

# Long-term water-energy demand prediction using a regression model: a case study of Addis Ababa city

Bedassa Dessalegn Kitessa , Semu Moges Ayalew,  
Geremew Sahilu Gebrie and Solomon Tesfamariam Teferi

## ABSTRACT

As part of sustainable urban planning, the demand for water and energy (WE) should also be addressed. The Waikato Environment for Knowledge Analysis (WEKA) modeling tool was employed to relate the historical WE consumptions with the population and economic growth scenarios using a linear regression model. The performance of the model was evaluated to properly identify the most influential drivers in each sector. The WE demand prediction was made for each year from 2016 up to 2050. Consequently, the long-term time interval for demand analysis is important rather than the consequent year for planning. The total electric energy demand including residential, street-lighting, commercial and industrial sectors was estimated to be around 14,000 and 53,000 Giga Watt hour (GWh) for the years 2030 and 2050, respectively. These years' forecasted petroleum demand was around 8840 and 30,140 for diesel, 13,860 and 52,700 for gasoline, and 1230 and 9890 GWh for kerosene and the water demand including residential, commercial and industrial sectors were 520 and 1600 million cubic meters (MCM). The proposed methodology can comfortably be used to predict the urban WE demand corresponding to economic (gross domestic product and per capita income) and population growth at different scenarios which could support policy makers.

**Key words** | consumption, demand, energy, regression, water, WEKA

**Bedassa Dessalegn Kitessa**  (corresponding author)

**Geremew Sahilu Gebrie**

**Solomon Tesfamariam Teferi**

Addis Ababa Institute of Technology,  
Addis Ababa University,  
P.O.Box 385, King George VI St, Addis Ababa,  
Ethiopia  
E-mail: [bedassa.dessalegn@aait.edu.et](mailto:bedassa.dessalegn@aait.edu.et)

**Semu Moges Ayalew**

School of Civil and Environmental Engineering,  
University of Connecticut,  
Storrs, CT,  
USA

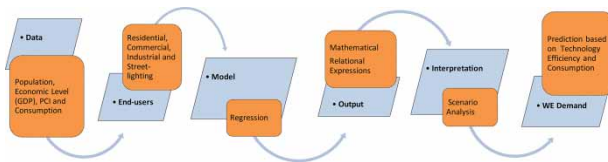
## HIGHLIGHTS

- Predicting long-term water-energy demand is important for planning.
- A linear regression model is used for long-term water-energy demand predicting.
- The water-energy demand in urban areas is increasing.
- Population and economic growth are the main factors which are highly affecting the urban water-energy demand.
- Identifying the most influential drivers on water-energy demand is important for water-energy supply planning.
- Technological factors (such as water loss, energy loss) are commonly considered in demand prediction.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

doi: 10.2166/wcc.2021.012

## GRAPHICAL ABSTRACT



## INTRODUCTION

Water-energy (WE) are the most critical resources to support socio-economic development, and are fundamentally linked and holistic to achieve socio-economic sustainability (Hao *et al.* 2020). Water and energy are interlinked with each other, particularly energy is intensively used in different sections of water supply systems (such as water transmission, treatment, distribution, extraction) and wastewater treatment sections (Kitessa *et al.* 2020). As strategic resources for the survival and development of humans, water-energy (WE) affects the stability and security of society. Different cities across the globe are stressed with WE supply due to natural and social factors including economic development, rapid population growth and climate change (Lee & Kim 2018). Ethiopia's urban population is expected to triple by 2037 and the rate of urbanization including Addis Ababa city is expected to accelerate at a rate of 5% annually (World Bank 2015a, 2015b, 2015c). In 2015, the city registered a GDP of about 4.32 billion USD and a per capita income of 1,359 USD (BoFED 2016).

To solve the problems of cities' WE supply, accurate WE demand prediction and the expansion and efficient operation of WE supply and distribution facilities are essential. Reliable and accurate WE demand prediction is essential to develop accurately WE supply expansion strategies cost-effectively. However, prediction of WE demand is a challenging task due to the availability of data, various influencing factors (e.g. social, economic and technological factors), different prediction time horizons and different prediction methods (Donkor *et al.* 2014). The urban WE demand is mainly affected by two socio-economic drivers, namely: rapid population growth and economic growth of the city (Luvimba Ramulongo 2017; Nhamo *et al.* 2018).

The WE demand prediction can be categorized as short-term forecasting (STF), medium-term forecasting (MTF) and long-term forecasting (LTF). The hourly to weekly, monthly, annual to decadal predictions are classified as STF, MTF and LTF respectively (Tiwari & Adamowski 2013). Long-term prediction is used for decision problems such as capacity expansion at the strategic planning level (Youngmin Seo 2018). This study is focused on long-term predicting based on the annual WE demand time series. However, a different study reveals that long-term WE predicting received less attention compared to short-term demand predicting (Swasti *et al.* 2018), this is because of the complexity involved in achieving accurate forecasts.

There are different parameters affecting demand forecasting. The main parameters are listed as follows (Hossein & Mohammad 2011): Time factors such as hours of the day, day of the week and time of the year, weather conditions, class of customers (end-users or sectors), socio-economic indicators (population, PCI, GDP, etc.), trends in using new technologies, WE price. However, when moving towards longer periods, the accuracies of some driving parameters drop. For instance, the prices and weather parameter for the STF is more accurate than that of the MTF. Due to inaccuracies involved in the long-term driving parameters, it is common practice to perform long-term demand forecasts using different scenarios (such as GDP, PCI and population scenarios) (Feinberg & Genethliou 2012).

Long-term WE forecasting requires observing the short-term fluctuations of the many variables that affect demand (e.g. temperature, politics, etc.). Therefore, for the long-term, instead of complex or hybrid models, a more practical

approach is needed to estimate the WE demand of any cities (Arturo Morales 2014).

Long-term demand predicting is based on the integration of concepts from theoretical foundations of economic theory with knowledge of financial, statistical, probability, and applied mathematics to make inferences about the demand growth (Swasti *et al.* 2018). Long-term WE demand prediction takes into account socio-economic factors like population growth, GDP and technological factors along with explicit factors like historical WE consumption (Swasti *et al.* 2018). The technological factor is related to the sustainable and appropriate use of technology to increase WE efficiency.

The relationship between energy and GDP or PCI has been well reported by different researchers (Cleveland 2000). A study made in the USA showed that PCI and energy consumption are interrelated (Kraft 1978). The energy consumption under model uncertainty was estimated using GDP and PCI as a proxy for wealth and the main driving force of energy consumption was observed to be GDP (Cserekyei 2012). For LTF, the GDP and population have been mostly used. Of all the LTF studies, about 41% studies used the GDP and 49% used population data as energy demand driving variables (Mir *et al.* 2020). This indicates, for LTF, that the GDP, population, and previous energy consumption data were the most commonly used demand determinants.

The long-term WE demand prediction approach can be classified into time-series, econometric, and end-use approaches (Donkor *et al.* 2014). These are broadly categorized into traditional (statistical) and non-traditional (artificial intelligence) based methodologies. The regression models and time series methods are some of the traditional techniques. In artificial intelligence (AI) based techniques, ANN is one of the most popular models. An AI-based technique was mostly used for STF demand (Mir *et al.* 2020). Many variants of regression analyses can be found in the literature such as linear regression (LR) or multiple linear regression (MLR), smooth transition autoregressive models, support vector machines (SVM) and so on. The SVM model appeared to be the winning entry for MTF (He *et al.* 2017). Moreover, the inclusion of regression analysis in some of the top entries of Gefcom2012 further vouches for its significance in predicting (Hong *et al.*

2014). The energy demand for both STF and LTF can be conveniently produced using MLR analysis (Hong 2016).

In many quantitative structure-property studies, the linear regression (LR) method is commonly used for simple and multiple linear issues whereas an artificial neural network (ANN) model can be used to resolve relatively complex non-linear issues (Ziyi Yin 2018). Recently, the LR model has been adopted by many researchers to develop WE demand predicting models (Al-Musaylh 2018; Quilty 2018). Neural networks and hybrid models are more suitable for STF of WE demand, while regression-based models are more appropriate for LTF (Haque 2018).

Different studies implemented an extreme machine learning (EML) model for predicting the standardized precipitation evapotranspiration index (SPEI) and compared its enactment to that of an LR, an ANN, and the least support vector regression (LSSVR) models (Shahabbodin Shamshirband 2020). The analysis between observed and predicted SPEI indicated the potential of the developed models is contributing more in understanding the potential of future predicted drought-risks in Australia. The three data-driven models, auto-regressive moving average, ANN and k-nearest-neighbors (K-NN), are applied to short-term rainfall predictions (Toth 2000). The interpretation showed that ANN performed the best in terms of the accuracy of runoff forecasting when the predicted rainfalls by the three models were used as inputs of a rainfall runoff model.

The other soft computing-based methods besides ANN include support vector regression (SVR) and fuzzy logic (FL), which are used for rainfall prediction for contemporary studies. ANN and fuzzy logic were employed to predict rainfall either using the different meteorological parameters or using only the rainfall time series (Chau 2013).

ANN has been widely used for the prediction of water resource variables in different hydrological contexts such as rainfall-runoff modeling or stream flow prediction (Riccardo & Kwok-Wing 2015). On the other hand, ANNs simulate the concept of biological neural networks to identify pattern determination, data building and modeling in many types of researches; ANNs were used to model environmental impacts and energy consumption (Najafi 2018). The two artificial intelligence (AI) methods, namely, ANN and the adaptive neuro-fuzzy inference system (ANFIS) model, were used for predicting life cycle

environmental impacts and the output energy of sugarcane production in planted farms (Ali Kaab 2019). The ANFIS model was also used to correlate the observed values and predicted value of the energy output in converting paddy to white rice in milling factories, indicating high accuracy in predicting the energy yield in milling factories (Ashkan Nabavi-Pelesaraei 2019). The LR model was used as the predicting energy for the different sectors as this has been noted to be the most appropriate statistical technique for LTF (Makridakis 1998). There have been previously reported studies where LR has been used for LTF of energy consumption in other countries, either for total consumption (Bianco 2009) or for a specific sector (Al-Ghandoora 2008).

Either an LR model or an ANN model can be employed to predict a WE demand (Ziyi Yin 2018). Most of the models used for demand prediction have limited accuracy since the socioeconomic input variables are measured annually and are uncertain (Donkor 2012). For LTF, the most popular LR model can be employed to give accurate results using socio-economic parameters (Sullivan 1977; Christian-Smith 2012). This model is a stochastic approach correlating the dependent variable  $y$  with one or more independent variables denoted by  $x$  (Autar & Egbu 2010). The regression analysis model is advantageous as it allows for the prediction of an outcome even when multiple predictors are correlated with each other and can give good results for small datasets (Isaac *et al.* 2017). A regression model, being non-black box in nature (Hong 2016), reveals insightful information about demand driving variables such as the GDP, PCI and population (Khan & Jayaweera 2018). The relationships between these drivers and WE demand can be instrumental for researchers and policy makers in devising WE policies and in demand side management. Meanwhile, concerns related to the accuracy level of this model may counterweigh its merits.

There are different commercial data mining tools such as Oracle DM, Microsoft Analysis Services, SPSS Clementine, and SAS Enterprise Miner (Nefeslioglu & Sezer 2010) to forecast WE demand. Additionally, the other best five open source data mining tools are Orange, Tanagra, YALE (Yet Another Learning Environment), WEKA (Waikato Environment for Knowledge Analysis), and KNIME (Konstanz Information Miner) (Sharma 2015). The WEKA toolkit is the best tool in terms of the ability to run the selected classifier followed by

Orange, Tanagra, and finally KNIME respectively (Sharma 2015). Written in Java, WEKA is a well-known suite of machine learning (ML) software ([www.cs.waikato.ac.nz](http://www.cs.waikato.ac.nz)) that supports several typical data mining tasks, particularly data preprocessing, clustering, classification, regression, visualization, and feature selection. WEKA is employed for data mining (DM) and machine learning (ML) and can be deployed on any given problem (Kunder 2001; Mohammed 2017).

