [-]	705±569.20	726.71±590.47	654.59±515.53	0.46	4.85
Total number of occupants' visits [-]	3689±2220.93	3777±2372	3483.4±1813.94	0.65	4.84
Kitchen presence, n (%)	40(11.6)	32(13.2)	8(7.7)	0.20	1.72
Max water flow on toilet faucets [l/min]	14.34±3.83	14.31±4.01	14.4±3.39	0.81	1.97
Number of toilet flushers [-]	14.82±9.21	15±9.50	14.35±8.52	0.67	4.13

In order to determine the most suitable approach for estimating water consumption in the preschool buildings, authors considered two groups of methods with six predictive models:

1) statistical-based model which includes
    - simple linear regression (SLR)
    -multiple linear regression (MLR) and

8) machine learning-based model which includes
    - decision table (DT),
    - random tree (RT),
    - random forest (RF), and
    -evolutionary assembled artificial neural network (EAANN).

The SLR was applied with the intention to assess a possibility to estimate a building water consumption using a single most influential factor that can be applied for all the analyzed buildings (total building presence). The SLR and MLR were developed using the MS Excel, while the following three machine learning models (DT, RT, and RF) were developed by using the Weka software (Witten, Frank, and Hall, 2011). The EAANN assumes employing the Genetic Algorithm (GA) to evolutionary configure ANN with respect to the selected criterion; in the present study authors aimed to maximize the correlation coefficient (R^2) (Vukićević et al., 2014). Regarding the EAANN settings, the number of GA generations was set to 50 and the size of the population was set to 200. Boundaries of

the ANN hyperparameters, which were considered for the optimization, were set to a minimum number of neurons in the hidden layer that was set to 3 and the maximum was set to 50, while the considered activation functions, training algorithms, and ANN error-functions were those originally proposed. EAANN model was developed by using the MATLAB 2017b (MathWorks, Natick, MA).

RESULTS AND DISCUSSION

For clarity, this section was split into two subsections, which separately present and discuss two major contributions of the present study. The first section is related to the assessment of the proposed consumption parameters (statistical significance of the considered parameters), while the second focuses on the interpretation of the developed predictive models (performances of the developed predictive models). All the data and predictive models developed in this study are publically available on the repository.[8]

Statistical Significance of The Considered Parameters

In general, all parameters related to the presence of kindergartens occupants have shown considerable linear correlation with the water demand. Additionally, it is important to emphasize that they have also shown to be intercorrelated to each other (i.e., the number of nurses is related to the number of children). In order to avoid multicollinearity, the authors calculated variance inflation factor (VIF) and selected independent and the most influential parameters (Table 3). Nursery monthly visits and a total number of occupants' visits have shown the greatest liner correlation with water consumption (Table 2). Consequently, the present study suggests that the number of nursery monthly visits has a great influence on the water consumption because children aged from 1 to 3 years require specific attention (related to daily showering, which is contrary to older groups). Maximal water flow on a single toilet faucet did not vary significantly during the sampling (±0.5 l/min). The remaining two selected parameters, the presence of kitchen in buildings and the number of toilets, represent static values that did not varied over the observed period.

Regarding the rest of the parameters, which were excluded from further analysis, the authors emphasize that some of them showed a minor individual influence on water consumption. These are the average monthly temperature and the monthly sum of daily max temperatures, which both showed a negligible impact on the water demand in preschool buildings in the city of Kragujevac. The presented findings are contrary to those obtained by the previous study that focused on the city of Campinas, São Paulo (Brazil) (Ilha et al., 2005), while they are in accordance with the study that observed all buildings in 103 Italian cities (Romano, Salvati, and Guerrini, 2016). It was interesting to note that the presence of laundry room in the kindergarten buildings was not a parameter of significance neither for linear nor nonlinear methods, although some buildings had industrial washing machines (K5 and K11) with capacities of 30–50kg per wash. Briefly, the washing process in such buildings is more efficient (expressed in the water consumed per kg of washed laundry) compared to other buildings that had conventional washing machines with capacities of 6 kg. The remaining parameters, that are not discussed in this section, have shown to be more correlated with the mentioned influential parameters than with the water demand itself. Thus, they were excluded from further analysis.

---

[8] The repository (DOI: 10.5281/zenodo.3908361) includes: 1) Excell file with data and SLR and MLR, 2) developed Weka models and input *.arff files for DT, RF, and RT models, and 3) Matlab script and *.mat files for running the developed EAANN model.

Performances of The Developed Predictive Models

Performances of predictive models are expressed by measuring their coefficients of determination (R^2), which assess the strength of the linear relationship between the real and predicted values, and the Mean Absolute Percentage Error (MAPE), as an additional measure of prediction accuracy. R^2 values that are close to 1 and MAPE values that are close to 0 indicate better forecasting results. A comparison of the most interpretable and most accurate statistical and machine learning predictive models is shown in following figures. SLR and MLR model predictions and the corresponding ground truth values are shown in Figure 4. MLR model performed better than SLR in both data sets (0.89 vs. 0.59 in terms of R^2 and 88% vs. 35% in terms of MAPE, on training and test data set, respectively). On the other hand, the SLR model is based on just one predictive variable (total number of occupants' visits) contrary to MLR which requires five different predictive parameters (Table 4). Thus, the SLR predictive formula is more intuitive than the one used for MLR predictions.
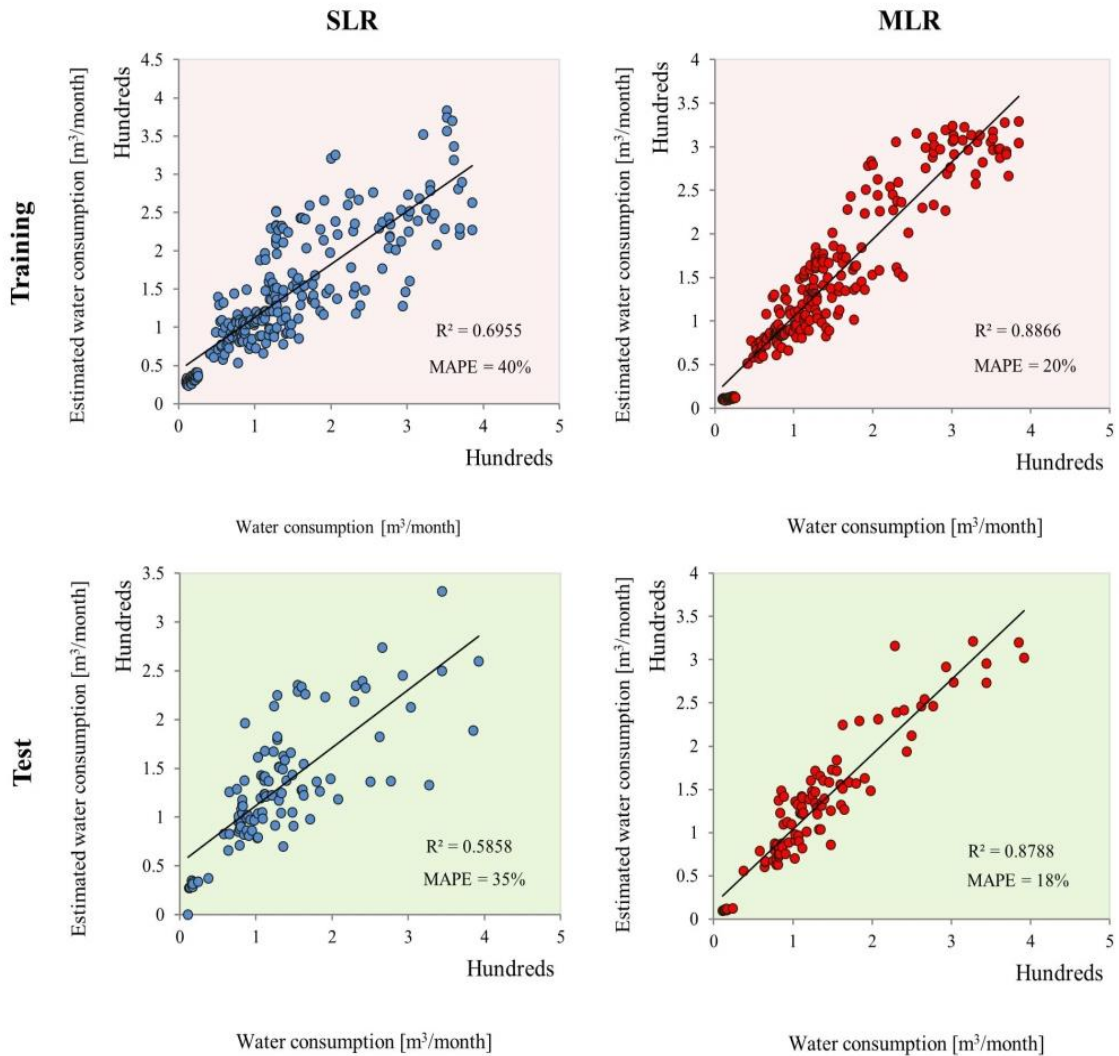

Figure 4. Correlations between ground truth and SLR and MLR predictions of water consumption in analyzed buildings.