Different performance indices are available to determine the most sensitive driving parameters and validate the regression model results. It is common to use multiple performance indices since there are pros and cons in each index (Krause *et al.* 2005). The most widely used accuracy measures are mean absolute error (MAE), Nash–Sutcliffe Efficiency (NSE) and water balance error (WBE) (Muhammad Shahid 2020). Similarly, mean absolute percentage error (MAPE), root mean square error (RMSE) and coefficient of determination (CoD) are employed to evaluate the performance of models (Renno 2016).

As discussed above, there have been various studies on ML forecasting techniques to predict WE demand. The present study aims at creating a more comprehensive dataset and deploy it with an ML linear regression model using the WEKA tool which has not yet been explored for long-term WE demand prediction for individual end-users or sectors in the open literature. The tool helps to determine the values of the most influential parameters such as population, GDP, PCI, that affect the LTF of WE demand for each sector. Consequently, the study combines technological or efficiency factors (water loss, energy loss), socio-economic factors and WE consumption to predict the WE demand.

---

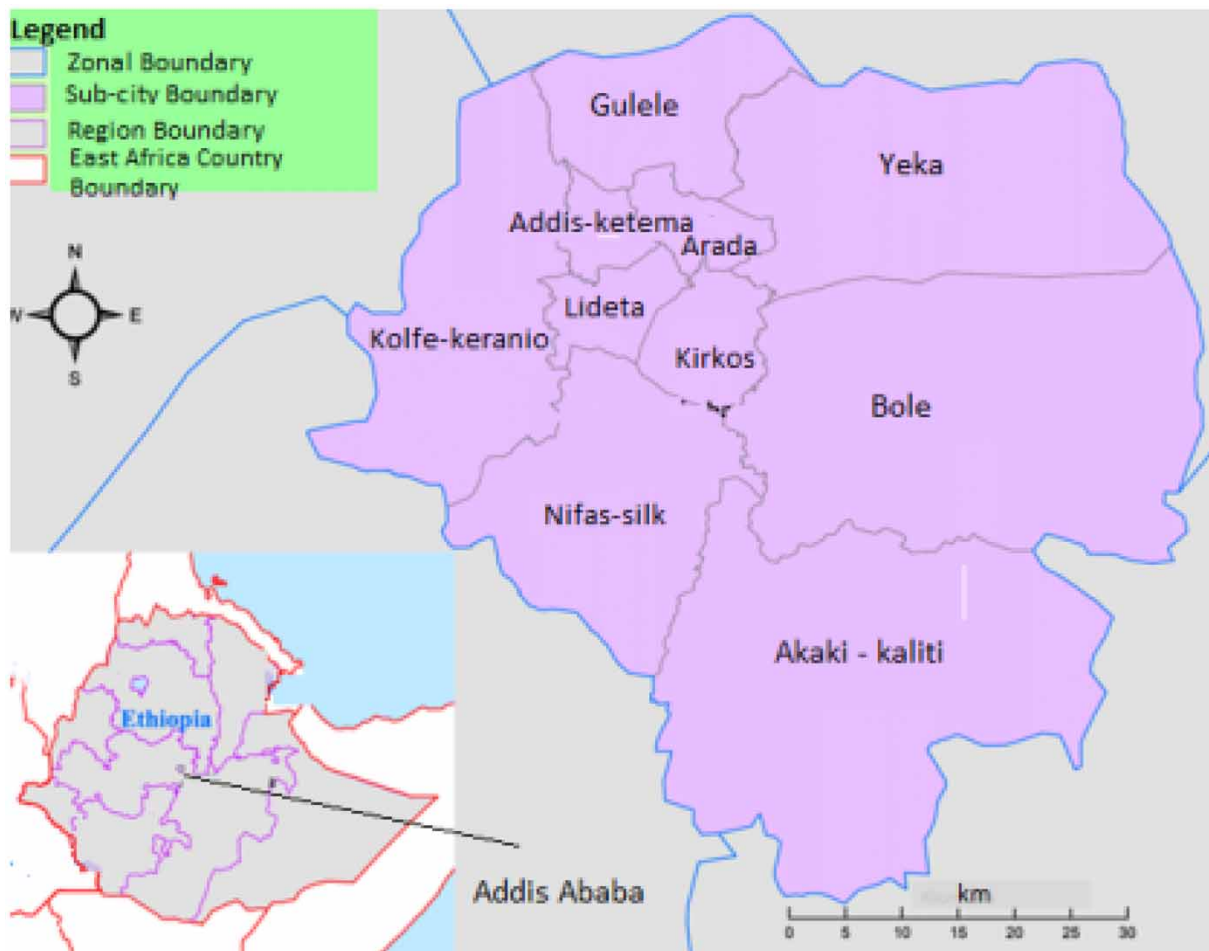
## METHODOLOGY

### Overview of Addis Ababa city

The case study was carried out focusing on Addis Ababa city since its populations and economic growth vary considerably compared to other cities in the country. The city currently covers an area of 540 km<sup>2</sup>, as obtained from the city map. The city lies between 2,000 and 3,000 m above sea level, enjoying a mild and warm temperature climate. The lowest and

highest annual average temperatures recorded are about 10 and 25 °C respectively. The average annual rainfall is around 1,250 mm. The city has insufficient WE supply for the growing demand (Ethiopia Ministry of Transport 2011) that has resulted from human factors (such as population growth and economic expansion). The city currently contributes approximately 50% towards the national GDP and is characterized as having the highest levels of WE consumption. Addis Ababa (38°44'E and 9°1'N) is home to 25% of the urban population in Ethiopia and is one of the fastest growing cities in Africa. Addis Ababa's economy is growing annually by 14% and it is noted to be among the fastest urbanizing metropolises in Africa. For this city, it is anticipated that the WE demand will outstrip the WE delivered by 2050. The total ground

and surface water supply is around 450,000 m<sup>3</sup>/day and 36.5% of the water is lost due to leakage (World Bank 2015a, 2015b, 2015c). The city water scarcity is expected to become significant due to rapid urbanization, increased individual water demand, and the impacts of climate change. There are enough energy resources in the country. However, there is a limited capacity of energy expansion, as well as transmission and distribution in the city. The energy supply is deprived due to the aging distribution system network and outage, which are less likely to provide an efficient and reliable service to end-users. Moreover petroleum is imported and supplied to different sectors like kerosene for households, diesel for industry and commerce, gasoline and diesel for transportation. The location of the study area is shown in Figure 1.



**Figure 1** | Location map of Addis Ababa city in Ethiopia.

## Methodological framework

This paper considers socio-economic perspectives that affect WE demand. The socioeconomic variables, including population, GDP and PCI, are included in an LR model to identify the factors affecting WE consumption. The method to enable determining the WE demand is framed as shown in Figure 2. The first step in forecasting the WE demand is correlating the independent socio-economic variables (GDP, population and PCI) with the dependent variable consumption or demand using the WEKA tool. The tool uses stochastic relation and relates the independent variables  $x$  with dependent variables  $y$ . This will first give consumption by sectors (residential, commercial, industries and street lightings).

WEKA as a modeling tool was used to generate different equations for all considered sectors to predict the WE consumption of the city. The driving scenarios quantify the longer-term values of predictors which were identified during the LR modeling. Each scenario produced its own set of sector consumption forecasts and adjusted for estimated WE losses, to make forecasts for the entire annual demand at each sector. The advantage of having the ability to forecast for every sector during a scenario is that one can assess the relevance and compatibility of the model corresponding to the scenario.

## Data used

Prediction accuracy strongly depends on the quality of available historical data. A poor history, composed only of anomalous or average events, may polarize the analysis

and affect the quality of the prediction values. For this study, data from different offices were collected and analyzed. Socio-economic data including GDP and population were collected from the Bureau of Finance and Economic Development (BoFED) and the Central Statistical Agency (CSA) respectively whereas the water and electric energy consumption data were collected from Addis Ababa Water and Sewerage Authority (AAWSA) and Ethiopian Electric Utility (EEU) respectively. Additionally, petroleum (kerosene, diesel and gasoline) energy consumption was gathered from Ethiopian Petroleum Enterprise (EPE).

Descriptive statistics of the original data with five data subsets, including mean ( $\mu$ ), standard deviation ( $S_x$ ), minimum ( $X_{min}$ ), and maximum ( $X_{max}$ ) and coefficient of variation (CoEV) show statistical characteristics of data (Ebru Eris 2019). For comparability, the annual WE consumption data were calculated in the unit of Peta Joule (PJ) for petroleum, GWh for electric energy and MCM for water. Characteristics of the data used are indicated in Table 1.

The distribution characteristics were also indicated. There are no unique and universally accepted probability distribution functions for fitting data. Some of the different forms of characteristic functions frequently used to show the probability distribution of WE consumption are Weibull, Gumbel, Pearson Type III, and log-normal distributions (Ebru Eris 2019). Weibull, Generalized Extreme Value (GEV) and Log-Pearson Type III (LP3) distributions were used to fit the WE consumption data after being ranked with statistical test results such as Kolmogorov–Smirnov and Anderson–Darling, see Table 2.

Based on the statistical distribution, all trends of water end-users are fitted by GEV, except industrial water

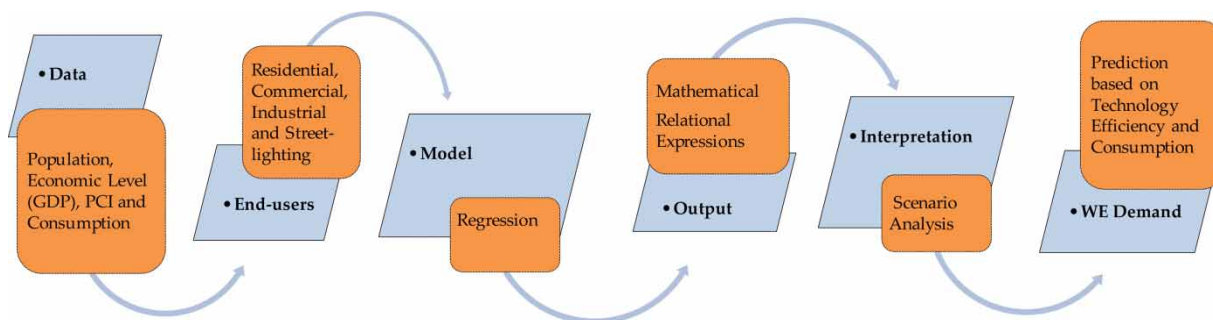


Figure 2 | A framework to predict the WE demand.

**Table 1** | Statistical characteristics of annual WE consumption data

| WE consumption    | Sectors                  | $X_{\min}$ | $X_{\max}$ | $\mu$ | $S_x$ | CoeV |
|-------------------|--------------------------|------------|------------|-------|-------|------|
| Energy (electric) | Commercial               | 556        | 1,167      | 1,139 | 242   | 83   |
|                   | Industrial               | 1,139      | 1,889      | 1,472 | 333   | 56   |
|                   | Residential              | 889        | 1,667      | 1,278 | 306   | 67   |
|                   | Street-lighting          | 8          | 17         | 14    | 3     | 67   |
| Diesel            | Industrial and transport | 6.5        | 11.8       | 8.7   | 1.9   | 0.22 |
| Gasoline          | Transport                | 4.6        | 13.3       | 7.8   | 3.2   | 0.4  |
| Kerosene          | Residential              | 3.4        | 3.7        | 3.6   | 0.12  | 0.03 |
| Water             | Commercial               | 30         | 60         | 40    | 10    | 240  |
|                   | Industrial               | 30         | 50         | 40    | 10    | 230  |
|                   | Residential              | 120        | 240        | 170   | 40    | 250  |

**Table 2** | Distribution characteristics for observed WE consumption

| WE consumption    | Sectors                  | Distribution | Kolmogorov Smirnov |      | Anderson Darling |      |
|-------------------|--------------------------|--------------|--------------------|------|------------------|------|
|                   |                          |              | Statistic          | Rank | Statistic        | Rank |
| Water             | Industrial               | Weibull      | 0.21               | 1    | 1.01             | 1    |
|                   | Commercial               | GEV          | 0.2                | 1    | 0.8              | 1    |
|                   | Residential              | GEV          | 0.08               | 1    | 0.16             | 1    |
| Energy (electric) | Commercial               | GEV          | 0.14               | 1    | 0.16             | 2    |
|                   | Industrial               | GEV          | 0.2                | 1    | 0.27             | 1    |
|                   | Residential              | GEV          | 0.13               | 1    | 0.14             | 2    |
|                   |                          | LP3          | 0.14               | 2    | 0.15             | 1    |
| Diesel            | Industrial and transport | GEV          | 0.12               | 1    | 0.17             | 1    |
| Gasoline          | Transport                | LP3          | 0.13               | 1    | 0.21             | 2    |
|                   |                          | GEV          | 0.14               | 2    | 0.22             | 1    |
| Kerosene          | Residential              | GEV          | 0.18               | 1    | 0.31             | 1    |

consumption which has Weibull characteristics. As well as energy consumption data also characterized by GEV distribution, except residential energy consumption and diesel which have properties of both GEV and LP3 distribution.