Decision Table and random forest model predictions and the corresponding ground truth values are shown in Figure 5.
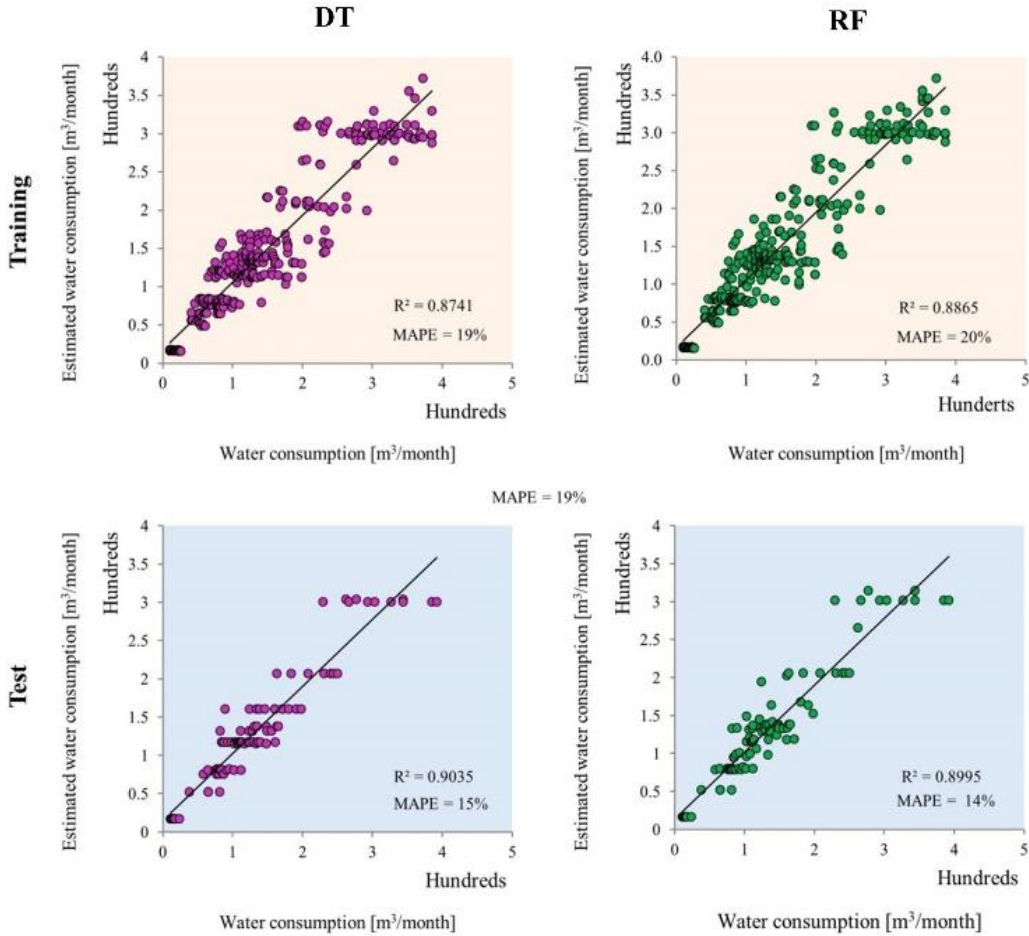

Figure 5. Correlations between ground truth and DT and RF predictions of water consumption in analyzed buildings.

Both models show more or less similar performances with slightly better test results in terms of MAPE in favor of RF (14% vs. 15%) and slightly better R^2 results in favor of DT (0.90 vs. 0.89). DT model was based on two predictive parameters (presence of kitchen and number of toilet flushers), while RF model required three predictive attributes (monthly sum of daily max temperatures, total number of faucets, and presence of preschool group).