To include the effects of the driving parameters, the prediction of the PCI, GDP, population and technology factor (WE efficiency) from 2016 to 2050 has to be determined first. The following population growth rate (%) data scenarios are taken into consideration to predict the population. Scenario 1 (Business as usual): The population growth rate of 3% (CSA 2015); this scenario assumes that despite Addis Ababa's economic boom that has attracted rural-urban and urban-urban migrants, the surge in secondary cities' population growth could ease some migration pressure away from Addis Ababa and keep its population growth at the current pace. Scenario 2 (Rapid population growth rate): The urban

population growth rate is just driven by demographic factors (death and birth rates and migration), but that it will also potentially be significantly influenced by policies such as megaprojects. Considering these factors, various estimates come to a 5.2% (MUDHCo 2015) urban population growth rate for Ethiopia and nation-wide. Scenario 3 (Lower population growth rate): The surging role of secondary cities (Mekelle, Hawasa, and Bahir Dar) in Ethiopia could have an impact on Addis Ababa by lowering population growth rates to 2.5% (UN HABITAT 2017a, 2017b).

Similar to the population forecast, the current GDP and GDP growth rate (%) were taken into consideration. According to the city revenue study, Growth and Transformation Plan II (GTP II) target, the GDP growth rate of the city is analyzed based on the following scenarios (UN HABITAT 2017a, 2017b):

Scenario 1 (Business as usual): This indicates a constant city GDP growth rate of about 18% annually; although more than 10% GDP growth annually by itself would be a major achievement and it is assumed that during the GTP II period the city could indeed repeat its 18%. However, as per the city's revenue study in 2015, there remain many untapped sources of municipal income, policy interventions, infrastructure facilities, and investments that will need to be activated if Addis Ababa is to realize the projected growth rate of 18%.

Scenario 2 (Lower GDP growth rate): The GDP growth rate of about 11% annually (planned by the city administration). After 2017, the city's GDP was expected to grow yearly by 11%. For the GTP II growth target, the city administration aims to reduce the pace of growth downwards from its 18% present level to 11% annually.

Technological factors: The high losses in the distribution system occur due to inefficient technology capacity. At present, the city governance is giving due attention to the improvement of energy efficiency. The electric energy loss in a distribution system is indicated in Table 3.

Water loss due to distribution technologies is a serious problem in Addis Ababa city causing both severe water shortage and huge financial losses. Therefore, water loss is considered in the estimation of water demand. AAWSA water loss analysis indicated the value of non-revenue water (NRW) ranges from 35 to 45% (2010–2015). NRW will decrease to 23% in 2030 (AAWSA 2019).

### WEKA data mining tool

The WEKA data mining tool provides a uniform interface to different learning algorithms, along with methods for pre- and post-processing and for evaluating the result of learning schemes on any given dataset. One way of using WEKA is to apply a learning method to a dataset and analyze its output

to learn more about the data, to use learned models to generate predictions on new instances and to apply several different learners and compare their performance to choose one for prediction. The learning methods are classifiers, and in the interactive WEKA interface, one of the wanted classifiers is selected. Many classifiers have tunable parameters, which are accessed through a property sheet. A common evaluation module is used to measure the performance of all classifiers (Ian 2005).

The WEKA tool contains a comprehensive set of useful algorithms to support data mining tasks. These include tools for data engineering (filters), algorithms for attribute selection, clustering, association rule learning and classification. Implementations of almost all mainstream classification algorithms are included such as Bayesian methods including naive Bayes; complement naive Bayes, multinomial naive Bayes, and Bayesian networks. Implementation of the tool includes MLR, simple linear regression (SLR), multi-layer perception (MLP) and support vector regression (SVR). The standard data mining framework that should be followed in various applications of data mining states that before data processing, selection and pre-processing should be performed to obtain clean and complete data records for processing. Then, the transformation of data represents a data extension conversion or compatibility conversion of data shape to an equivalent data mining tool format to perform data mining algorithm implementation and generate the desired knowledge outcome. The data-mining framework is shown in Figure 3.

### Regression model-based WE demand prediction

Every discipline utilizes LR analyses as a basis for comparing other models (Leng *et al.* 2017). The LR analysis is one of the foremost widely used methodologies for expressing the dependence of a response variable on several independent variables (Abdul-Wahab *et al.* 2005). The LR works on any size of datasets and provides information about the relevance of the features. Moreover, LR equations are considered consistent with two criteria: the equation had to be cogent (i.e. it had to supply a fundamental explanation for an observed trend), and it needed to possess high goodness of fit (Lei & Wang 2008). The primary step in LR analysis is to consider independent variables for

**Table 3** | Electric energy distribution loss (JICA 2018)

| Energy loss   | Unit | Year |      |
|---------------|------|------|------|
|               |      | 2017 | 2034 |
| Technical     | %    | 13   | 8    |
| Non-technical | %    | 3    | 1    |
| Total         | %    | 16   | 9    |



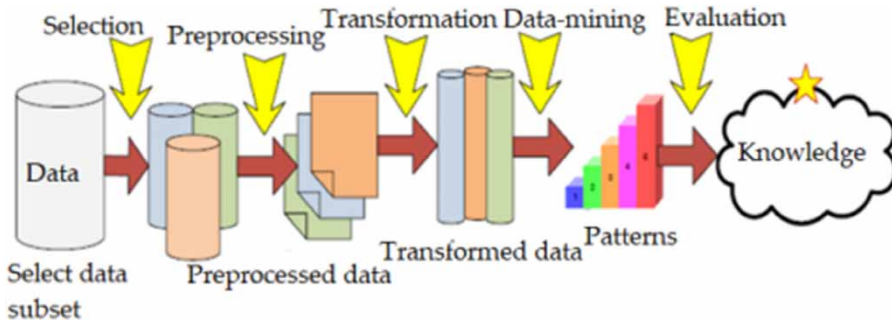


Figure 3 | Data-mining (DM) framework.

constructing a model. Here, the important peculiarities are: (1) to select out adequate dependent variables, (2) to exist linear cause–result relationship between a variable and independent variables, (3) to incorporate only relevant independent variables within the model. While handling a sizeable amount of independent variables, its importance is to work out the simplest combination of those variables to predict variables (Cevik 2007).

The flowchart of a regression model to forecast a long-term WE demand is shown in Figure 4.

The variables and the final governing LR equations should be obtained as part of the model results to predict the WE demand. Since the model was once fixed as linear, either MLR or SLR equations can be used for

demand forecasting. The SLR model describes the relationship between one input variable  $x$  and the output variable  $y$  and is given by Equation (1):

$$y = f(x, a_0, a_1) = a_0 + a_1x \tag{1}$$

when geometrically interpreted, the linear line shows the relationship between  $y$  and  $x$ . The shape of the straight line is determined by the regression parameters  $a_0$  and  $a_1$ . For given measurements  $x_1, x_2, \dots, x_n$  and  $y_1, y_2, \dots, y_n$  of the variables  $x$  and  $y$ , the parameters are calculated such that the mean quadratic distance between the actual  $y_i$  ( $i = 1, \dots, n$ ) and the predicted values  $\hat{y}_i$  on the straight line is minimized or minimum. This shows the optimization

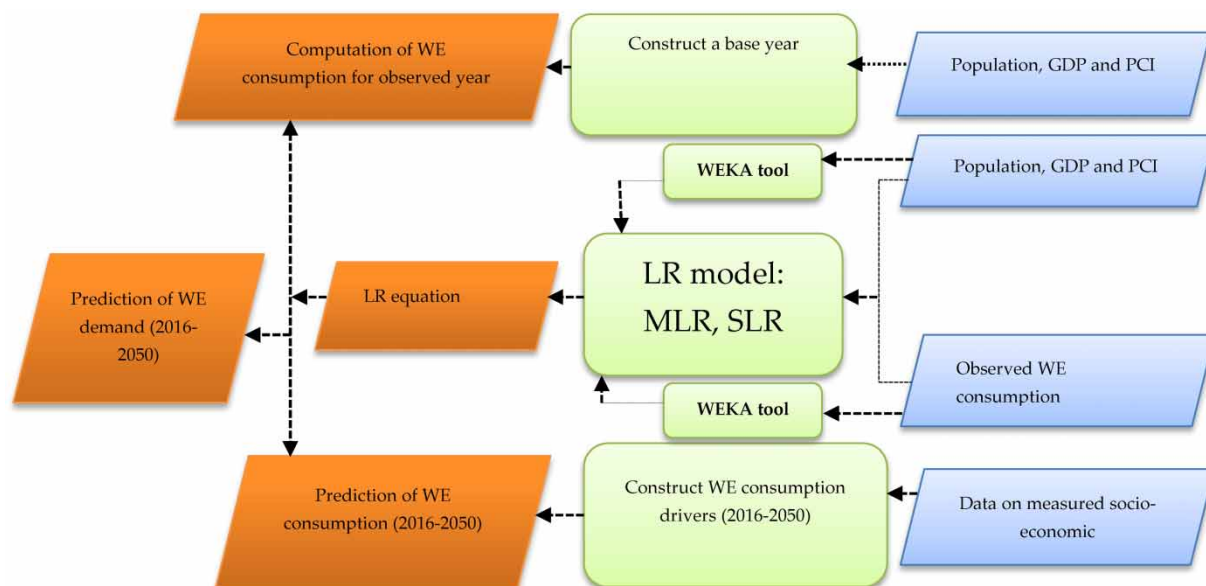


Figure 4 | Diagram illustrating the proposed LR model procedure.

problem is computed as Equation (2):

$$Q(a_0, a_1) = \sum_{i=1}^n (y_i - f(x_i, a_0, a_1))^2 \rightarrow \text{Min}_{a_0, a_1} \quad (2)$$

The calculated regression parameters represent a least squares estimation of the fitting problem. The regression model can extend to multiple linear relationships where the output variable  $y$  is influenced by inputs  $x_1, x_2, \dots, x_p$ . The MLR models describe the relationship between more than one input variable  $x$  and the output variable  $y$ , which is given by Equation (3):

$$y = f(x, a) = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_px_p \quad (3)$$

where  $a_1, \dots, a_n$  and  $a_0$  are regression coefficients and constant respectively and  $x_1, \dots, x_n$  are independent variables (population, GDP and PCI), whereas  $y$  is the dependent variable (WE consumption).

The regression notation is given by Equation (4) (Xin Yan 2009):

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \quad (4)$$

where  $y$  contains the output variable,  $a$  represents the vector of the regression parameters, and the matrix  $x$  contains the value of  $x_{ij}$  of the  $i$ th observation of the input  $x_j$ .

The LR model is fast, reliable, and simple to implement and provides the importance of each predictor variable and the uncertainty of the regression coefficients. Furthermore, the results are relatively robust. The LR analysis is carried out for each seasonal cluster following the algorithm indicated in Figure 5.

### Basic assumptions of regression analysis

The LR model requires checking the assumptions about the data that will be included in the model to verify that the regression model will be robust. The linearity and multicollinearity of the data should be checked first before developing the model.