In addition to these performances, the authors assessed all predictive models with respect to three ranges of the water consumption in the considered locality (low< 50 m³/month, medium 50 - 200 m³/month, and high> 200 m³/month). Performances obtained on the training and test sets are shown in Table 4. In order to make more objective conclusions, the discussion was based on the results obtained from the independent test set.

Table 4. Performances of the developed predictive models on training and test set.

Method	Selected attributes (no. of attributes from Table 2)	Data set	Water consumption [m³/month]				
			Overall		< 40	40 - 200	200 >
			R^2	MAPE			
SLR	(4)	Training	0.69	40%	98%	32%	23%
		Test	0.59	35%	106%	28%	24%
MLR	(2) (4) (7) (14) (16)	Training	0.89	20%	30%	20%	13%
		Test	0.88	18%	30%	17%	12%
DT	(7) (16)	Training	0.89	20%	34%	20%	14%
		Test	0.90	14%	20%	14%	13%
RT	(17) (21)	Training	0.89	19%	34%	20%	11%
		Test	0.88	15%	20%	15%	11%
RF	(10) (19) (21)	Training	0.89	20%	34%	20%	12%
		Test	0.90	14%	20%	14%	12%
ANN	(2) (3) (13) (14) (15) (17) (20)	Training	0.92	16%	37%	14%	10%
		Test	0.90	26%	127%	14%	11%

In terms of R^2, all performing models show more or less similar results ranging from 0.88 to 0.90. Referring to the MAPE, for the buildings with high water consumption, the top-performing algorithms were ANN and RT, (11%). For the buildings with medium water consumption, the best performances were ANN, RF, and DT (14%). And for the buildings with low water consumption, RF, RT, and DT algorithm (20%) were the best. In general, RF and DT resulted in the lowest MAPE (14%). Considering the overall metrics, the obtained results showed that there is no statistically significant difference between top-performing algorithms (for most of the models, R^2 was in the range 0.88-90). Thus, besides performances, the selection of the most valuable predictive models can be performed by accounting their usability and interpretability.

The usability of the developed predictive models represents an important aspect from the viewpoint of practitioners and audience that is less familiar with statistics and data science. The most intuitive is the SLR model, which is relatively simple to employ and interpret, even by non-experts. In this study, the SLR has shown to be considerably less accurate (R^2=0.59, MAPE=35%) compared to competing models. Compared to the SLR, all other methods, achieved better predictions in terms of R^2 values - obtaining 0.89 and 0.91, respectively. As may be noted from Table 4, the gap between performances achieved on two data sets varies among predictive models. In particular, in terms of R^2 the gap between training and data set of the developed models is relatively negligible, except in the case of the SLR model that achieved a performance gap of around 10%. Compared to the SLR, the MLR requires more advanced analysis to interpret the influence of its parameters. In literature, a similar study proposed the MLR - which was developed using six input parameters of kindergarten in the city of Campinas (Brazil) and resulted in R^2=0.67. Another study (Flores et al. 2018) focused on residential buildings in southern Brazil explains more than 70% of water consumption in buildings and per capita consumption. However, dealing

with water consumption in residential buildings is less complicated compared to public buildings (Farina et al. 2011). Regarding the ANN, its development was done automatically using the EAANN algorithm that converged after approximately 40 generations reaching $R^2=0.92$. Although the best performing in terms of accuracy, the downside of the ANN model is that it requires experience with the software that is not commonly used in management offices. Moreover, since it works as a black box, it is not possible to intuitively interpret the influence of particular input variables on the obtained output.

CONCLUSION

The efficient management of water resources in preschool buildings is a nontrivial challenge, which advancements could make considerable savings in city budgets. By using the data collected from 13 public preschool buildings during the three-year period, this study: 1) assessed 21 parameters that influenced the water consumption and 2) developed 6 predictive models for the estimation of water consumption. By considering achieved performances and usability of the developed predictive models, multiple linear regression (MLR) and random forest (RF) were recommended as the most suitable for predicting water consumption (leaving for potential users to choose between slightly better precision of RF and usability of MLR).Black-box based ANN algorithm, although the most challenging to develop, did not perform better than the other machine learning methods. Moreover, ANN resulted in greater MAPE values in low range consumption predictions than other the methods.

With the ongoing progress towards the Industry 4.0 and the concept of smart cities, it is expected that eased access and availability of data will enable breakthrough of data science and predictive modeling into the water management of public buildings.