In LR analysis, there are assumptions about the model which can include multicollinearity and linearity (Waters 2002). However, multicollinearity will cause a problem when it is moderate or high. Multicollinearity is a statistical phenomenon in which two or more explanatory variables in a LR model are highly correlated (Ramirez 2013). This shows that the relationship between the independent and dependent variables is distorted and regression cannot be computed (George & Dallas 2002). Multicollinearity is identified by examining the tolerance for each predictor variable. Tolerance is the amount of variability in one independent variable that is not explained by the other independent variables, and it is  $1 - r^2$ . Tolerance values less than 0.1 indicate collinearity of socio-economic variables. The Variance

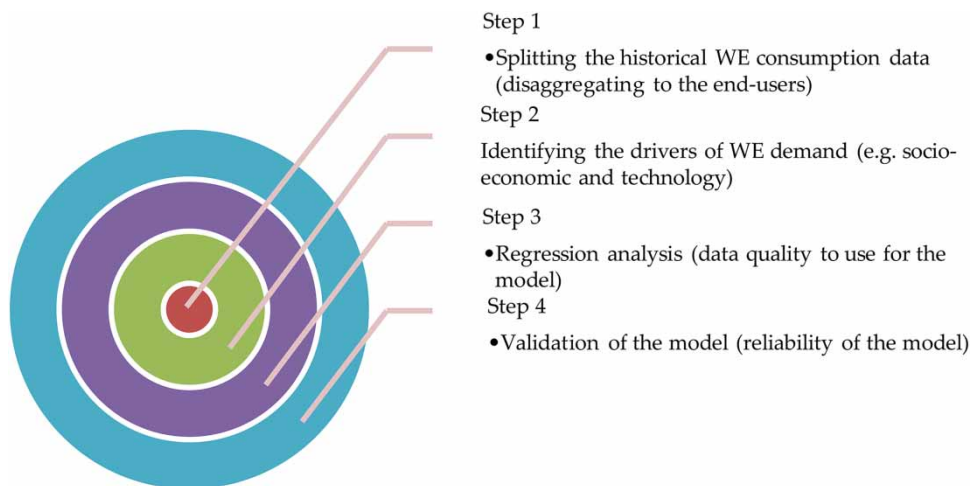


Figure 5 | Algorithm steps in the regression model.

Inflation Factor (VIF) also measures the multicollinearity of independent (explanatory) variables. Studies have shown that if  $VIF \geq 10$ , it indicates the presence of multicollinearity between explanatory variables (Debbie & Maria-Pia 2013). However, the WEKA tool used in this study by itself eliminates the multicollinearity variable while the LR model is developed. The VIF formula is given in Equation (5).

$$VIF = \frac{1}{1 - r^2} \quad (5)$$

In this study, most of the independent variables (socio-economic drivers) have a value of greater than 0.1 for  $1 - r^2$  and VIF of less than 10. This shows that the explanatory variables can be implemented in a regression model for the prediction of WE demand. The scenarios used in the model are justified based on the multicollinearity test as indicated in Table 4.

The tolerance of Test 3 is 0.03, which is much lower than the acceptable value. Hence, the results of this test are not considered for use as a driver in the LR modeling. Therefore, the combined driver (population and GDP) is eliminated from the analysis when forecasting WE demand considering a LR model.

Linearity requires the dependent variable to be a linear function of the independent variables. Therefore, linearity is the most important assumption in any LR model as it directly relates to the bias of the results of the whole analysis (Balance 2012). If linearity is violated, all the estimates of the regression and its statistical output may be biased resulting in serious error in the predicted values (Balance 2012). On

the other hand, when a linear relationship exists between the dependent and the independent variables, the SLR or MLR can accurately estimate the dependent variable (Balance 2012). The linearity between the dependent variable, independent variable, and as well as the dependent variable with the independent variable, is evaluated using correlation coefficient as shown in Table 5.

To use the driver's data in an LR model, the linearity between explanatory and dependent variables are considered for LTF of WE demand. The relationship between explanatory and dependent variables reflected their positive correlation (greater than 0.41).

### Sensitivity analysis and input variable selection

The selection of explanatory variables is important in the modeling of LR methods because the input parameters determine both the equation form and the regression coefficient values that affect the output variable (Moslem et al. 2019). This is widely recognized by the scientific community and input data selection is applied in many works (Lahouar & Le 2015). The Pearson and Spearman correlation coefficients were applied to identify the strongest correlation between the explanatory variable and dependent variable (Dimitrios et al. 2017). The parameters that mostly influence the WE demands of the end-users (sectors) are studied; the authors applied the Pearson correlation coefficient ( $r$ ) analysis. This method is used to relate the explanatory variables and the dependent variable. It is one of the simplest and fastest methods for selecting and identifying the most influential input parameter (variables) useful for the LR model (Moslem et al. 2019). It is a dimensionless index that ranges from  $-1$  to  $1$  with  $1$  corresponding to ideal correlation. Given two statistical variables (WE consumption and socio-economic factors), the Pearson correlation  $r$  coefficient is defined as the ratio between the covariance of the two variables and the standard deviation of each as indicated in the following Equation (6):

$$r = \frac{CV(x, y)}{\sigma_x \sigma_y} \quad (6)$$

where  $CV(x, y)$  indicates the covariance between the  $x$  and the  $y$  variables, while  $\sigma_x$  and  $\sigma_y$  are their respective standard deviations

**Table 4** | Multicollinearity test of the predictor data

| Multicollinearity test | Scenarios/drivers/factors |       |
|------------------------|---------------------------|-------|
| Test 1                 | GDP                       | PCI   |
| Tolerance              | 0.31                      | 0.31  |
| VIF                    | 3.18                      | 3.18  |
| Test 2                 | Population                | PCI   |
| Tolerance              | 0.29                      | 0.29  |
| VIF                    | 3.44                      | 3.44  |
| Test 3                 | Population                | GDP   |
| Tolerance              | 0.03                      | 0.03  |
| VIF                    | 29.25                     | 29.25 |

**Table 5** | Linearity test using correlation coefficient between the variables

|                 | Driver     |      |      | Petroleum energy          |            |            |
|-----------------|------------|------|------|---------------------------|------------|------------|
|                 | Population | GDP  | PCI  | Gasoline                  | Kerosene   | Diesel     |
| Population      | 1.00       | 0.98 | 0.84 | 0.99                      | 0.63       | 0.99       |
| GDP             | 0.98       | 1.00 | 0.85 | 0.96                      | 0.68       | 0.96       |
| PCI             | 0.84       | 0.85 | 1.00 | 0.87                      | 0.41       | 0.87       |
| Gasoline        | 0.99       | 0.96 | 0.87 | 1.00                      | 0.59       | 1.00       |
| Kerosene        | 0.63       | 0.68 | 0.41 | 0.59                      | 1.00       | 0.59       |
| Diesel          | 0.99       | 0.96 | 0.87 | 1.00                      | 0.59       | 1.00       |
|                 | Driver     |      |      | Water end-users           |            |            |
|                 | Population | GDP  | PCI  | Residential               | Commercial | Industrial |
| Population      | 1.00       | 1.00 | 0.94 | 1.00                      | 0.90       | 0.97       |
| GDP             | 0.99       | 1.00 | 0.93 | 0.99                      | 0.90       | 0.97       |
| PCI             | 0.93       | 0.93 | 1.00 | 0.94                      | 0.92       | 0.93       |
| Residential     | 1.00       | 0.99 | 0.94 | 1.00                      | 0.90       | 0.97       |
| Commercial      | 0.90       | 0.90 | 0.92 | 0.90                      | 1.00       | 0.89       |
| Industrial      | 0.97       | 0.97 | 0.93 | 0.97                      | 0.89       | 1.00       |
|                 | Driver     |      |      | Electric energy end-users |            |            |
|                 | Population | GDP  | PCI  | Street-lighting           | Commercial | Industrial |
| Population      | 1.00       | 1.00 | 0.99 | 0.93                      | 1.00       | 0.99       |
| GDP             | 1.00       | 1.00 | 0.98 | 0.92                      | 1.00       | 0.99       |
| PCI             | 0.99       | 0.98 | 1.00 | 0.91                      | 0.99       | 0.97       |
| Street-lighting | 0.93       | 0.92 | 0.91 | 1.00                      | 0.93       | 0.88       |
| Commercial      | 1.00       | 1.00 | 0.99 | 0.93                      | 1.00       | 0.99       |
| Industrial      | 0.99       | 0.99 | 0.97 | 0.88                      | 0.99       | 1.00       |

The need for sensitivity parameter analysis (SA) is an important element to determine how the output value changes when the input is varied (Kamínski *et al.* 2018). The chosen approach to determine the influential input parameter involves using the LR model. The CoD used to evaluate the performance of each input is determined to achieve the best output.

### Performance evaluation indices

For evaluating the performance of the model, the best model evaluation parameter includes the MAE, CoD, relative absolute error (RAE), relative squared error (RSE), and RMSE (Miller & Kane 2001). RMSE is sensitive to

forecasting errors for high data values, and this can be good performance indices for high data values. In contrast, MAE is evaluated based on all deviations without being weighted to lower and higher data values. MAE and RMSE have the advantage that they can represent the size of a typical error effectively since they are evaluated in the same unit as the original data. The RMSE and MAE are some of the most widely used approaches to test and validate model outcomes. For a perfect model, MAE and RMSE would be zero (Willmott & Matsuura 2005). MAE and RAE belong to relative errors. RAE is a good index for low data values since it is more affected by forecasting errors for low data values. It has the advantage that it is less sensitive to skewed error distributions and outliers.

For a perfect model, MAE and RAE would be zero (Willmott & Matsuura 2005). The indices are discussed as follows.

RMSE: is the square root of mean squared error and it is given by Equation (7):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - P_i)^2} \quad (7)$$

RMSE is the standard deviation of the model errors. First, the difference between an actual value and the predicted value is calculated, and then the difference is squared and averaged over all data items before the root of the mean value is computed.

MAE is calculated as the average of absolute differences between the actual values and model predicted values. This is expressed in Equation (8).

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - P_i| \quad (8)$$

RAE: The percentage error is a useful tool for determining the precision of the calculations and the model. It is the relative absolute difference between actual and forecasted values which is given in Equation (9).

$$RAE = \frac{\sum_{i=1}^n |P_i - A_i|}{\sum_{i=1}^n |A_i - \bar{A}|} \quad (9)$$

CoD shows the performance of the predicting model where zero means the model is random while 1 means there is a perfect fit, see Equation (10).

$$CoD = 1 - \frac{\sum_{i=1}^n (P_i - A_i)^2}{\sum_{i=1}^n (A_i - \bar{A})^2} \quad (10)$$

where  $A_i$  is the actual value,  $P_i$  is predicted value and  $\bar{A}$  is the mean of the actual value.

RSE: This normalizes the entire squared error of the predicted values, see Equation (11).

$$RSE = \sqrt{\frac{\sum_{i=1}^n (A_i - P_i)^2}{\sum_{i=1}^n |A_i - \bar{A}|}} \quad (11)$$

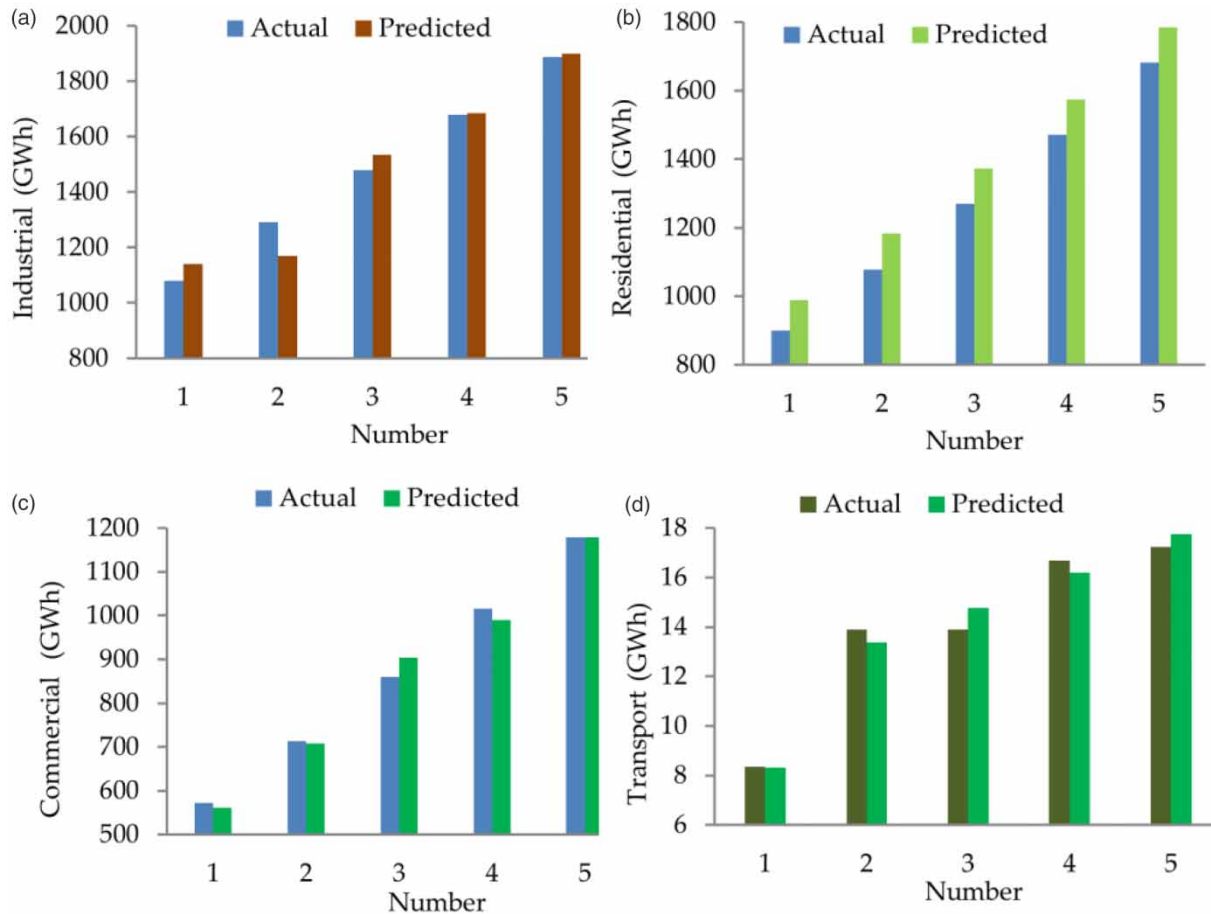
## RESULTS AND DISCUSSION

### Model evaluation

To ensure the strength and reasonableness of the model, evaluation analyses are employed. If there is no significant change in the behavior of the WE consumption with some perturbation in some system parameters, the model is relatively robust. The model evaluation indices are used to verify whether the results of the model are fitting with the historical actual values. As MAE and RMSE are results of an absolute error indicating the comparison between the actual and forecasted WE demand, no absolute criterion exists for both measures reliability. Therefore, a smaller value of MAE and RMSE indicates that the forecasted WE demand during the validation period was closer to the actual values, relative to the other driving factors (explanatory variables). The proposed LR model showed superior performance for annual RAE by achieving the lowest RAE value of less than or equal to 10%. When a RAE value is less than 10% corresponding to each year, it is interpreted as highly accurate forecasting (Lewis 1982) or acceptable forecasting accuracy (Hamzacebi & Es 2014), and when a RAE value is less than 10%, it is interpreted as near-perfect forecasting (Hamzacebi & Es 2014). So, the LR model was proven to yield superior performance in terms of RAE. The predicted values of electric energy demand based on the main socio-economic drivers have been plotted versus the actual values to visualize the difference between the two values. The measured and predicted electric energy consumption are illustrated in Figure 6.

### Socio-economic analysis

Socio-economic factors are the main factors that influence future urban WE consumption. The fast population growth



**Figure 6** | Measured and predicted electric energy consumption.

rate is due to the economic attractiveness of the city compared to the rural areas. The city population growth rate estimated at par with the national urban growth under the second scenario (high population growth rate) (MUDHC 2015) is reasonable.

According to AAWSA, population prediction for a high growth rate scenario is 6.3 million in 2030 and there is a 7% gap with this study due to the difference in growth rate value. The population projected for the three scenarios is shown in Table 6.

The population of Addis Ababa is expected to increase in the next 15 years at an average annual growth rate of approximately 4%, reaching almost 9 million in 2035 (UN 2018). The city population is expected to be around 5.3 million in 2025 (Un-Habitat 2008) and this number is expected to rise up to 5.3 and 6.3 million in 2025 and 2030, respectively (CSA 2006; UNESCO 2006). The results

**Table 6** | Addis Ababa city projected population in millions

| Scenario                    | Year |      |      |      |       |       |      |
|-----------------------------|------|------|------|------|-------|-------|------|
|                             | 2016 | 2025 | 2030 | 2035 | 2040  | 2045  | 2050 |
| Business as usual (BAU)     | 3.53 | 4.47 | 5.18 | 6.01 | 6.97  | 8.08  | 9.36 |
| High population growth rate | 3.53 | 5.29 | 6.8  | 8.79 | 11.32 | 14.59 | 18.8 |
| Low population growth rate  | 3.53 | 4.3  | 4.86 | 5.05 | 6.23  | 7.05  | 7.97 |

shown in Table 6 do not have significant variation with these results. The World Bank (2015a, 2015b, 2015c) forecasted that in 2040 the population will be around 10.3 million. The city GDP has grown at an annual average of 11% (UN-Habitat 2017a, 2017b), which is used as a driver in WE demand estimation. The two scenarios of GDP growth rate data were taken from analysis of GTP II

growth target and the City revenue study (UN-Habitat 2017a, 2017b).

As indicated in Table 7, the GDP estimated under low GDP growth rate is lower by 40 and 12% as compared to BAU growth rate for the years 2030 and 2050 respectively. The reason for this difference is the GDP growth rate variation. However, high growth of GDP increases the WE consumption whereas the decrease in GDP growth activity caused WE consumption to decline (Nasir & Ur Rehman 2011). The annual economic growth rate was assumed to be 11%.

Moreover, this study intended to understand the impact of socio-economics on WE consumption by changing socio-economic drivers (such as population, GDP and PCI). In the following, from the perspective of indicators or drivers, the changes of variables in the different scenarios were compared with the results in the base period (historical).

### Scenario in water consumption

The prediction methodology involved applying the quantified (statistical) relationships, identified from historical patterns, to different future scenarios concerning expected economic and demographic changes, to achieve the resulting forecasts for water usage. Since the forecasts are

directly linked to scenarios (GDP and population) and other predictor variables (population and PCI) or (GDP and PCI). The forecasted water values have to be interpreted within the context of those scenarios. For instance, if a specific scenario was compiled using GDP and PCI growth figures that were very far away from the particular growth patterns observed, then the water forecasts generated for this scenario can also be unrealistically high. While scenario thinking can support planning and forecasting well, there are certain pitfalls to avoid when generating scenarios (Sayim 2013). A statistical evaluation of scenarios was computed as indicated in Table 8.

Based on the value of CoD, MAE, RMSE, RAE and RSE, commercial and residential water consumption is more influenced by population and PCI variables, whereas the industrial sector is affected by GDP and PCI explanatory variables. The GDP and PCI are highly significant in explaining industrial water consumption. A considerable large effect of GDP and PCI was noted to influence the industrial water consumption. If the city GDP increases by 11%, then industrial water consumption will increase by 4%. The regression coefficients and constants used to fit the water consumption by sectors with scenarios are tabulated in Table 9.

The increase in GDP was seen as a positive indicator for city development. However, the increase in water

**Table 7** | Projected GDP (Billion ETB) of Addis Ababa city

| Scenario                | Year |        |          |          |          |           |           |
|-------------------------|------|--------|----------|----------|----------|-----------|-----------|
|                         | 2016 | 2025   | 2030     | 2035     | 2040     | 2045      | 2050      |
| Business as usual (BAU) | 107  | 474.59 | 1,085.76 | 2,483.95 | 5,682.66 | 13,000.56 | 29,742.13 |
| Low GDP growth rate     | 101  | 258.36 | 435.36   | 733.6    | 1,236.16 | 2,082.99  | 3,509.96  |

**Table 8** | Evaluation of drivers for water consumption prediction

| Sectors     | Scenario | Drivers            | Parameters |      |      |         |         |
|-------------|----------|--------------------|------------|------|------|---------|---------|
|             |          |                    | CoD        | MAE  | RMSE | RAE (%) | RSE (%) |
| Commercial  | 1        | Population and PCI | 0.96       | 2.40 | 2.80 | 31.00   | 31.00   |
|             | 2        | GDP and PCI        | 0.96       | 2.60 | 2.90 | 33.00   | 32.00   |
| Industrial  | 1        | Population and PCI | 0.76       | 2.10 | 2.60 | 33.00   | 34.00   |
|             | 2        | GDP and PCI        | 0.78       | 2.10 | 2.50 | 32.00   | 34.00   |
| Residential | 1        | Population and PCI | 0.94       | 2.60 | 3.10 | 7.80    | 7.90    |
|             | 2        | GDP and PCI        | 0.92       | 3.00 | 4.00 | 10.20   | 10.50   |

**Table 9** | Estimated values of the regression coefficient and constant

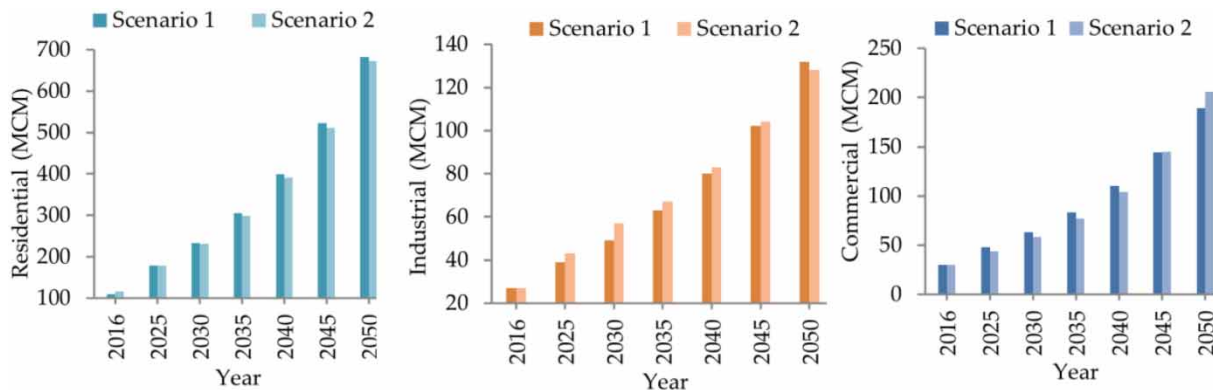
| Sector      | Scenario | Regression coefficient | Values of coefficients | Constant | Values of constant |
|-------------|----------|------------------------|------------------------|----------|--------------------|
| Commercial  | 1        | $a_1$                  | 0.006                  | $a_0$    | 0.003              |
|             |          | $a_3$                  | 0.2                    |          |                    |
|             | 2        | $a_2$                  | 0.05                   | $a_0$    | 0.02               |
|             |          | $a_3$                  | 0.3                    |          |                    |
| Industrial  | 1        | $a_1$                  | 0.005                  | $a_0$    | 0.006              |
|             |          | $a_3$                  | 0.1                    |          |                    |
|             | 2        | $a_2$                  | 0.05                   | $a_0$    | 0.02               |
|             |          | $a_3$                  | 0.2                    |          |                    |
| Residential | 1        | $a_1$                  | 0.03                   | $a_0$    | -0.007             |
|             |          | $a_3$                  | 0.3                    |          |                    |
|             | 2        | $a_2$                  | 0.3                    | $a_0$    | 0.06               |
|             |          | $a_3$                  | 0.9                    |          |                    |

Note:  $a_1$ ,  $a_2$  and  $a_3$  are coefficients of the independent variable for population, GDP and PCI, respectively and water consumption is in billion cubic meters (BCM), population (billion), GDP (1000 billion ETB), and PCI (1000 ETB per capita).

consumption can be seen as a negative indicator as well. The results show the causality running from water consumption to GDP. The population and PCI mainly affect the water consumption of commercial and residential end-use in Addis Ababa city. The consumption is positively affected by the rise of the population and PCI. Moreover, GDP and PCI that is positively affected by the increase of both GDP and PCI affects industrial water consumption. The commercial water consumption will increase by 5% with an increase in 6% of PCI value, holding other factors constant. Also, note that residential water consumption will rise by 5% annually if the PCI increases by around 5%. This estimated increase is not especially small. This shows these consumption growths are directly proportional to population growth. The predicted water consumption (2016–2050) for the end-users is shown in Figure 7.