At the moment, challenges that need to be addressed to enable wider usage of predictive models in water management are related to two directions. The first is the education of public building managers, who need to be train how to use, interpret and benefit from data-driven expert systems. The second is related to data accessibility, which is a prerequisite for both development and exploitation of predictive models. Regarding the training of predictive algorithms, it is important to emphasize that, so far, this process assumed a long period of data collection. However, with the ongoing rapid progress of computer technology and digitalization (Open data initiative, 5G networks, etc.), it is expected that the data collection will be less issue for further studies. As an addition to that, equipment automation will provide more frequent sampling that will allow daily or hourly water demand predictions. Considering these trends, it is expected that the impact of artificial intelligence and ICT will constantly grow resulting in better management of water consumption in kindergartens.

ACKNOWLEDGMENT

REFERENCES

Ahuja, V. 2016. Water Consumption Norms and Utilities Management. 10th World Aqua Congress, New Delhi, India.

Aleisa, E. &Alshayji, K. 2019. "Analysis on Reclamation and Reuse of Wastewater in Kuwait." Journal of Engineering Research 7 (1): 1–13.

Bernhard, M. Florian, M. Rüthrich, W. Kluge, T. Deffner, J. Götz, K. Liehr, Michel, S. 2007. Water demand forecast 2030 for Hamburg (in German). Frankfurt am Main/Darmstadt, Germany.

Davis, J. Miller, M. Wendy, B. & Megan, G. 2008. The impact and potential of water education in early childhood care and education settings: A report of the Rous Water early childhood water aware center program. Brisbane, Australia.

Farina, M. Maglionico, M. Pollastri, M. &Stojkov, I. 2011. Water consumptions in public schools. Procedia Engineering 21:929–38.

Federal Ministry for Economic Affairs and Energy (Federal Republic of Germany). 2021. PlattformIndustrie 4.0 – Startseite. Retrieved July 16, 2021 (https://www.plattform-i40.de/PI40/Navigation/DE/Home/home.html).

Flores, T. Kalbusch, A. & Henning, E. 2018. Factors influencing water consumption in buildings in southern Brazil. Journal of Cleaner Production 184:160–67.

Gamarra, A. R. Istrate, I. R. Herrera, I. Lago, C. Lizana, J. &Lechón, Y. 2018. Energy and water consumption and carbon footprint of school buildings in hot climate conditions. Results from life cycle assessment. Journal of Cleaner Production 195:1326–37.

Ilha, M. S. O. Gonçalves, O. M. & De Amorim, S. V. 2005. Estimating water consumption in nursery and elementary public schools. (1):1–10.

Oduro-Kwarteng, S. Nyarko, K. B. Odai S. N. &Aboagye-Sarfo, P. 2009. Water conservation potential in educational institutions in developing countries: case study of a university campus in Ghana. Urban Water Journal 6(6):449–55.

Otaki, Y. Ueda, K. & Sakura, O. 2017. Effects of feedback about community water consumption on residential water conservation. Journal of Cleaner Production 143:719–30.

Romano, G. Salvati, N. &Guerrini, A. 2016. An empirical analysis of the determinants of water demand in Italy. Journal of Cleaner Production 130:74–81.

Sánchez-Torija, J. G. Gómez-Rubiera E. L. &Frutos, C. B. 2017. The incorporation of the study into water consumption in energy audits in schools. Revista de La Construcción 16(3):361–73.

Vatin, N. Gamayunova, O. &Nemova, D. 2014. An energy audit of kindergartens to improve their energy efficiency. 4th International Conference on Civil Engineering and Building Materials, CEBM 2014 305–8. Saint Petersburg, Russian Federation.

Vatin, N. Gamayunova, O. &Nemova, D. 2014. Conducting an energy audit in kindergartens in order to improve their energy efficiency (in Russian). Construction of Unique Buildings and Structures 9(24):71–83.

Vukićević, A. Jovičić, G. Stojadinović, M. Prelević, R. &Filipović, N. 2014. Evolutionary assembled neural networks for making medical decisions with minimal regret: Application for predicting advanced bladder cancer outcome. Expert Systems with Applications 41(18):8092–8100.

Witten, I. Frank, E. & Hall, M. 2011. Data mining: Practical machine learning tools and techniques (Google EBook).