## Water demand

The technology performance indicator, such as water efficiency, in the distribution system affects water demand. NRW should be less than 25% according to the World Bank recommendations. NRW for the baseline is about 45% (2016) (AAWSA 2019) and reaches 22% in 2050, the real loss decreases by 75% of NRW. Therefore, by decreasing NRW, the real loss also decreases to 17.5% for NRW of 23%, and 16.75% for NRW of 22%. However, a baseline real loss is considered in this study, which is 75% of NRW (33.6%). The distribution system needs technology improvement and it needs a plan by AAWSA for water loss reduction. Scenario 1 (population and PCI) is a paramount independent variable for commercial and residential water consumption prediction due to its accuracy relative to the

**Figure 7** | Water consumption based on socio-economic drivers.



other scenarios considered. Similarly, for industrial water consumption, scenario 2 (GDP and PCI) is considered as the most influential parameter. The study reveals the fact that population and PCI are the factors that could be considered as important factors to affect the water demand of the commercial and residential sectors, whereas GDP and PCI are the main factors for industrial water demand. Using these factors, the equation used to determine the water consumption is given in Table 10, and as well as considering the water loss the predicted value of water demand is given in Table 11.

Due to the high population growth, the total water demand is reflected to be 580 MCM in 2030 (Rooijen 2011). Consequently, for a high population and low GDP growth rate, this study has estimated the total water demand of 518 MCM in 2030. From the results of this study, in 2030 and 2050, the water demand of the city is expected to reach 518 and 1568 MCM respectively. The percentage share of water demand in the residential, commercial, and industrial sectors in year 2030 were estimated to be 68, 17 and 15% respectively whereas these

values in 2050 were estimated to be 64, 15 and 21% respectively.

The residential water demand in the city in 2030 and 2050 will be 149 and 158 liters per capita per day (LPCD) respectively, which is greater than the World Health Organization recommend (110 LPCD). The present study of the total per capita water demand in 2025 is 221 LPCD, which is insignificantly underestimating 229 LPCD (AAWSA 2008).

The study results showed that the city's water demand would be around 886 MCM in 2040, which is lower compared to the study results of Mengistu (2010). In 2050, the total per capita water use of Addis Ababa city for an expected population of 18 million is around 230 LPCD. This is high as compared to the international average water use of 173 LPCD (Rookmoney et al. 2019). Ethiopia is one of the developing countries and the government has a long-term plan to be a middle-income country, and this indicates that the PCI (socio-economic measures) rises yearly. Consequently, it embraces the domestic water demand that would continue to rise to meet the drinking water needs of an increasingly urban population. The percent growth rate of water demand in commercial, industrial and residential sectors is shown in Figure 8.

**Table 10** | Variables and the final governing LR equations to predict the water demand

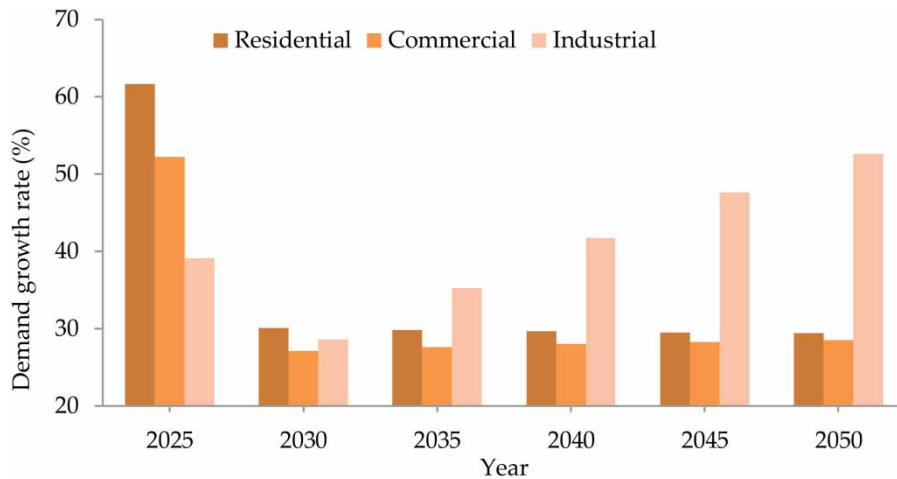
| Notation | Transport electric energy | Regression | Mathematical relational expressions         |
|----------|---------------------------|------------|---|
| COMWAT   | Commercial water          | MLR        | COMWAT = 0.006*Population + 0.2*PCI + 0.003 |
| INDWAT   | Industrial water          | MLR        | INDWAT = 0.05*GDP + 0.2*PCI + 0.02          |
| RESWAT   | Residential water         | MLR        | RESWAT = 0.03*Population + 0.3*PCI - 0.007  |

**Table 11** | Water demand (MCM) prediction result

| Sectors     | Year |      |      |      |      |       |       |
|-------------|------|------|------|------|------|-------|-------|
|             | 2016 | 2025 | 2030 | 2035 | 2040 | 2045  | 2050  |
| Residential | 168  | 272  | 354  | 459  | 595  | 771   | 997   |
| Commercial  | 45   | 68   | 87   | 111  | 142  | 182   | 233   |
| Industrial  | 44   | 61   | 78   | 105  | 150  | 221   | 337   |
| Total       | 256  | 401  | 518  | 675  | 886  | 1,173 | 1,568 |

## Scenarios in energy consumption

To predict the city energy demands, the most influential variables were conducted using CoD, MAE, RAE, RMSE and RSE parameters. Scenario or input variables that showed least values of MAE, RAE, RMSE and RSE are used for demand prediction. Based on these values, in the street-lighting sector, energy consumption was estimated for scenarios such as scenario 1: (Population and PCI), and scenario 2 (Population). Scenario 1 is more realistic to the energy consumption for street lighting relative to scenario 2 and is used for the analysis of the future demand. In the cases of commercial, residential and industrial sectors, the energy consumption was estimated based on scenario 1 (Population) and scenario 2 (PCI). Both scenarios were given nearly the same value and the average of these results is expected to be the future energy consumption. The fuel energy demand of kerosene is based on population and GDP, whereas population and PCI influence diesel and



**Figure 8** | Growth rate of water demand for each end-user.

gasoline demand. The evaluation of the scenario results from the regression model for energy consumption is indicated in Table 12.

The population and PCI mainly affect the energy consumption of industrial, street-lighting, commercial and residential end-use. The consumptions are positively affected by the rise of the population and PCI, except street-lighting is negatively affected by the increase of population and positively affected by the PCI. Similarly, a stronger relationship exists between the population and the city electric energy demand (Sailor & Muñoz 1997).

The energy regression results illustrate that the coefficients of all explanatory variables are statistically significant with expected signs. PCI influences energy consumption in a positive way. An increase in each PCI causes additional energy consumption for each person. For example, economic growth may promote energy consumption in cities (Sarwar et al. 2017).

The regression coefficient and constant of LR used in energy consumption prediction were estimated as indicated in Table 13. The coefficient  $a_1$  and  $a_2$  mentioned in the table represent population and PCI respectively in electric energy

**Table 12** | Performance of fit evaluation for scenarios used in energy consumption prediction

| Sectors and fuel | Drivers            | Parameter |       |       |         |         |
|------------------|--------------------|-----------|-------|-------|---------|---------|
|                  |                    | CoD       | MAE   | RMSE  | RAE (%) | RSE (%) |
| Transport        | Population and PCI | 0.95      | 0.44  | 0.56  | 18.60   | 18.30   |
|                  | Population         | 0.93      | 1.00  | 1.19  | 41.10   | 39.80   |
| Commercial       | Population         | 0.94      | 21.11 | 25.83 | 13.70   | 15.40   |
|                  | PCI                | 0.94      | 21.39 | 26.11 | 13.8    | 15.60   |
| Residential      | Population         | 0.95      | 55.56 | 69.44 | 25.900  | 31.00   |
|                  | PCI                | 0.93      | 58.33 | 69.44 | 26.00   | 31.10   |
| Industrial       | Population         | 0.94      | 52.78 | 66.67 | 24.20   | 28.20   |
|                  | PCI                | 0.92      | 58.33 | 72.22 | 26.40   | 31.10   |
| Diesel           | Population and PCI | 0.94      | 0.05  | 0.01  | 6.50    | 6.60    |
|                  | GDP and PCI        | 0.97      | 0.11  | 0.01  | 12.80   | 12.90   |
| Gasoline         | Population and PCI | 0.98      | 0.09  | 0.01  | 6.80    | 6.90    |
|                  | GDP and PCI        | 0.93      | 0.25  | 0.04  | 21.00   | 24.00   |
| Kerosene         | Population and PCI | 0.53      | 0.035 | 0.004 | 70.00   | 75.00   |
|                  | GDP and PCI        | 0.56      | 0.030 | 0.004 | 59.70   | 66.10   |

**Table 13** | Regression coefficient and constant in the energy model

| Sector and fuel | Scenarios | Regression coefficient | Value | Constant | Value |
|-----------------|-----------|------------------------|-------|----------|-------|
| Transport       | 1         | $a_1$                  | -0.4  | $a_o$    | 0.03  |
|                 |           | $a_2$                  | 49.8  |          |       |
| Commercial      | 2         | $a_1$                  | 0.1   | $a_o$    | -0.1  |
|                 |           |                        |       |          |       |
| Residential     | 1         | $a_1$                  | 2.9   | $a_o$    | 7.6   |
|                 |           | $a_2$                  | 328.5 | $a_o$    | -7.1  |
| Industrial      | 2         | $a_1$                  | 3.82  | $a_o$    | -9.3  |
|                 |           | $a_2$                  | 424.7 | $a_o$    | -8.6  |
| Diesel          | 1         | $a_1$                  | 4     | $a_o$    | -9.2  |
|                 |           | $a_2$                  | 420.7 | $a_o$    | -7.7  |
| Gasoline        | 2         | $a_1$                  | 5.7   | $a_o$    | -11.6 |
|                 |           | $a_2$                  | 58.7  |          |       |
|                 | 1         | $a_1$                  | 3.9   | $a_o$    | -7.13 |
|                 |           | $a_3$                  | 31.8  |          |       |
| Kerosene        | 1         | $a_1$                  | 11.5  | $a_o$    | -29.3 |
|                 |           | $a_2$                  | -32.3 |          |       |
|                 |           | $a_3$                  | -10.4 |          |       |
| Kerosene        | 2         | $a_1$                  | 11.8  | $a_o$    | -30.3 |
|                 |           | $a_3$                  | -10.4 |          |       |
|                 |           | $a_1$                  | -1.2  | $a_o$    | 6.2   |
| Kerosene        | 2         | $a_3$                  | 14.8  |          |       |
|                 |           | $a_3$                  | 7.3   | $a_o$    | 3.3   |
|                 |           | $a_1$                  | -12.1 |          |       |

consumption. The regression coefficient values represent the mean change in the response variable for one unit of change in the predictor variable while holding other predictors in the model constant, then, the greater the coefficient, the steeper the slope, the greater change in the dependent variable. In the commercial, industrial and residential electric energy consumption, the regression coefficient of the population is much greater with a great change in consumption.

The energy consumption by sector is greatly affected by changes in population and changes in PCI. Under high PCI, energy consumption is high. Energy consumption is high when the population rate is high. The results show a high relationship between the projected population and predicted energy consumption. The 5% increase in the Addis Ababa population will result in a 4% increase in kerosene consumption. The predicted energy consumption is given in Table 14.

### Energy demand

Finally, for demand prediction, the drivers that showed the most significant statistical results are used. The electric energy demand is also influenced by technology factors

that include energy efficiency (e.g. distribution loss) and consumption. In Addis Ababa city, the loss is around 19% in 2016 and is expected to decrease in 2019 to 14.4% (DVRPC 2011). The Addis Ababa Distribution Master Plan (AADMP) has planned to improve the distribution loss as it decreases to 9% in 2034 and is expected to reach 6.65% in 2050. Based on the most influential socio-economic parameter that affects the energy consumption, the governing equation used to predict the energy demand is given in Table 15 and the values of predicted energy demand is shown in Table 16.

As shown in Table 16, the energy demand showed strong fluctuations, according to increasing GDP, PCI and population growth. The percent growth rates of electric energy demand for commercial, industrial, residential and transport sectors are indicated in Figure 9.

Electric energy demand growth rate will decrease from 2030 to 2050 for commercial, industrial and residential and from 2040 to 2050 for street-lighting (transport). The demand will decrease from 2030 to 2050 for commercial, industrial residential and street-lighting (transport) by 18, 13, 15 and 7% respectively. For the same year, the petroleum energy

**Table 14** | Energy consumption (GWh) for sectors

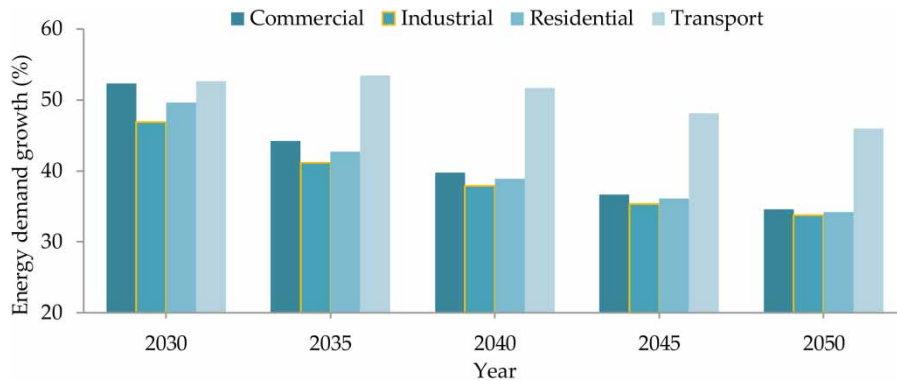
| Sector and fuel | Scenarios | Year  |       |        |        |        |        |        |
|-----------------|-----------|-------|-------|--------|--------|--------|--------|--------|
|                 |           | 2016  | 2025  | 2030   | 2035   | 2040   | 2045   | 2050   |
| Commercial      | 1         | 708   | 2,348 | 3,636  | 5,294  | 7,432  | 10,186 | 13,734 |
|                 | 2         | 708   | 2,373 | 3,707  | 5,452  | 7,734  | 10,719 | 14,621 |
| Industrial      | 1         | 1,169 | 3,416 | 5,125  | 7,360  | 10,282 | 14,104 | 19,102 |
|                 | 2         | 1,169 | 3,494 | 5,238  | 7,485  | 10,380 | 14,111 | 18,918 |
| Residential     | 1         | 1,022 | 3,225 | 4,950  | 7,206  | 10,156 | 14,013 | 19,059 |
|                 | 2         | 1,022 | 3,193 | 4,857  | 7,001  | 9,763  | 13,323 | 17,909 |
| Transport       | 1         | 14    | 33    | 53     | 81     | 125    | 186    | 272    |
|                 | 2         | 14    | 39    | 61     | 89     | 125    | 171    | 231    |
| Diesel          | 1         | 2,706 | 6,136 | 8,841  | 12,331 | 16,834 | 22,644 | 30,141 |
|                 | 2         | 2,713 | 6,286 | 9,564  | 14,403 | 21,673 | 32,786 | 50,046 |
| Gasoline        | 1         | 2,608 | 8,906 | 13,863 | 20,248 | 28,471 | 39,062 | 52,702 |
|                 | 2         | 2,610 | 8,791 | 13,502 | 19,366 | 26,574 | 35,275 | 45,499 |
| Kerosene        | 1         | 1,010 | 1,014 | 1,232  | 1,803  | 3,026  | 5,422  | 9,892  |
|                 | 2         | 1,022 | 1,283 | 1,591  | 2,129  | 3,059  | 4,658  | 7,395  |

**Table 15** | Variables and the final governing LR equations to predict the energy demand

| Notation | Transport electric energy   | Regression | Mathematical relational expressions               |
|----------|-----------------------------|------------|---|
| TRAE LC  | Transport electric energy   | MLR        | $TRAE LC = -0.4 * Population + 49.8 * PCI + 0.03$ |
| COMELC   | Commercial electric energy  | MLR        | $COMELC = 1.5 * Population + 164.3 * PCI + 0.25$  |
| RESEL C  | Residential electric energy | MLR        | $RESEL C = 1.9 * Population + 212.3 * PCI - 8.9$  |
| INDEL C  | Industrial electric energy  | MLR        | $INDEL C = 2 * Population + 210.4 * PCI - 8.4$    |
| KERC     | Kerosene consumption        | MLR        | $KERC = -1.2 * Population + 14.8 * GDP + 6.2$     |
| DSL C    | Diesel consumption          | MLR        | $DSL C = 5.7 * Population + 58.7 * PCI - 11.6$    |
| GSL C    | Gasoline consumption        | MLR        | $GSL C = 11.5 * Population - 32.3 * PCI - 29.3$   |

**Table 16** | Energy demand (GWh) prediction result

| Sector and fuel | Year  |       |        |        |        |        |        |
|-----------------|-------|-------|--------|--------|--------|--------|--------|
|                 | 2016  | 2025  | 2030   | 2035   | 2040   | 2045   | 2050   |
| Commercial      | 874   | 2,689 | 4,095  | 5,905  | 8,255  | 11,282 | 15,188 |
| Industrial      | 1,444 | 3,936 | 5,780  | 8,157  | 11,247 | 15,228 | 20,364 |
| Residential     | 1,262 | 3,655 | 5,470  | 7,806  | 10,842 | 14,754 | 19,801 |
| Transport       | 17    | 38    | 58     | 89     | 135    | 200    | 292    |
| Diesel          | 2,706 | 6,136 | 8,841  | 12,331 | 16,834 | 22,644 | 30,141 |
| Gasoline        | 2,608 | 8,906 | 13,863 | 20,248 | 28,471 | 39,062 | 52,702 |
| Kerosene        | 1,010 | 1,014 | 1,232  | 1,803  | 3,026  | 5,422  | 9,892  |



**Figure 9** | Predicted energy demand growth rate.

demand growth rate for diesel and gasoline will decrease by 11 and 21% respectively whereas kerosene will increase by 61%.

## CONCLUSIONS

As the residential, industrial, commercial, street-lighting sector contributes significantly to WE consumption and their proportion of consumption is on the rise, developing a forecasting model capable of making accurate predictions is important for WE system planning. In this regard, this paper conducted a method of WE demand model for a city considering the socioeconomic factors by building an LR model that relates the consumption to economic and social factors. The study began with identifying the possible socio-economic factors which will affect the demand. This study differs from the other method reported for forecasting the long-term demand at a city level. Here, the future demands were forecasted for each sector separately to fill a specific gap. The regression model has good accuracy in the forecasting of WE demand, as evidenced by the model evaluation tools used. The evaluation of population, GDP and PCI, as well as technology efficiency, will determine the future WE demand. The WE demand was forecasted for each year starting from 2016 to 2050. In 2030 and 2050, the electric energy demand was estimated in GWh to be 4095 and 15,188 for the commercial sector; 5780 and 20,364 for the industrial sector; and 5470 and 19,801 for the residential sector, whereas it was 58 and 292 for the street-lighting sector. Respectively, for the two years mentioned above, the water demand was estimated to be 87 and 233 MCM for the commercial, 354 and 997

MCM for the residential, and 78 and 337 MCM for the industrial sector. The present model can effectively utilize and provide relevant data on the projected socio-economic and technology factors. These data can be used as an input in determining the urban WE supply to meet up with the demand.

The results of this study can be used as input for policy makers and researchers for their further study. Since the LR model gave acceptable values, one can comfortably use the model implemented in this paper to predict the long-term water and energy demand.

## ACKNOWLEDGEMENTS

The authors would like to thank Addis Ababa Water and Sewerage Authority (AAWSA), Central Statistical Agency (CSA), Ethiopian Electric Utility (EEU), and Ethiopian Petroleum Enterprise (EPE) for providing data.

## DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

## REFERENCES

AAWSA 2008 *Water Production and Distribution in Addis Ababa (Unpublished Data)*. Addis Ababa Water Supply and Sewerage Authority, Addis Ababa.

- AAWSA 2019 *Consultancy Service for Addis Ababa Water Distribution and Operation Management and Hydraulic Modeling (Unpublished Report)*. Addis Ababa, Ethiopia.
- Abdul-Wahab, S., Bakheit, C. & Al-Alawi, S. 2005 Principle component and multiple regression analysis in modeling of ground-level ozone and factors affecting its concentration. *Environmental Modelling and Software* **20**, 1263–1271.
- Al-Ghandoora, A. A.-H. 2008 Electricity consumption and associated GHG emissions of the Jordanian industrial sector. Empirical analysis and future projection. *Energy Policy* **36**, 258–267.
- Ali Kaab, M. S.-P.-w. 2019 Combined life cycle assessment and artificial intelligence for prediction of output energy and environmental impacts of sugarcane production. *Science of the Total Environment* **664**, 1005–1019.
- Al-Musaylh, M. S. 2018 Advanced engineering informatics short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Advanced Engineering Informatics* **35**, 1–16.
- Arturo Morales, A. 2014 Forecasting future energy demand: electrical energy in Mexico as an example case. *Energy Procedia* **57**, 782–790.
- Ashkan Nabavi-Pelesaraei, S. R. 2019 Comprehensive model of energy, environmental impacts and economic in rice milling factories by coupling adaptive neuro-fuzzy inference system and life cycle assessment. *Journal of Cleaner Production* **217**, 742–756.
- Autar, K. & Egwu, E. 2010 *Numerical Methods with Applications*. Lulu Press, Inc., North Carolina, USA.
- Balance, D. 2012 *Assumptions in Multiple Regression*. Available from: [www.dianneballanceportfolio.com/uploads/1/2/8/2/12825958/assumptions\\_in\\_multipleregression](http://www.dianneballanceportfolio.com/uploads/1/2/8/2/12825958/assumptions_in_multipleregression)
- Bianco, V. M. 2009 Electricity consumption forecasting in Italy using linear regression models. *Energy* **34**, 1413–1421.
- BoFED 2016 *Revenue Enhancement Plan of the City Administration*. Addis Ababa. Available from: [www.unhabitat.org](http://www.unhabitat.org)
- Cevik, A. 2007 Unified formulation for web crippling strength of cold-formed steel sheeting using stepwise regression. *Journal of Constructional Steel Research* **63**, 1305–1316.
- Chau, C. W. 2013 Prediction of rainfall time series using modular soft computing methods. *Engineering Applications of Artificial Intelligence* **26**, 997–1007.
- Christian-Smith, J. H. 2012 *Urban Water Demand in California to 2100: Incorporating Climate Change*. Pacific Institute, California, USA.
- Cleveland, C. R. 2000 Aggregation and the role of energy in the economy. *Ecological Economics* **32**, 301–318.
- CSA 2006 *Census of Ethiopia 2006. Section B: Population, Central Statistics Authority*. Census Administrative Report, Addis Ababa, Ethiopia.
- CSA 2015 *Urban Unemployment Survey*. Census Administrative Report, Addis Ababa, Ethiopia.
- Csereklyei, Z. A. 2012 *Modeling Primary Energy Consumption Under Model Uncertainty*. Department of Economics WP 147, Vienna, Austria.
- Debbie, J. D. & Maria-Pia, V. 2013 Robust VIF regression with application to variable selection in large data sets. *Applied Statistics* **7**, 319–341.
- Dimitrios, S. K., Eleni, M. & Donal, P. F. 2017 Input variable selection for thermal load predictive models of commercial buildings. *Energy and Buildings* **137**, 13–26.
- Donkor, E. M. 2012 Urban water demand forecasting: a review A review of methods and models. *Journal of Water Resources Planning and Management* **140**, 146–159. [Pub. ahead of print doi: [http://dx.doi.org/10.1061/\(ASCE\)WR.1943-5452.0000314](http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000314)].
- Donkor, E., Mazzuchi, T., Soyer, R. & Roberson, J. 2014 Urban water demand forecasting: review of methods and models. *Journal of Water Resources Planning and Management* **140**, 146–159.
- DVRPC 2011 *Greater Philadelphia's Food System Plan*. Delaware Valley Regional Planning Commission, Philadelphia, USA.
- Ebru Eris, H. A. 2019 Frequency analysis of low flows in intermittent and non-intermittent rivers. *Water Science & Technology Water Supply* **19**, 30–39.
- Ethiopia Ministry of Transport 2011 *Transport Policy of Addis Ababa*. Available from: <https://chilot.me/wp-content/uploads/2011/08/> (accessed 19/6/2018).
- Feinberg, E. A. & Genethliou, D. 2005 Load forecasting. In: *Applied Mathematics for Power Systems*. State University of New York, Stony Brook, Chapter 12.
- George, A. M. & Dallas, E. J. 2002 *Analysis of Messy Data*. Chapman and Hall/CRC, New York, USA.
- Hamzacebi, C. & Es, H. 2014 Forecasting the annual electricity consumption of Turkey using an optimized grey. *Energy* **70**, 165–171.
- Hao, L., Yuhuan, Z., Jia-Ning, K. & Song, W. 2020 Identifying sectoral energy-carbon-water nexus characteristics of China. *Journal of Cleaner Production* **249**, 1–13.
- Haque, M. R. 2018 Assessment of variable selection methods in urban water demand forecasting. *Water* **10**, 1–15.
- He, Y., Jiao, J., Chen, Q., Ge, S., Chang, Y. & Xu, Y. 2017 Urban long term electricity demand forecast method based on system dynamics of the new economic normal: the case of Tianjin. *Energy* **133**, 9–22.
- Hong, T. F. 2016 Probabilistic electric load forecasting: a tutorial review. *International Journal of Forecasting* **32**, 914–938.
- Hong, T., Pinson, P. & Fan, S. 2014 Global energy forecasting competition. *International Journal of Forecasting* **30**, 357–363.
- Hossein, S. & Mohammad, S. S. 2011 *Electric Power System Planning*. Springer, Heidelberg Dordrecht London New York, pp. 45–50.
- Ian, H. A. 2005 *Data Mining Practical Machine Learning Tools and Techniques*, 2nd edn. University of Waikato, New Zealand.
- Isaac, A. S., Emmanuel, A. & Ishioma, O. 2017 A comparative study of regression analysis and artificial neural network

- methods for medium-term load forecasting. *Journal of Science and Technology* **10**, 1–8.
- JICA 2018 *Addis Ababa Distribution Master Plan (Unpublished Report)*. Final report Available from: [http://open\\_jicareport.jica.go.jp/pdf](http://open_jicareport.jica.go.jp/pdf)
- Kamínski, B., Jakubczyk, M. & Szufel, P. 2018 A framework for sensitivity analysis of decision trees. *Central European Journal* **26**, 135–159.
- Khan, Z. & Jayaweera, D. 2018 Approach for forecasting smart customer demand with significant energy demand variability. In: *Proceedings of the 2018 1st International Conference on Power, Energy and Smart Grid (ICPESG)*, Mirpur Azad Kashmir, Pakistan, pp. 1–5.
- Kitessa, B. D., Ayalew, S. M., Gebrie, G. S. & Teferi, S. 2020 Quantifying water-energy nexus for urban water systems: a case study of Addis Ababa city. *AIMS Environmental Science* **7** (6), 486–504.
- Kraft, J. A. 1978 Relationship between Energy and GNP. *Journal of Energy and Development* **3**, 401–403.
- Krause, P., Boyle, D. & Båse, F. 2005 Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* **5**, 89–97.
- Kunder, D. 2001 *Multi-resolution Digital Watermarking Algorithms and Implications for Multimedia Signals. PhD Thesis*, University of Toronto, Canada.
- Lahouar, A. & Le, B. 2015 Day-ahead load forecast using random forest and expert input selection. *Energy Conversion & Management* **103**, 1040–1052.
- Lee, J. & Kim, J. 2018 Assessing strategies for urban climate change adaptation: the case of six metropolitan cities in South Korea. *Sustainability* **10**, 1–31.
- Lei, K. & Wang, Z. 2008 Energy synthesis and simulation for Macao. *Energy* **33**, 613–625.
- Leng, X., Wang, J., Ji, H., Wang, Q., Li, H., Qian, X. & Yang, M. 2017 Prediction of size-fractionated airborne particle-bound metals using MLR, BP-ANN and SVM analyses. *Chemosphere* **180**, 513–522.
- Lewis, C. 1982 *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. Butterworth Scientific, London, UK. Boston, MA, USA.
- Luvimba Ramulongo, D. N. 2017 Nature of urban household and water demand and consumption in Makhado Local Municipality: a case study of Makhado Newtown. *Procedia Environmental Science* **37**, 182–194.
- Makridakis, S. W. 1998 *Forecasting: Methods and Applications*, 3rd edn. John Wiley & Sons Inc., New York, USA.
- Mengistu, T. 2010 *Assessing Water Conservation and Demand Management Option for Addis Ababa City*. Addis Ababa University Libraries, Addis Ababa, Ethiopia.
- Miller, T. & Kane, M. 2001 The precision of change scores under absolute and relative interpretations. *Applied Measurement in Education* **14**, 307–327.
- Mir, A. A., Alghassab, M., Ullah, K., Khan, Z. A., Lu, Y. & Imran, M. 2020 A review of electricity demand forecasting in low and middle income countries: the demand determinants and horizons. *Sustainability* **12**, 1–35.
- Mohammed, J. 2017 *Long Term Demand-Forecasting and Generation Expansion Planning of the Ethiopian Electric Power (EEP)*. Addis Ababa University Libraries, Addis Ababa, Ethiopia.
- Moslem, A., Mehdi, Y. H., Seyyed, H. & Amir, H. 2019 Artificial neural networks and multiple linear regression as potential methods for modeling seed yield of safflower (*Carthamus tinctorius* L.). *Industrial and Crops Products* **127**, 185–194.
- MUDHCo 2015 *National Report on Housing and Sustainable Urban Development*. Federal Democratic Republic of Ethiopia. Available from: <http://www.mudc.gov.et>
- Muhammad Shahid, K. B. 2020 Impact assessment of land use and climate changes on the variation of runoff in Margalla Hills watersheds, Pakistan. *Arabian Journal of Geosciences* **13**, 1–15.
- Najafi, B. A. 2018 Application of ANNs, ANFIS and RSM to estimating and optimizing the parameters that affect the yield and cost of biodiesel production. *Engineering Applications of Computational Fluid Mechanics* **12** (1), 611–624.
- Nasir, M. & Ur Rehman, F. 2011 Environmental Kuznets Curve for carbon emissions in Pakistan: an empirical investigation. *Energy Policy* **39**, 1857–1864.
- Nefeslioglu, H. A. & Sezer, E. 2010 Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. *Mathematical Problems in Engineering* **2010**, 1–15.
- Nhamo, L., Ndlela, B. & Nhemachena, C. 2018 The water-energy-food nexus: climate risks and opportunities in Southern Africa. *Water* **10**, 1–18.
- Quilty, J. 2018 Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. *Journal of Hydrology* **563**, 336–353.
- Ramirez, D. A. 2013 Variance inflation in regression. *Advances in Decision Sciences* **2013**, 1–15.
- Renno, C. P. 2016 ANN model for predicting the direct normal irradiance and the global radiation for a solar application to a residential building. *Journal of Cleaner Production* **135**, 1298–1316.
- Riccardo, T. & Kwok-Wing, C. 2015 ANN-based interval forecasting of stream flow discharges using the LUBE method and MOFIPS. *Engineering Applications of Artificial Intelligence* **45**, 429–440.
- Rooijen, V. 2011 *Implications of Urban Development for Water Demand, Wastewater Generation and Reuse in Water-Stressed Cities: Case Studies From South Asia and sub-Saharan Africa*. Thesis IDRC. IDL-53371.pdf
- Rookmoney, T., Geoffrey, H., Surendra, T. & Stanley, O. 2019 Factors contributing towards high water usage within poor communities in Kwazulu-Natal, South Africa. *WIT Transactions on Ecology and the Environment* **239**, 1–10.

- Sailor, D. & Muñoz, J. 1997 Sensitivity of electricity and natural gas consumption to climate in the USA – methodology and results for eight states. *Energy* **22**, 987–998.
- Sarwar, S., Chen, W. & Waheed, R. 2017 Electricity consumption, oil price and economic growth: global perspective. *Renewable and Sustainable Energy Reviews* **76**, 9–18.
- Sayim, K. A. 2013 Scenarios as channels of forecast advice. *Technological Forecasting and Social Change* **80**, 772–788.
- Shahabbodin Shamshirband, S. H. 2020 Predicting standardized streamflow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics* **14** (1), 339–350.
- Sharma, A. K. 2015 Evaluating WEKA over the Open Source Web Data Mining Tools. *International Journal of Engineering Research and Technology* **8**, 128–132.
- Sullivan, R. 1977 *Power System Planning*. McGraw-Hill, New York, USA.
- Swasti, R. K., Jose, L. R. & Mart, A. M. 2018 Long term load forecasting considering volatility using multiplicative error model. *Energies* **11**, 1–19.
- Tiwari, M. & Adamowski, J. 2013 Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resources Research* **49**, 6486–6507.
- Toth, E. B. 2000 Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of Hydrology* **239**, 132–147.
- UN 2018 *Department of Economic and Social Affairs, Population Division (2018). World Population Prospects*. United Nations, New York, USA.
- UNESCO 2006 *Water A Shared Responsibility. The United Nations World Water Development Report 2*. United Nations Educational Scientific and Cultural Organization, Paris.
- Available from: <http://www.unesco.org/water/wwap/wwdr/wwdr2> (accessed 17 June 2010).
- Un-Habitat 2008 *State of the World's Cities 2008/2009. Harmonious Cities*. UN-Habitat, Nairobi. Available from <http://www.unhabitat.org/pmss/listItemDetails.aspx?publicationID=2562> (accessed 10 June 2010).
- UN-Habitat 2017a *The State of Addis Ababa*. UN-Habitat, Nairobi, Kenya.
- UN-Habitat 2017b *United Nations Human Settlements Programme, Addis Ababa, Ethiopia*. UN-Habitat, Nairobi, Kenya.
- Waters, J. W. 2002 Four assumptions of multiple regressions that researchers should always test. *Journal of Practical Assessment, Research, and Evaluation* **8**, 1–5.
- Willmott, C. J. & Matsuura, K. 2005 Advantages of the mean absolute error (MAE) over advantages of the mean absolute error (MAE) over average model performance. *Climate Research* **30**, 79–82.
- World Bank 2015a *Enhancing Urban Resilience in Addis Ababa: Resilient Cities Program; Document Number 100980*. World Bank Group, Addis Ababa, Ethiopia.
- World Bank 2015b *Enhancing Urban Resilience*. GFDRR, Addis Ababa.
- World Bank 2015c *Ethiopia Urbanization Review: Urban Institutions for A Middle-Income Ethiopia*. Available from: <http://www.cs.waikato.ac.nz/ml/weka/>.
- Xin Yan, X. G. 2009 *Theory and Computing: Linear Regression Analysis*. World Scientific Publishing, Singapore.
- Youngmin Seo, S. K. 2018 Short-term water demand forecasting model combining variational mode decomposition and extreme learning machine. *Hydrology* **5**, 1–19.
- Ziyi Yin, B. J. 2018 Comprehensive forecast of urban water-energy demand based on a neural network model. *Water* **10**, 1–16.

First received 13 January 2021; accepted in revised form 23 March 2021. Available online 28 April 2021